PENGANTAR INTELEGENSI BISNIS

Dr. Joseph Teguh Santoso, S.Kom., M.Kom.

YAYASAN PRIMA AGUS TEKNIK
UNIVERSITAS STEKOM

PENGANTAR INTELEGENSI BISNIS

Dr. Joseph Teguh Santoso, S.Kom., M.Kom.

Kerjasama Penerbit Dengan UNIVERSITAS STEKOM

YAYASAN PRIMA AGUS TEKNIK

PENERBIT :
YAYASAN PRIMA AGUS TEKNIK
Jl. Majapahit No. 605 Semarang
Telp. (024) 6723456. Fax. 024-6710144
Email : penerbit_ypat@stekom.ac.id

PENGANTAR INTELEGENSI BISNIS

Penulis :
Dr. Joseph Teguh Santoso, S.Kom., M.Kom.

ISBN : 9 786239 504267

Editor :
Dr. Joseph Teguh Santoso, S.Kom., M.Kom.

Penyunting :
Dr. Mars Caroline Wibowo. S.T., M.Mm.Tech

Desain Sampul dan Tata Letak :
Irdha Yunianto

Penebit :
Yayasan Prima Agus Teknik Bekerja sama dengan
Universitas STEKOM

Redaksi :
Jl. Majapahit no 605 Semarang
Telp. (024) 6723456
Fax. 024-6710144
Email : penerbit_ypat@stekom.ac.id

Distributor Tunggal :
Universitas STEKOM
Jl. Majapahit no 605 Semarang
Telp. (024) 6723456
Fax. 024-6710144
Email : info@stekom.ac.id

Kata Pengantar

Puji syukur kami panjatkan kehadirat Tuhan Yang Maha Esa karena dengan rahmat,karunia, serta taufik dan hidayah-Nya kami dapat menyelesaikan penyusunan buku ***PENGANTAR INTELEGENSI BISNIS*** dengan harapan untuk dapat dipergunakan oleh kalangan para akademisi.

Tujuan utama penyusunan buku ini adalah untuk memudahkan mahasiswa dalam memahami dan menguasai dasar dasar ilmu Intelegensi Bisnis. Buku teks ini memberikan gambaran tentang Intelegensi Bisnis yang disertai berbagai analisis yang lengkap, sehingga memudahkan para mahasiswa untuk memahami, menerapkan aplikasi perangkat lunak yang berhubungan dengan analisis statistik data.

Semoga buku ini dapat dipahami bagi siapapun yang membacanya. Sekiranya buku yang telah disusun ini dapat berguna bagi kami sendiri maupun orang yang membacanya. Sebelumnya kami mohon maaf apabila terdapat kesalahan kata-kata yang kurang berkenan dan kami memohon kritik dan saran yang membangun demi perbaikan di masa depan.

Semarang, Desember 2020

Dr. Joseph Teguh Santoso, S.Kom, M.Kom
Penulis

DAFTAR ISI

Business Intellegince dan Analisis

Pendahaluan

Data merupakan suatu kumpulan yang terdiri dari fakta-fakta untuk memberikan gambaran yang luas terkait dengan suatu keadaan. Seseorang yang akan mengambil sebuah kebijakan atau keputusan umumnya akan menggunakan data sebagai bahan pertimbangan. Melalui data seseorang dapat menganalisis, menggambarkan, atau menjelaskan suatu keadaan. Intelijen bisnis mengekstrak informasi dari data mentah melalui alat-alat seperti penggalian data, analisis perspektif, pemrosesan analitik online, dll. IB bertujuan untuk memudahkan interpretasi jumlah data yang besar tersebut. Dengan mengidentifikasi kesempatan baru dan mengimplementasikan strategi yang efektif berdasarkan wawasan dapat membuat bisnis memiliki keuntungan pasar yang kompetitif dan stabilitas jangka panjang. Buku teks akan menyediakan informasi komprehensif kepada pembaca tentang intelijen bisnis dan analitik. Buku ini mengeksplorasi semua aspek penting dari intelijen bisnis dan analitik di masa kini skenario. Topik-topik yang dicakup dalam buku ini membahas banyak hal pokok bisnis intelijen. Ini bertujuan untuk berfungsi sebagai panduan sumber daya bagi siswa dan memfasilitasi pembelajaran disiplin.

Berikut ini adalah deskripsi bab demi bab yang di bahas dalam buku ini:
Bab 1- Strategi dan perencanaan yang tergabung dalam bisnis apa pun dikenal sebagai intelejensi bisnis. Ini juga dapat mencakup produk, teknologi, analisis, dan presentasi informasi bisnis. Bab ini akan memberikan pemahaman bisnis yang terintegrasi intelijen.

Bab 2- Analisis adalah pemahaman dan komunikasi pola-pola data yang signifikan. Analytics diterapkan dalam bisnis untuk meningkatkan kinerja mereka. Beberapa aspek dijelaskan dalam teks ini adalah analisis perangkat lunak, analisis tertanam, analitik pembelajaran, dan media social analitik. Bagian analitik menawarkan fokus wawasan, dengan mengingat kompleksnya materi pelajaran.

Bab 3- Proses memahami pola yang ditemukan dalam set data besar dikenal sebagai penggalian data. Beberapa aspek penggalian data yang telah dijelaskan di bagian berikut adalah pembelajaran aturan asosiasi, analisis cluster, analisis regresi, summarization otomatis dan contoh-contoh penggalian data. Bab tentang penggalian data menawarkan focus wawasan, menjaga dalam pikiran subjek yang kompleks.

Bab 4- Gudang data adalah inti dari intelijen bisnis. Ini digunakan untuk pelaporan dan menganalisis data. Data mart, manajemen data master, dimensi, dimensi yang berubah perlahan dan skema bintang. Teks ini menjelaskan teori dan prinsip penting dari pergudangan data.
Bab 5- Upaya yang dilakukan untuk mengumpulkan informasi yang terkait dengan pelanggan atau pasar dikenal sebagai riset pasar. Riset pasar adalah bagian penting dari strategi bisnis. Segmentasi pasar, tren pasar, analisis SWOT dan riset pasar adalah beberapa topik yang dijelaskan dalam hal ini bab.

Bab 6- Aspek penting dari intelijen bisnis adalah analisis konteks, kinerja bisnis manajemen, penemuan proses bisnis, sistem informasi, kecerdasan organisasi dan proses penggalian. Metode untuk menganalisis lingkungan bisnis apa pun dikenal sebagai konteks analisis. Topik yang dibahas dalam bagian ini sangat penting untuk memperluas yang sudah ada pengetahuan tentang intelijen bisnis.

Bab 7- Kecerdasan operasional memiliki sejumlah aspek yang telah dijelaskan dalam hal ini bab. Beberapa fitur ini adalah pemrosesan acara yang kompleks, manajemen proses bisnis, metadata dan analisis akar penyebab. Komponen yang dibahas dalam teks ini sangat penting untuk memperluas pengetahuan yang ada tentang intelijen operasional.

Bab 1 : Pengantar Intelejensi Bisnis

1.1 Intelejensi Bisnis

Perkembangan usaha bisnis baik bisnis produk maupun bisnis jasa berkembang sangat pesat. Untuk mencapai tujuan bisnis tanpa strategi hanyalah mimpi. Usaha yang dibangun tidak akan bertahan lama tanpa strategi yang terencana dengan baik. Dengan meningkatnya persaingan, strategi dalam bisnis menjadi sangat penting untuk perencanaan dan pengembangan bisnis yang berkelanjutan. Diperlukan landasan dan berbagai macam teknik untuk meningkatkan bisnis yang dijalankan agar menjadi semakin besar. Salah satunya adalah dengan menguasai dan memahami Intelejensi Bisnis.

Strategi dan perencanaan yang dimasukkan dalam bisnis apa pun dikenal sebagai *Business Intelligence/*BI atau Intelijen Bisnis (IB). Mungkin juga termasuk produk, teknologi dan analisis dan presentasi informasi bisnis. Bab ini akan memberikan pemahaman terpadu tentang intelijen bisnis.

Bagi perusahaan, IB dapat digunakan untuk mendukung keputusan bisnis mulai dari operasi sampai strategis. Keputusan operasi termasuk penempatan dan harga produk, sedangkan keputusan strategis termasuk prioritas, tujuan, dan arah pada tingkat yang lebih luas. IB akan jadi lebih efektif bila digabungkan dengan data yang didapat dari pasar tempat perusahaan beroperasi (data eksternal) dengan data dari sumber internal bisnis perusahaan seperti data operasi dan finansial (data internal). Bila digabungkan, data eksternal dan internal bisa menyediakan gambaran yang lebih lengkap, yang efeknya, menciptakan “inteligensi” yang tidak dapat diturunkan dari kumpulan data tunggal manapun.

Pengertian

Business Intelligence (BI) atau Inteligensi Bisnis (IB) adalah sekumpulan teknik dan alat untuk mentransformasi data mentah menjadi informasi yang berguna dan bermakna untuk tujuan analisis bisnis. Teknologi ini dapat menangani data yang tak terstruktur dalam jumlah yang sangat besar untuk membantu mengidentifikasi, mengembangkan, dan membuat strategi bisnis yang baru. Fungsi BI sering di kaitkan dengan stilah “*data surfacing*”. Teknologi BI mampu menangani sebagian besar data terstruktur dan kadang-kadang tidak

terstruktur untuk membantu mengidentifikasi, mengembangkan dan menciptakan strategi peluang bisnis baru. Tujuan BI adalah untuk memudahkan penafsiran volume data yang besar ini. Identifikasi peluang baru dan menerapkan strategi yang efektif berdasarkan wawasan dapat memberikan keunggulan bisnis di pasar ekonomi yang kompetitif dan menjaga stabilitas jangka panjang.

Teknologi BI memberikan pandangan historis, terkini, dan prediksi operasi bisnis. Fungsi utama teknologi intelijen bisnis adalah pelaporan, pemrosesan analisis online, analisis, data pertambangan, proses pertambangan, kompleks peristiwa pengolahan, bisnis kinerja pengelolaan, pembandingan, teks pertambangan, prediktif analisis dan preskriptif analisis.

BI dapat digunakan untuk mendukung berbagai keputusan bisnis mulai dari operasional hingga strategis. Keputusan operasi dasar termasuk produk posisi atau harga. Bisnis strategis keputusan termasuk prioritas, tujuan dan arah di terluas tingkat. Dalam semua kasus, BI paling efektif kapan itu menggabungkan data yang berasal dari pasar di mana perusahaan beroperasi (data eksternal) dengan data dari sumber perusahaan internal ke bisnis seperti data keuangan dan operasi (internal data). Ketika digabungkan, data eksternal dan internal dapat memberikan gambaran yang lebih lengkap, di mana efek, menciptakan “kecerdasan” yang tidak dapat diturunkan oleh set data tunggal. Di antara banyak sekali menggunakan, alat BI memberdayakan organisasi untuk mendapatkan wawasan tentang pasar baru, menilai permintaan dan kesesuaian produk dan layanan untuk pasar yang berbeda segmen dan mengukur dampak pemasaran upaya.

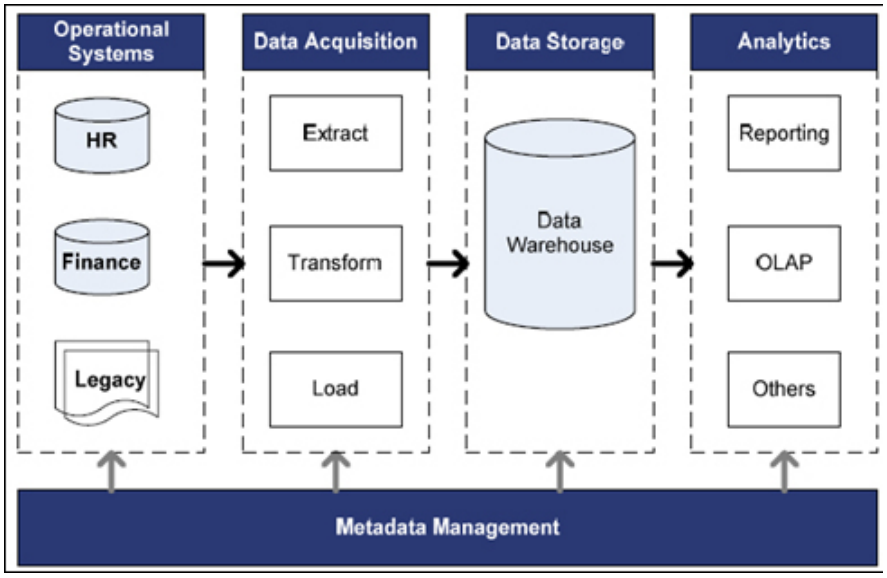
Komponen Intelejensi Bisnis

Pada umumnya sistem Intelejensi Bisnis terdiri dari empat level komponen dan modul manajemen metadata. Arsitektur general dari sistem Intelejensi Bisnis terlampir pada gambar 1. Komponen-komponen saling berinteraksi untuk memfasilitasi fungsi dasar Intelejensi Bisnis: mengekstrak data dari sistem operasional perusahaan, menyimpan data yang sudah diekstrak kedalam datawarehouse, dan menarik data yang disimpan untuk berbagai aplikasi analisis bisnis.

Intelejensi Bisnis terdiri dari sejumlah komponen yang meningkat termasuk:

- Alokasi dan agregasi multidimensi
- Denormalisasi, penandaan dan standarisasi
- Pelaporan seketika dengan peringatan analitis

- Sebuah metode antarmuka terhadap sumber data tak terstruktur
- Perkiraan konsolidasi grup, anggaran dan perpindahan (pegawai)
- Inferensi statistik dan simulasi probabilitas
- Optimalisasi kunci indikator performansi
- Pengontrolan versi dan manajemen proses
- Manajemen *Open Item*



Gambar 1.1 *Komponen Utama Intelejensi Bisnis*

Sejarah Awal Intelejensi Bisnis

Istilah BI atau business intelligence ini pada awalnya ditemukan pada tahun 1865 oleh Richard Millar Devens dalam bukunya yang berjudul *Cyclopedia of Commercial and business Anecdotes*. Istilah yang digunakan oleh Devens ini berawal ketika seorang banker bernama Sir Henry Furnese memperoleh profit dengan cara memainkan sebuah informasi mengenai lingkungannya terlebih dahulu sebelum melangkah pada kompetitornya. Kemampuan untuk mengumpulkan dan mengenal baik lingkungannya inilah yang membuat bisnis furnese menjadi berkembang. Seringkali banyak orang yang memulai bisnis menggunakan pedoman bahwa mereka harus lebih mengenal kompetitor terlebih dahulu ketimbang lingkungan tempat usaha yang akan dibangun tersebut.

Dalam artikel tahun 1958, peneliti dari IBM Hans Peter Luhn menggunakan istilah inteligensi bisnis. Dia menggunakan definisi kamus Webster tentang inteligensi yaitu : "kemampuan untuk memahami hubungan mendalam dari fakta yang ada dengan suatu cara sebagai panduan aksi terhadap tujuan yang diinginkan."

Business Intelligence seperti yang dipahami sekarang dikatakan telah berkembang dari Sistem Pendukung Keputusan (SPK) yang mulai dari tahun 1960-an dan berkembang sepanjang pertengahan 1980-an. SPK berasal dari model dibantu-komputer yang dibuat untuk membantu dalam pembuatan keputusan dan perencanaan. Dari SPK, gudang data, Sistem Informasi Eksekutif, OLAP dan inteligensi bisnis muncul menjadi fokus pada akhir 80-an.

Di tahun 1988, konsorsium Itali-Belanda-Prancis-Inggris melaksanakan pertemuan internasional tentang Analisis Data Ragamcara di Roma. Tujuan utamanya yaitu untuk mereduksi beragam dimensi menjadi satu atau dua (dengan mendeteksi pola pada data) yang dapat dipresentasikan pada pembuat-keputusan manusia.

Pada tahun 1989, Howard Dresner (kemudian sebagai analis Gartner Group) mengajukan "Business Intelligence" sebagai istilah umum untuk menjelaskan "konsep dan metode untuk meningkatkan pembuatan keputusan bisnis dengan menggunakan sistem bantu berdasar-fakta. Baru pada akhir 1990-an penggunaan ini menyebar luas.

Gudang Data / *Data Warehouse*

Seringkali aplikasi BI menggunakan data yang dikumpulkan dari *data warehouse* (DW) atau dari *data mart*, dan data konsep BI dan DW terkadang digabungkan dengan sebutan “BI / DW” atau sebagai “BIDW”. Gudang data berisi salinan data analisis yang memfasilitasi dukungan keputusan. Namun, tidak semua gudang data melayani untuk intelijen bisnis, juga tidak semua aplikasi intelijen bisnis memerlukan *data warehouse*.

Untuk membedakan antara konsep intelijen bisnis dan gudang data, *Forrester Research* mendefinisikan intelijen bisnis dengan dua cara :

1. Definisi luas

   “*Intelejensi Bisnis* adalah seperangkat metodologi, proses, arsitektur, dan teknologi yang menguba h data mentah menjadi informasi yang bermakna dan berguna biasanya aktifkan lebih banyak efektif strategis, taktis, dan wawasan operasional dan pengambilan keputusan. “ Dibawah ini definisi, bisnis intelijen juga termasuk teknologi seperti itu sebagai integrasi data, kualitas data, pergudangan data, manajemen data master, teks dan analisis konten, dan banyak yang lainnya itu pasar terkadang benjolan ke informasi Pengelolaan” segmen. Karena itu, Forrester mengacu pada data persiapan

dan data pemakaian sebagai dua yang terpisah tapi rapat ditautkan segmen dari Intelejensi Bisnis arsitektur tumpukan.

2. Definisi Sempit

   Dalam arti yang lebih sempit *Forrester* mendefinisikan pasar intelijen bisnis sebagai, "... sesuatu yang hanya merujuk pada bagian utama dari tumpukan arsitektur BI seperti pelaporan, analisis, dan dasbor. "

Intelejensi Bisnis dengan Kecerdasan Kompetitif

Istilah Intelejensi Bisnis terkadang merupakan sinonim dari intelijen kompetitif (karena keduanya mendukung pengambilan keputusan), BI menggunakan teknologi, proses, dan aplikasi untuk menganalisis sebagian besar internal, data terstruktur dan proses bisnis sementara kompetitif intelijen mengumpulkan, menganalisis dan menyebarluaskan informasi dengan fokus topikal pada perusahaan pesaing. Jika dipahami secara luas, intelijen bisnis dapat mencakup subset intelijen kompetitif.

Intelejensi Bisnis dengan Analisis Bisnis

Intelejensi Bisnis dan analisis bisnis terkadang digunakan secara bergantian, tetapi ada definisi alternatif. Satu definisi kontras keduanya, menyatakan bahwa istilah Intelejensi Bisnis mengacu pada pengumpulan data bisnis untuk menemukan informasi terutama melalui mengajukan pertanyaan, melaporkan, dan proses analisis online. Analisis bisnis, di sisi lain, menggunakan alat-alat secara statistik dan kuantitatif untuk pemodelan penjelas dan prediksi.

Dalam definisi alternatif, Thomas Davenport, profesor teknologi dan manajemen informasi di *Babson College* berpendapat bahwa intelijen bisnis seharusnya dibagi menjadi query, pelaporan, *Online analytical processing* (OLAP), sebuah alat "Peringatan" dan analisis bisnis.Dalam pengertian ini, analisis bisnis adalah himpunan bagian dari BI yang berfokus pada statistik, prediksi, dan optimisasi, bukan hanya sebagai fungsi pelaporan.

Penerapan Intelejensi Bisnis dalam Perusahaan

Intelejensi Bisnis dapat diterapkan untuk tujuan bisnis berikut, untuk mendorong nilai bisnis.

1. Pengukuran - program yang menciptakan hierarki metrik kinerja dan tolok ukur- yang menginformasikan para pemimpin bisnis tentang kemajuan menuju sasaran bisnis (proses bisnis pengelolaan).

2. Analisis - program yang membangun proses kuantitatif agar suatu bisnis mencapai optimal keputusan dan untuk melakukan penemuan pengetahuan bisnis. Sering melibatkan: penggalian data (*data mining*), proses pengumpulan informasi (*mining process*), analisis statistik, analisis prediktif, prediktif pemodelan, Pemodelan Proses Bisnis, *data lineage*, kompleksitas pemrosesan kegiatan dan analisis preskriptif
3. Pelaporan / pelaporan perusahaan - program yang membangun infrastruktur untuk pelaporan strategis dalam pelayanan manajemen strategis bisnis, bukan pelaporan operasional. Program ini melibatkan visualisasi data, sistem informasi eksekutif dan OLAP.
4. Kolaborasi / kolaborasi platform - program yang mendapat area yang berbeda (baik di dalam dan di luar bisnis) untuk bekerja bersama melalui pertukaran data dan pertukaran data elektronik.
5. Manajemen pengetahuan - program untuk membuat data perusahaan dijalankan melalui strategi dan praktik untuk mengidentifikasi, membuat, mewakili, mendistribusikan, dan memungkinkan adopsi wawasan dan pengalaman yang merupakan pengetahuan bisnis sejati. Manajemen pengetahuan mengarah pada pembelajaran kepatuhan manajemen dan peraturan.

Selain hal di atas, intelijen bisnis dapat diterapkan sebagai pendekatan proaktif, seperti peringatan fungsionalitas yang segera memberi tahu pengguna akhir jika kondisi tertentu terpenuhi. Misalnya, jika beberapa metrik bisnis melebihi ambang yang telah ditentukan, metrik akan disorot dalam standar laporan, dan analis bisnis dapat diberitahu melalui email atau layanan pemantauan lainnya. Ini proses end-to-end membutuhkan tata kelola data, yang harus ditangani oleh ahli.

Intelejensi Bisnis sebagai Prioritas Proyek

Mungkin sulit untuk memberikan kasus tentang penggunaan intelijen bisnis sebagai sebuah inisiatif dan seringkalisebuah prioritas proyek harus ditentukan melalui inisiatif strategis. Proyek BI dapat menjadi sebuah prioritas yang lebih tinggi dalam organisasi jika manajer mempertimbangkan hal-hal sebagai berikut:

- Manajer BI harus menentukan manfaat nyata dari penerapan IB ini, misalnya dapat memangkas bahkan menghilangkan biaya produksi Report Pengolahan.
- Pemberlakuan akses data bagi seluruh organisasi Dengan cara ini bahkan jika hanya sebentar, seperti menghemat beberapa "menit", dapat membuat perbedaan ketika dikalikan dengan jumlah karyawan di seluruh organisasi.

- Manajer juga harus mempertimbangkan bahwa proyek BI dapat mendorong memberikan ruang terhadap inisiatif bisnis lain yang akan menghasilkan bisnis yang bermutu. Untuk mendukung pendekatan ini, organisasi harus memiliki arsitek-arsitek pemograman yang dapat mengidentifikasi proyek bisnis yang sesuai bagi perusahaan.
- Menggunakan metodologi terstruktur dan kuantitatif untuk membuat prioritas yang dapat dipertahankan sejalan dengan kebutuhan aktual organisasi, seperti matriks keputusan berbobot.

1.2 Faktor Keberhasilan Penerapan IB

Menurut Kimball et al., Ada tiga area kritis yang harus dinilai oleh organisasi sebelum benar-benar siap untuk melakukan proyek BI:

1. Tingkat komitmen dan dukungan dari manajemen puncak.
2. Tingkat kebutuhan bisnis untuk membuat penerapan BI.
3. Jumlah dan kualitas data bisnis yang tersedia.

Dukungan Manajemen

Komitmen dan sponsor dari manajemen puncak menurut Kimball dkk. adalah aspek yang paling penting. Dengan adanya dukungan manajemen yang kuat akan membantu mengatasi kekurangan di bagian lain dalam proyek. Namun, seperti Kimball juga menyatakan bahwa: “bahkan sistem DW / BI yang dirancang paling elegan sekalipun tidak dapat mengatasi kurangnya dukungan [manajemen] perusahaan”.

Personil yang berpartisipasi dalam proyek ini harus memiliki visi dan gagasan tentang manfaat dan kelemahan penerapan sistem BI. Dukungan penuh manajemen harus memiliki pengaruh organisasi dan harus terhubung dengan baik di dalam organisasi. Sangat umum bahwa dukungan manajemen tidak hanya menuntut tetapi juga mampu bersikap realistis dan mendukung jika pelaksanaannya mengalami penundaan atau kekurangan. Sponsor manajemen juga harus mampu menanggung bersama secara pendanaan dan bertanggung jawab atas kegagalan dan kemunduran proyek. Dukungan dari beberapa anggota manajemen memastikan proyek tidak gagal jika satu orang meninggalkan grup pengarah. Namun, memiliki banyak manajer yang bekerja sama dalam proyek juga dapat berarti bahwa ada beberapa kepentingan berbeda yang mencoba menarik proyek ke arah yang berbeda, seperti jika departemen yang berbeda ingin lebih menekankan pada penggunaannya. Masalah ini dapat diatasi dengan analisis awal dan spesifik dari area bisnis yang paling diuntungkan dari penerapannya. Semua pemangku kepentingan dalam

proyek harus berpartisipasi dalam analisis ini agar mereka merasa diinvestasikan dalam proyek dan menemukan kesamaan.

Masalah manajemen lain yang mungkin ditemui sebelum dimulainya implementasi adalah sponsor bisnis yang terlalu agresif. Masalah pengujian sifat mulur terjadi ketika sponsor meminta kumpulan data yang tidak ditentukan dalam fase perencanaan awal.

Kebutuhan Bisnis

Masih terkait dengan dukungan manajemen, aspek penting lain yang harus dinilai sebelum proyek dimulai adalah ada tidaknya kebutuhan bisnis dan apakah ada keuntungan bisnis yang jelas dengan melakukan penerapan BI. Kebutuhan dan manfaat implementasi terkadang didorong oleh persaingan dan kebutuhan untuk mendapatkan keuntungan di pasar. Alasan lain untuk pendekatan berbasis bisnis untuk implementasi BI adalah akuisisi organisasi lain yang memperbesar organisasi aslinya. Terkadang penerapan DW atau BI dapat bermanfaat untuk menciptakan pengawasan yang lebih baik.

Perusahaan yang menerapkan BI seringkali merupakan organisasi multinasional yang besar dengan anak perusahaan yang beragam. Solusi BI yang dirancang dengan baik memberikan pandangan terkonsolidasi dari data bisnis utama yang tidak tersedia di tempat lain dalam organisasi, memberikan visibilitas dan kontrol manajemen atas tindakan yang tidak akan ada.

Jumlah dan Kualitas Data yang Tersedia

Tanpa data yang tepat, atau dengan kualitas data yang terlalu sedikit, implementasi BI akan mengalami kegagalan, tidak peduli seberapa besar dukungan manajeman atau motivasi yang digerakkan oleh perusahaan. Sebelum BI diterapkan, ada baiknya untuk melakukan pembuatan profil data. Analisis ini mengidentifikasi "konten, konsistensi, dan struktur [..]" dari data. Ini harus dilakukan sedini mungkin dalam proses dan jika analisis menunjukkan bahwa data kurang, tunda proyek sementara sementara departemen TI mencari cara untuk mengumpulkan data dengan benar. Saat merencanakan data bisnis dan persyaratan kecerdasan bisnis, selalu disarankan untuk mempertimbangkan skenario tertentu yang berlaku untuk organisasi tertentu, lalu pilih fitur kecerdasan bisnis yang paling cocok untuk skenario tersebut.

Seringkali, skenario berkisar pada proses bisnis yang berbeda, masing-masing dibangun di atas satu atau beberapa sumber data. Sumber-sumber ini digunakan oleh fitur-fitur yang

menyajikan data tersebut sebagai informasi kepada pekerja intelektual, yang kemudian bertindak atas informasi tersebut. Kebutuhan bisnis organisasi untuk setiap proses bisnis yang diadopsi sesuai dengan langkah-langkah penting intelijen bisnis. Langkah-langkah penting ini mencakup:

1. Memeriksa sumber data bisnis untuk mengumpulkan data yang dibutuhkan
2. Mengubah data bisnis menjadi informasi dan menyajikannya dengan tepat
3. Menggali dan menganalisis data
4. Menindak lanjuti data yang dikumpulkan

Aspek kualitas dalam intelijen bisnis harus mencakup semua proses dari sumber data ke pelaporan akhir. Pada setiap langkah, *quality gate* akan sangat berbeda, misalnya:

1. Sumber Data
    - Standardisasi Data: membuat data dapat dibandingkan (unit yang sama, pola yang sama ...)
    - Manajemen Data Master (*Master Data Management*/MDM): referensial unik
2. Penyimpanan Data Operasional (*Operational Data Store*/ODS)
    - Pembersihan Data (*Data Cleansing*): mendeteksi & memperbaiki data yang tidak akurat
    - *Data Profiling*: Pemeriksaan terhadap nilai-nilai yang tidak sesuai, nol / kosong
3. Gudang Data (*Data Warehouse*)
    1. Kelengkapan: memeriksa apakah semua data yang diharapkan dapat termuat
    2. Integritas referensial: referensi unik dan yang ada atas semua sumber
    3. Konsistensi antar sumber: memeriksa data gabungan vs sumber data
    4. Pelaporan
        - Keunikan indikator: hanya satu kamus rujukan sebagai menerjemahkan indikator
        - Akurasi formula: formula pelaporan lokal harus dihindari dan diperiksa

Aspek Pengguna

Beberapa pertimbangan harus dibuat agar berhasil mengintegrasikan penggunaan sistem intelijen bisnis di perusahaan. Pada akhirnya sistem BI harus diterima dan dimanfaatkan oleh pengguna agar dapat menambah nilai bagi organisasi. Jika kegunaan sistem buruk, pengguna mungkin menjadi frustrasi dan menghabiskan banyak waktu untuk mencari tahu bagaimana menggunakan sistem atau mungkin tidak dapat benar-benar menggunakan sistem. Jika sistem tidak menambah nilai pada misi pengguna, mereka tidak akan menggunakannya.

Untuk meningkatkan penerimaan pengguna atas sistem BI, disarankan untuk berkonsultasi dengan pengguna bisnis pada tahap awal siklus hidup DW / BI, misalnya pada tahap pengumpulan persyaratan. Ini dapat memberikan wawasan tentang proses bisnis dan apa yang dibutuhkan pengguna dari sistem BI. Ada beberapa metode untuk mengumpulkan informasi ini, seperti kuesioner dan sesi wawancara.

Saat mengumpulkan persyaratan dari pengguna bisnis, departemen TI lokal juga harus dikonsultasikan untuk menentukan sejauh mana mungkin untuk memenuhi kebutuhan bisnis berdasarkan data yang tersedia. Mengambil pendekatan yang berpusat pada pengguna selama tahap desain dan pengembangan selanjutnya dapat meningkatkan kemungkinan adopsi pengguna yang cepat dari sistem BI. Selain berfokus pada pengalaman pengguna yang ditawarkan oleh aplikasi BI, hal ini juga dapat memotivasi pengguna untuk memanfaatkan sistem dengan menambahkan elemen persaingan. Kimball menyarankan untuk menerapkan fungsi di situs portal Business Intelligence tempat laporan penggunaan sistem dapat ditemukan. Dengan demikian, manajer dapat melihat seberapa baik kinerja departemen mereka dan membandingkan diri mereka dengan yang lain dan ini dapat memacu mereka untuk mendorong staf mereka untuk lebih memanfaatkan sistem BI.

Dalam artikel tahun 2007, H. J. Watson memberikan contoh bagaimana elemen kompetitif dapat bertindak sebagai insentif. Watson menjelaskan bagaimana pusat panggilan besar menerapkan dasbor kinerja untuk semua agen panggilan, dengan bonus insentif bulanan yang terkait dengan metrik kinerja. Selain itu, agen dapat membandingkan kinerja mereka dengan anggota tim lainnya. Penerapan jenis pengukuran kinerja dan persaingan ini secara signifikan meningkatkan kinerja agen.

Peluang keberhasilan BI dapat ditingkatkan dengan melibatkan manajemen senior untuk membantu menjadikan BI sebagai bagian dari budaya organisasi, dan dengan menyediakan alat, pelatihan, dan dukungan yang diperlukan kepada pengguna. Pelatihan mendorong lebih banyak orang untuk menggunakan aplikasi BI.

Memberikan dukungan pengguna diperlukan untuk memelihara sistem BI dan menyelesaikan masalah pengguna. Dukungan pengguna dapat digabungkan dengan banyak cara, misalnya dengan membuat situs web. Situs web harus berisi konten dan alat hebat untuk menemukan informasi yang diperlukan. Selain itu, dukungan helpdesk dapat digunakan. Help desk dapat diawaki oleh power user atau tim proyek DW / BI.

1.3 Portal BI

Portal Intelejensi Bisnis (BI portal) adalah antarmuka akses utama untuk Data Warehouse (DW) dan aplikasi Intelejensi Bisnis. Portal BI adalah kesan pertama pengguna dari DW / Sistem BI. Ini biasanya merupakan aplikasi browser, dari mana pengguna memiliki akses ke semua individu jasa dari itu DW / BI sistem, laporan dan fungsi analitis lainnya. Itu Portal BI harus menjadi diimplementasikan di misalnya seperti itu Itu mudah untuk para pengguna dari DW / BI aplikasi untuk panggil itu Kegunaan aplikasi.

Fungsi utama portal BI adalah untuk menyediakan sistem navigasi aplikasi DW / BI. Ini berarti bahwa portal harus diimplementasikan sedemikian rupa sehingga pengguna memiliki akses ke semua fungsi aplikasi DW / BI. Cara paling umum untuk mendesain portal adalah menyesuaikannya dengan proses bisnis organisasi untuk yang itu DW / BI aplikasi dirancang, karena cara portal bisa terbaik cocok dengan kebutuhan dan persyaratan penggunanya. Portal BI harus mudah digunakan dan dipahami, dan jika mungkin memiliki tampilan dan nuansa yang serupa aplikasi atau konten web organisasi yang dirancang untuk aplikasi DW / BI (konsistensi). Berikut ini adalah daftar fitur yang diperlukan bagi portal web secara umum dan portal BI secara khusus :

1. Terpakai
   Pengguna harus dengan mudah menemukan apa yang mereka butuhkan dalam alat IB.
2. Kaya isi/konten
   Portal tidak hanya alat pencetakan laporan, ia harus berisi fungsi lebih seperti saran, bantuan, informasi pendukung dan dokumentasi.
3. Bersih
   Portal harus dirancang supaya mudah dipahami dan tidak terlalu kompleks sehingga membingungkan pengguna.
4. Terbaru
   Portal harus diperbarui secara teratur.
5. Interaktif
   Portal harus diimplementasikan supaya mudah bagi pengguna menggunakan fungsinya dan mendorong mereka menggunakan portal. Skalabilitas dan kostumisasi membuat pengguna dapat menyesuaikan portal sesuai kebutuhan mereka.
6. Berorientasi nilai
   Sangat penting bahwa pengguna merasakan bahwa aplikasi GD/IB memiliki sumber nilai yang patut dipakai.

Marketplace

Ada sejumlah vendor business intelligence, sering dikategorikan ke dalam independen "murni-play" vendor tersisa dan konsolidasi "megavendors" yang telah memasuki pasar melalui tren terbaru Akuisisi di industri BI.

Beberapa perusahaan mengadopsi software BI memutuskan untuk memilih dari penawaran produk yang berbeda (best-of-breed) daripada membeli satu solusi terintegrasi yang komprehensif (layanan penuh).

Industri Khusus

Pertimbangan khusus untuk sistem intelijen bisnis harus diambil di beberapa sektor seperti peraturan perbankan pemerintah atau perawatan kesehatan. Informasi yang dikumpulkan oleh lembaga perbankan dan dianalisis dengan software BI harus dilindungi dari beberapa kelompok atau individu, sementara tersedia sepenuhnya untuk kelompok atau individu lain. Oleh karena itu, solusi BI harus peka terhadap kebutuhan tersebut dan cukup fleksibel untuk beradaptasi dengan peraturan baru dan perubahan hukum yang ada.

Data Semi-Terstruktur atau Tidak Terstruktur

Bisnis membuat sejumlah besar informasi berharga dalam bentuk email, memo, catatan dari pusat panggilan, berita, grup pengguna, obrolan, laporan, halaman web, presentasi, file gambar, file video, dan materi pemasaran dan berita. Menurut Merrill Lynch, lebih dari 85% dari semua informasi bisnis ada dalam formulir ini. Jenis informasi ini disebut data semi-terstruktur atau tidak terstruktur. Namun, organisasi seringkali hanya menggunakan dokumen ini satu kali.

Pengelolaan data semi-terstruktur diakui sebagai masalah utama yang belum terpecahkan dalam industri teknologi informasi. Menurut proyeksi dari Gartner (2003), pekerja kerah putih menghabiskan 30 sampai 40 persen dari waktu mereka untuk mencari, menemukan dan menilai data yang tidak terstruktur. BI menggunakan data terstruktur dan tidak terstruktur, tetapi yang pertama mudah dicari, dan yang terakhir berisi sejumlah besar informasi yang diperlukan untuk analisis dan pengambilan keputusan. Karena kesulitan dalam mencari, menemukan, dan menilai data tidak terstruktur atau semi-terstruktur dengan benar, organisasi tidak boleh memanfaatkan sumber informasi yang sangat besar ini, yang dapat memengaruhi keputusan, tugas, atau proyek tertentu. Hal ini pada akhirnya dapat menyebabkan pengambilan keputusan yang kurang tepat.

Oleh karena itu, ketika merancang solusi kecerdasan bisnis / DW, masalah spesifik yang terkait dengan data semi-terstruktur dan tidak terstruktur harus diakomodasi sebaik mungkin seperti data terstruktur.

Data Tidak Terstruktur vs. Semi-Terstruktur

Data tidak terstruktur dan semi-terstruktur memiliki arti yang berbeda bergantung pada konteksnya. Dalam konteks sistem basis data relasional, data yang tidak terstruktur tidak dapat disimpan dalam kolom dan baris yang diurutkan sesuai prediksi. Satu jenis data tidak terstruktur biasanya disimpan dalam *binary large object* / BLOB (kumpulan data biner), yaitu aplikasi yang digunakan terutama untuk menampung objek multimedia seperti gambar, video, dan suara, meskipun mereka juga dapat digunakan untuk menyimpan program atau bahkan potongan kode. Data tidak terstruktur juga dapat mengacu pada pola kolom yang berulang secara tidak teratur atau acak yang bervariasi dari baris ke baris dalam setiap file atau dokumen.

Banyak dari tipe data ini, seperti email, file teks pengolah kata, PPT, file gambar, dan file video sesuai dengan standar yang menganjurkan kemungkinan penggunaan metadata. Metadata dapat menyertakan informasi seperti penulis dan waktu pembuatan, dan ini dapat disimpan dalam database relasional. Oleh karena itu, mungkin lebih akurat untuk membicarakan hal ini sebagai dokumen atau data semi-terstruktur, tetapi tampaknya tidak ada konsensus khusus yang telah dicapai. Data tidak terstruktur juga bisa menjadi pengetahuan yang dimiliki pengguna bisnis tentang tren bisnis masa depan. Perkiraan bisnis secara alami sejalan dengan sistem BI karena pengguna bisnis memikirkan bisnis mereka secara agregat. Menangkap pengetahuan bisnis yang mungkin hanya ada di benak pengguna bisnis memberikan beberapa poin data terpenting untuk solusi BI yang lengkap.

Masalah dalam Data Semi-Terstruktur atau Tidak Terstruktur

Ada beberapa tantangan dalam mengembangkan BI dengan data semi terstruktur. Menurut Inmon & Nesavich, beberapa di antaranya adalah:

1. Secara fisik mengakses data tekstual tak-terstruktur - data tak terstruktur disimpan dalam berbagai format.
2. Terminologi - Di antara peneliti dan analis, ada kebutuhan untuk mengembangkan termilogi yang standar.

3. Volume data - Sebagaimana yang dinyatakan sebelumnya, sampai 85% dari semua data yang ada adalah semi-terstruktur. Gabungkan hal tersebut dengan kebutuhan untuk analisis semantik dan kata-per-kata.
4. Pencarian dari data tekstual tak-terstruktur - Pencarian sederhana pada beberapa data, misalnya apel, menghasilkan tautan yang memiliki acuan terhadap istilah yang dicari. Sebagai contoh: "suatu pencarian dilakukan untuk istilah tindak pidana. Dalam pencarian sederhana, istilah tindak pidana digunakan, dan di mana pun ada suatu acuan ke kata tindak pidana, sampai pada dokumen tak terstruktur. Tapi pencarian yang sederhana adalah kasar. Ia tidak menemukan referensi ke kriminal, aksi pembakaran, pembunuhan, penggelapan, kematian karena tabrakan, dan lainnya, walaupun jenis kejahatan ini adalah tipe dari tindak pidana."

Penggunaan Metadata

Untuk menangani masalah pencarian dan penilaian dari data, sangat diperlukan untuk mengetahui tentang isinya. Hal ini bisa dilakukan dengan menambahkan konteks lewat penggunaan metadata. Banyak sistem telah menggunakan metadata (misalnya, nama berkas, penulis, ukuran, dll), tetapi yang lebih berguna tentu metadata tentang apa yang ada dalam isi—misalnya, kesimpulan, topik, orang atau perusahaan yang disebutkan. Dua teknologi dirancang untuk menghasilkan metadata tentang yaitu kategorisasi otomatis dan ekstraksi informasi.

Tren Masa Depan

Sebuah makalah 2009 memprediksi perkembangan ini di pasar intelijen bisnis:

- Karena kurangnya informasi, proses, dan alat, hingga 2012, lebih dari 35 persen perusahaan global teratas secara rutin gagal membuat keputusan mendalam tentang signifikansi perubahan dalam bisnis dan pasar mereka.
- Pada 2012, unit bisnis akan mengendalikan setidaknya 40 persen dari total anggaran untuk intelijen bisnis.
- Pada 2012, sepertiga dari aplikasi analisis yang diterapkan pada proses bisnis akan dikirimkan melalui mashup aplikasi berbutir kasar.
- BI memiliki ruang lingkup yang luas dalam Kewirausahaan namun sebagian besar pengusaha baru mengabaikannya potensi.

Laporan khusus Information Management tahun 2009 memprediksi tren teratas dari IB: "komputasi hijau, jasa jaringan sosial, visualisasi data, IB seluler, analitis prediktif, aplikasi komposit, komputasi awan dan multi-sentuh. "Penelitian yang dilakukan tahun 2014

mengindikasikan bahwa karyawan lebih mungkin memiliki akses ke, dan lebih mungkin lagi terlibat dengan, perangkat IB berbasis-*cloud* daripada perangkat tradisional.

Tren intelijen bisnis lainnya meliputi:

- Produk SOA-IB pihak ketiga yang menangani masalah ETL yang besar.
- Perusahaan menerapkan pemrosesan dalam memory, prosesor 64-bit, dan pra-paket aplikasi IB analitis.
- Aplikasi operasional memiliki komponen IB, dengan peningkatan pada waktu respon, skala, dan konkurensi.
- Analitis IB yang tepat atau mendekati seketika adalah ekpektasi dasar.
- Perangkat lunak sumber-berbuka IB menggantikan penawaran dari vendor.

Jalur penelitian lain termasuk studi gabungan intelijen bisnis dan data yang tidak pasti. Di konteks ini, data yang digunakan tidak dianggap tepat, akurat dan lengkap. Sebaliknya, data dianggap tidak pasti dan karenanya ketidakpastian ini disebarkan ke hasil yang dihasilkan oleh BI.

Menurut kajian dari Aberdeen Group, ada peningkatan ketertarikan yang mencapai dua kali lipat oleg beberapa organisasi atau perusahaan dalam penerpan dan penggunaan *Software-as-a-Service* (SaaS) selama beberapa tahun terakhir, dimana setahun lalu - 15% pada tahun 2009 dibandingkan 7% pada tahun 2008.

Sebuah artikel oleh InfoWorld, Chris Kanaracus, menunjukkan data pertumbuhan serupa dari perusahaan riset IDC, yang memperkirakan pasar SaaS BI akan tumbuh 22 persen setiap tahun hingga 2013 berkat peningkatan kecanggihan produk, anggaran TI yang tegang, dan faktor-faktor lainnya.

1.4 Intelejensi Bisnis Seluler

Mobile business intelligence atau dikenal dengan *Mobile BI atau Mobile Intelligence* adalah distribusi dari data bisnis ke sebuah alat mobile seperti ponsel pintar (smartphone). Bisa juga dianggap sebagai sistem BI yang tertanam pada sebuah alat *mobile*.

“Mobile BI adalah salah satu bagian dari teka-teki BI. Jika BI ingin membuat keputusan yang lebih baik dengan menggunakan data yang benar, maka Mobile BI akan memastikan bahwa setiap orang – terutama pekerja jarak jauh – memiliki akses ke data tersebut kapan saja, di mana saja.

Meskipun konsep komputasi seluler telah lazim selama lebih dari satu dekade, Mobile BI telah menunjukkan momentum / pertumbuhan baru-baru ini. Perubahan ini antara lain didorong oleh perubahan dari 'dunia kabel' ke dunia nirkabel dengan keunggulan smartphone yang membawa era baru komputasi bergerak, khususnya di bidang BI.

Menurut Aberdeen Group, sejumlah besar perusahaan dengan cepat menjalankan Mobile BI karena sejumlah besar tekanan pasar seperti kebutuhan akan efisiensi yang lebih tinggi dalam proses bisnis, peningkatan produktivitas karyawan (misalnya, waktu yang dihabiskan untuk mencari informasi), lebih baik dan pengambilan keputusan yang lebih cepat, layanan pelanggan yang lebih baik, dan pengiriman akses data dua arah secara real-time untuk membuat keputusan kapan saja dan di mana saja. Namun terlepas dari keuntungan nyata dari pengiriman informasi seluler, Mobile BI masih dalam fase 'prototipe awal'. Beberapa CFO tetap skeptis terhadap manfaat bisnis dan dengan kurangnya kasus penggunaan bisnis tertentu dan ROI yang nyata, adopsi Mobile BI masih tertinggal dibandingkan dengan aplikasi seluler perusahaan lainnya.

Sejarah

Pengiriman Informasi ke Perangkat Seluler

Metode utama untuk mengakses informasi BI adalah menggunakan perangkat lunak berpemilik atau browser Web pada komputer pribadi untuk menyambung ke aplikasi BI. Aplikasi BI ini meminta data dari database. Mulai akhir 1990-an, sistem BI menawarkan alternatif untuk menerima data, termasuk email dan perangkat seluler.

Menggunakan Data Statis

Awalnya, perangkat seluler seperti pager dan ponsel menerima data yang didorong menggunakan layanan pesan singkat (SMS) atau pesan teks. Aplikasi ini dirancang untuk perangkat seluler tertentu, berisi informasi dalam jumlah minimal, dan tidak menyediakan interaktivitas data. Akibatnya, perancangan dan pemelihatraan aplikasi Mobile BI terdahulu dinilia mahal dan hanya memberikan nilai informasi yang terbatas, sehingga sepi peminat.

Akses Data Melalui Browser Seluler

Peramban seluler pada ponsel cerdas, komputer genggam yang terintegrasi dengan ponsel, menyediakan sarana untuk membaca tabel data sederhana. Ruang layar yang kecil, browser seluler yang belum matang, dan transmisi data yang lambat tidak dapat

memberikan pengalaman BI yang memuaskan. Aksesibilitas dan bandwidth dapat dianggap sebagai masalah dalam hal teknologi seluler, tetapi solusi BI menyediakan fungsionalitas lanjutan untuk memprediksi dan mengungguli tantangan potensial tersebut. Meskipun solusi Mobile BI berbasis web memberikan sedikit atau bahkan tidak ada kendali atas pemrosesan data dalam jaringan, solusi BI terkelola untuk perangkat seluler hanya menggunakan server untuk operasi tertentu. Selain itu, laporan lokal dikompresi selama transmisi dan pada perangkat, memungkinkan fleksibilitas yang lebih besar untuk penyimpanan dan penerimaan laporan ini. Dalam lingkungan seluler, pengguna memanfaatkan akses mudah ke informasi karena aplikasi seluler beroperasi dalam lingkungan otorisasi tunggal yang memungkinkan akses ke semua konten BI (dengan menghormati keamanan yang ada) terlepas dari bahasa atau lokalnya. Selain itu, pengguna tidak perlu membuat dan memelihara penyebaran Mobile BI yang terpisah. Selain itu, Mobile BI membutuhkan lebih sedikit bandwidth untuk fungsionalitas. Mobile BI menjanjikan footprint laporan kecil pada memori, enkripsi selama transmisi serta pada perangkat, dan penyimpanan data terkompresi untuk dilihat dan digunakan secara offline..

Mobile Client Application (MoCA)

Pada tahun 2002, *Research in Motion* merilis smartphone *BlackBerry* pertama yang dioptimalkan untuk penggunaan email nirkabel. Email nirkabel terbukti menjadi "aplikasi pembunuh" yang mempercepat popularitas pasar ponsel cerdas. Pada pertengahan 2000-an, *Research in Motion's BlackBerry* telah memperkuat cengkeramannya di pasar ponsel pintar dengan organisasi perusahaan dan pemerintah. Ponsel cerdas *BlackBerry* menghilangkan hambatan terhadap kecerdasan bisnis seluler. *BlackBerry* menawarkan perlakuan data yang konsisten di banyak modelnya, menyediakan layar yang jauh lebih besar untuk melihat data, dan memungkinkan interaktivitas pengguna melalui *thumbwheel* dan *keyboard*. Vendor BI memasuki pasar kembali dengan penawaran yang mencakup berbagai sistem operasi seluler (*BlackBerry, Windows, Symbian*) dan metode akses data. Dua opsi akses data paling populer adalah:

- untuk menggunakan *browse*r seluler untuk mengakses data, mirip dengan komputer *desktop*, dan
- untuk membuat aplikasi asli yang dirancang khusus untuk perangkat seluler.

Research in Motion terus kehilangan pangsa pasar ke smartphone Apple dan Android. Dalam tiga bulan pertama tahun 2011 *OS Android Google* memperoleh 7 poin pangsa pasar. Selama periode waktu yang sama, pangsa pasar RIM jatuh dan turun hampir 5 poin.

Aplikasi Mobile BI yang Dirancang Secara khusus

Apple dengan cepat menetapkan standar untuk perangkat seluler dengan diperkenalkannya iPhone. Dalam tiga tahun pertama, Apple menjual lebih dari 33,75 juta unit. Demikian pula, pada 2010, Apple menjual lebih dari 1 juta iPad hanya dalam waktu kurang dari tiga bulan. Kedua perangkat memiliki tampilan layar sentuh interaktif yang merupakan standar de facto pada banyak ponsel dan komputer tablet.

Pada tahun 2008, Apple menerbitkan SDK di mana pengembang dapat membangun aplikasi yang berjalan secara native di iPhone dan iPad daripada aplikasi berbasis Safari. Aplikasi asli ini dapat memberikan pengguna pengalaman yang kuat, lebih mudah dibaca dan lebih mudah dinavigasi. Yang lain dengan cepat bergabung dalam keberhasilan unduhan perangkat seluler dan aplikasi. *Google Play Store* sekarang memiliki lebih dari 700.000 aplikasi yang tersedia untuk perangkat seluler yang menjalankan sistem operasi Android.

Lebih penting lagi, kemunculan perangkat seluler telah secara radikal mengubah cara orang menggunakan data di perangkat seluler mereka. Ini termasuk Mobile BI. Aplikasi kecerdasan bisnis dapat digunakan untuk mengubah laporan dan data menjadi dasbor seluler, dan mengirimkannya secara instan ke perangkat seluler apa pun.

Android Google Inc. telah melampaui *iOS Apple Inc.* di arena pengunduhan aplikasi yang berkembang pesat. Pada kuartal kedua 2011, 44% dari semua aplikasi yang diunduh dari pasar aplikasi di seluruh web adalah untuk perangkat Android sementara 31% untuk perangkat Apple, menurut data baru dari *ABI Research*. Aplikasi yang tersisa adalah untuk berbagai sistem operasi seluler lainnya, termasuk *BlackBerry* dan *Windows Phone 7*. Aplikasi Mobile BI telah berevolusi dari aplikasi klien untuk melihat data menjadi aplikasi yang dibuat untuk tujuan yang dirancang untuk memberikan informasi dan alur kerja yang diperlukan untuk membuat keputusan bisnis dan mengambil tindakan dengan cepat.

Aplikasi Web vs. Aplikasi Khusus Perangkat untuk Mobile BI

Pada awal 2011, ketika pasar perangkat lunak Mobile BI mulai matang dan adopsi mulai tumbuh dengan kecepatan yang signifikan baik di perusahaan kecil maupun besar, sebagian besar vendor mengadopsi strategi aplikasi khusus perangkat yang dibuat khusus (mis. Aplikasi iPhone atau Android, diunduh dari *iTunes* atau *Google Play Store*) atau strategi aplikasi web (berbasis browser, bekerja di sebagian besar perangkat tanpa

memasang aplikasi di perangkat). Perdebatan ini terus berlanjut dan ada keuntungan dan kerugian dari kedua metode tersebut. Salah satu solusi potensial adalah adopsi yang lebih luas dari HTML5 pada perangkat mobile yang akan memberikan aplikasi web banyak karakteristik dari aplikasi khusus sambil tetap memungkinkan mereka untuk bekerja pada banyak perangkat tanpa aplikasi terinstal.

Microsoft telah mengumumkan strategi Mobile BI mereka. Microsoft berencana untuk mendukung aplikasi berbasis browser seperti *Reporting Services* dan *PerformancePoint* di iOS pada paruh pertama tahun 2012 dan aplikasi berbasis sentuh di iOS dan Android pada paruh kedua tahun 2012. Terlepas dari persepsi populer bahwa Microsoft hanya mengakui keberadaannya sendiri , langkah baru-baru ini menunjukkan bahwa perusahaan sadar bahwa itu bukan satu-satunya pemain dalam ekosistem teknologi. Alih-alih mencoba memadamkan persaingan atau menyatakan bahwa perkembangan teknologi baru itu konyol, perusahaan tersebut malah memutuskan untuk membuat teknologinya dapat diakses oleh khalayak yang lebih luas.

Ada banyak perangkat dan platform seluler yang tersedia saat ini. Daftar ini terus berkembang dan begitu juga dengan dukungan platformnya. Ada ratusan model yang tersedia saat ini, dengan berbagai kombinasi perangkat keras dan perangkat lunak. Perusahaan harus memilih perangkat dengan sangat hati-hati. Perangkat target akan memengaruhi desain Mobile BI itu sendiri karena desain untuk smartphone akan berbeda dengan tablet. Ukuran layar, prosesor, memori, dll. Semuanya bervariasi. Program Mobile BI harus memperhitungkan kurangnya standarisasi perangkat dari penyedia dengan terus menguji perangkat untuk aplikasi Mobile BI. Beberapa praktik terbaik selalu dapat diikuti. Misalnya, smartphone adalah calon yang baik untuk operasional Mobile BI. Namun, untuk analisis dan analisis bagaimana-jika, tablet adalah pilihan terbaik. Karenanya, pemilihan atau ketersediaan perangkat memainkan peran besar dalam implementasinya.

Permintaan

Analis Gartner Ted Friedman percaya bahwa model aplikasi tujuan penyampaian BI seluler adalah tentang kepraktisan, penyampaian informasi taktis yang diperlukan untuk membuat keputusan segera - “Nilai terbesar adalah dalam operasional BI adalah informasi dalam konteks aplikasi - bukan dalam mendorong banyak data ke telepon seseorang . ”

Mengakses Internet melalui perangkat seluler seperti smartphone juga dikenal sebagai Internet seluler atau Web seluler. IDC memperkirakan tenaga kerja seluler AS akan

meningkat 73% pada tahun 2011. Morgan Stanley melaporkan bahwa Internet seluler meningkat lebih cepat daripada pendahulunya, Internet desktop, memungkinkan perusahaan untuk menyampaikan pengetahuan kepada tenaga kerja seluler mereka untuk membantu mereka membuat keputusan yang lebih menguntungkan .

Michael Cooney dari Gartner telah mengidentifikasi bawa-teknologimu sendiri-di tempat kerja sebagai norma, tidak terkecuali. Pada 2015 pengiriman tablet media akan mencapai sekitar 50% dari pengiriman laptop dan Windows 8 kemungkinan akan berada di posisi ketiga di belakang Android dan Apple. Hasil akhirnya adalah pangsa Microsoft dari platform klien, baik itu PC, tablet, atau smartphone, kemungkinan akan berkurang hingga 60% dan bisa turun di bawah 50%.

Manfaat dalam Bisnis

Dalam Magic Quadrant for Business Intelligence Platforms terbaru, Gartner memeriksa apakah platform tersebut memungkinkan pengguna untuk "sepenuhnya berinteraksi dengan konten BI yang dikirimkan ke perangkat seluler". Frasa "berinteraksi penuh" adalah kuncinya. Kemampuan untuk mengirim peringatan yang disematkan dalam email atau pesan teks, atau tautan ke konten statis dalam pesan email hampir tidak mewakili kecanggihan dalam analisis seluler. Agar pengguna dapat memanfaatkan Mobile BI, mereka harus dapat menavigasi dasbor dan analisis terpandu dengan nyaman — atau senyaman yang dimungkinkan oleh perangkat seluler, yaitu tempat perangkat dengan layar resolusi tinggi dan antarmuka sentuh (seperti iPhone dan Android) ponsel berbasis) memiliki keunggulan yang jelas, katakanlah, edisi BlackBerry sebelumnya. Sama pentingnya untuk mundur selangkah untuk menentukan tujuan dan pola adopsi Anda. Pengguna bisnis mana yang paling diuntungkan dari analisis seluler — dan apa sebenarnya kebutuhan mereka? Anda tidak memerlukan analisis seluler untuk mengirim beberapa peringatan atau laporan ringkasan ke perangkat genggam mereka — tanpa interaktivitas, Mobile BI tidak dapat dibedakan dari sekadar email atau pesan teks yang informatif.

1.5 Aplikasi Mobile BI

Mirip dengan aplikasi konsumen, yang telah menunjukkan pertumbuhan yang terus meningkat selama beberapa tahun terakhir, permintaan yang konstan kapan pun, di mana pun akses ke BI mengarah ke sejumlah pengembangan aplikasi seluler kustom. Bisnis juga mulai mengadopsi solusi seluler untuk tenaga kerja mereka dan segera menjadi komponen kunci dari proses bisnis inti. Dalam survei Aberdeen yang dilakukan pada Mei

2010, 23% dari perusahaan yang berpartisipasi menunjukkan bahwa mereka sekarang telah memiliki aplikasi atau dasbor Mobile BI, sementara 31% lainnya menunjukkan bahwa mereka berencana untuk menerapkan beberapa bentuk Mobile BI di tahun depan.

Jenis−Jenis Aplikasi Mobile BI

Aplikasi Mobile BI dapat dibagi menjadi :

- ***Mobile Browser App***
  Hampir semua perangkat seluler mengaktifkan aplikasi BI berbasis web, thin client, dan hanya HTML. Namun, aplikasi ini statis dan menyediakan sedikit interaksi data. Data dilihat seperti halnya melalui *browser* dari komputer pribadi. Sedikit upaya tambahan diperlukan untuk menampilkan data, tetapi browser seluler biasanya hanya dapat mendukung sebagian kecil interaktivitas *browser web*.
- ***Custome App***
  Langkah maju dari pendekatan ini adalah membuat setiap (atau semua) laporan dan dasbor dalam format khusus perangkat. Dengan kata lain, berikan informasi khusus tentang ukuran layar, optimalkan penggunaan bidang layar, dan aktifkan kontrol navigasi khusus perangkat. Contohnya termasuk roda jempol atau tombol jempol untuk BlackBerry, panah atas / bawah / kiri / kanan untuk Palm, manipulasi gerakan untuk iPhone. Pendekatan ini membutuhkan lebih banyak usaha daripada sebelumnya tetapi tidak ada perangkat lunak tambahan.
- ***Mobile Client App (MoCA)***
  Yang paling canggih, aplikasi klien menyediakan interaktivitas penuh dengan konten BI yang dilihat di perangkat. Selain itu, pendekatan ini menyediakan caching data secara berkala yang dapat dilihat dan dianalisis bahkan secara offline.

Perusahaan diberbagai bidang usaha, dari ritel hingga organisasi nirlaba mesti menyadari benar pentingnya aplikasi mobile BI yang cocok untuk usaha mereka.

Pengembangan

Mengembangkan aplikasi BI berbasir seluler merupakan sebuah tantangan, terutama yang berkaitan dengan rendering tampilan data dan interaktivitas pengguna. Pengembangan Aplikasi Mobile BI biasanya membutuhkan waktu dan juga pembiayaan yang mahal. Pengembangan ini melakukannya tulisan-tulisan dan bentuk-bentuk peringatan, mereka juga membutuhkan informasi yang disesuaikan untuk bidang pekerjaan mereka yang dapat digunakan untuk berinteraksi dan menganalisis untuk mendapatkan dan menggali informasi lebih dalam.

Aplikasi Mobile BI seringkali merupakan aplikasi yang membutuhkan kode khusus dalam pengoperasiannya yang mendasarinya sistem yang dijalankan. Sebagai contoh, aplikasi iPhone memerlukan pengkodean di Objective-C sementara aplikasi Android membutuhkan coding di Java. Selain fungsionalitas pengguna aplikasi, aplikasi harus diberi kode untuk berfungsi dengan infrastruktur server pendukung yang diperlukan untuk melayani data ke aplikasi Mobile BI. Sementara pelanggan aplikasi kode kustom menawarkan opsi hampir tak terbatas, keahlian pengkodean perangkat lunak dan infrastruktur khusus bisa mahal untuk dikembangkan, dimodifikasi, dan dirawat.

Aplikasi Mobile BI dengan Kode Khusus (*Custome Code*)

Aplikasi Mobile BI sering kali merupakan aplikasi dengan kode khusus khusus untuk sistem operasi seluler yang mendasarinya. Misalnya, aplikasi iPhone memerlukan pembuatan kode *Objective-C* sementara aplikasi Android memerlukan pembuatan kode *Java*. Selain fungsionalitas pengguna aplikasi, aplikasi harus diberi kode untuk bekerja dengan infrastruktur server pendukung yang diperlukan untuk menyajikan data ke aplikasi Mobile BI. Sementara aplikasi berkode khusus menawarkan opsi yang hampir tidak terbatas, keahlian dan infrastruktur pengkodean perangkat lunak khusus membutuhkan biaya yang sangat mahal untuk dikembangkan, dimodifikasi, dan dipelihara.

Aplikasi BI dengan Acuan Tetap (*Fixed-form*)

Data bisnis dapat ditampilkan di klien/server Mobile BI (atau browser web) yang berfungsi sebagai antarmuka pengguna ke platform BI yang ada atau sumber data lain, sehingga tidak perlu sumber data master baru dan infrastruktur server khusus. Opsi ini menawarkan visualisasi data tetap dan dapat dikonfigurasi seperti bagan, tabel, tren, KPI, dan tautan, dan biasanya dapat diterapkan dengan cepat menggunakan sumber data yang ada. Namun, visualisasi data tidak terbatas dan tidak selalu dapat diperluas melebihi apa yang tersedia dari vendor.

Aplikasi Mobile BI Berbasis Grafis

Aplikasi Mobile BI juga dapat dikembangkan menggunakan grafis, lingkungan pengembangan *drag-and-drop* (atau"seret dan lepas") dari platform BI. Keuntungannya antara lain sebagai berikut:

1. Aplikasi dapat dikembangkan tanpa pengkodean,
2. Aplikasi dapat dengan mudah dimodifikasi dan dipelihara menggunakan alat manajemen perubahan platform BI,

3. Aplikasi dapat menggunakan berbagai visualisasi data dan tidak terbatas pada beberapa saja,
4. Aplikasi dapat menggabungkan alur kerja bisnis tertentu, dan
5. Platform BI menyediakan infrastruktur server.

Menggunakan alat pengembangan BI grafis dapat memungkinkan pengembangan aplikasi BI seluler lebih cepat ketika aplikasi kustom diperlukan.

Keamanan Aplikasi Mobile BI

Tingkat adopsi yang tinggi dan ketergantungan pada perangkat seluler membuat komputasi seluler yang aman menjadi perhatian penting. Studi Pasar *Mobile Business Intelligence* menemukan bahwa keamanan adalah masalah nomor satu (63%) bagi organisasi.
Solusi keamanan seluler yang komprehensif harus memberikan keamanan pada hal-hal berikut berikut:

- Perangkat
- Penularan
- Otorisasi, Otentikasi, dan Keamanan Jaringan

Keamanan Perangkat

Seorang analis senior di firma riset Burton Group merekomendasikan cara terbaik untuk memastikan data tidak akan dirusak adalah untuk tidak menyimpannya di perangkat klien (perangkat seluler). Dengan demikian, tidak ada salinan lokal hilang jika perangkat seluler dicuri dan data dapat berada di server di dalam data pusat dengan akses hanya diizinkan melalui jaringan. Sebagian besar produsen smartphone menyediakan serangkaian fitur keamanan lengkap termasuk enkripsi disk penuh, enkripsi email, serta manajemen jarak jauh yang mencakup kemampuan untuk menghapus konten jika perangkat hilang atau dicuri. Selain itu, beberapa perangkat telah memasang perangkat lunak antivirus dan firewall pihak ketiga seperti BlackBerry RIM.

Keamanan Transmisi

Keamanan transmisi mengacu pada tindakan yang dirancang untuk melindungi data dari intersepsi yang tidak sah, analisis lalu lintas, dan penipuan tiruan. Langkah-langkah ini termasuk Secure Sockets Layer (SSL), iSeries Access untuk Windows, dan koneksi jaringan pribadi virtual (VPN). Transmisi data yang aman harus memungkinkan identitas pengirim dan penerima diverifikasi dengan menggunakan sistem kunci bersama cryptographic serta melindungi data untuk dimodifikasi oleh pihak ketiga ketika melintasi

jaringan. Ini dapat dilakukan menggunakan AES atau Triple DES dengan terowongan SSL terenkripsi.

Otorisasi, Otentikasi, dan Keamanan Jaringan

Otorisasi mengacu pada tindakan menentukan hak akses untuk mengontrol akses informasi kepada pengguna. Otentikasi mengacu pada tindakan menetapkan atau mengonfirmasi pengguna sebagai benar atau asli. Keamanan jaringan mengacu pada semua ketentuan dan kebijakan yang diterapkan oleh administrator jaringan untuk mencegah dan memantau akses yang tidak sah, penyalahgunaan, modifikasi, atau penolakan jaringan komputer dan sumber daya yang dapat diakses jaringan. Mobilitas menambah tantangan keamanan yang unik. Karena data diperdagangkan di luar firewall perusahaan menuju wilayah yang tidak diketahui, memastikan bahwa data tersebut ditangani dengan aman adalah yang terpenting. Menuju ini, otentikasi yang tepat dari koneksi pengguna, kontrol akses terpusat (seperti Direktori LDAP), mekanisme transfer data terenkripsi dapat diterapkan.

Peran BI dalam Keamanan Aplikasi Seluler

Untuk memastikan standar keamanan yang tinggi, platform perangkat lunak BI harus memperluas opsi otentikasi dan kontrol kebijakan ke platform seluler. Platform perangkat lunak intelijen bisnis perlu memastikan a kunci aman terenkripsi untuk penyimpanan kredensial. Kontrol administratif atas kebijakan kata sandi harus memungkinkan pembuatan profil keamanan untuk setiap pengguna dan mulus integrasi dengan direktori keamanan terpusat untuk mengurangi administrasi dan pemeliharaan pengguna.

Produk

Sejumlah vendor BI dan vendor perangkat lunak niche menawarkan solusi Mobile BI. Beberapa contoh vendor penyedian aplikasi Mobile BI antara lain:

- CollabMobile
- Cognos
- Cherrywork
- Dimensional Insight
- InetSoft
- Infor
- Information Builders
- MicroStrategy
- QlikView

- Roambi
- SAP
- Tableau Software
- Sisense
- TARGIT Business Intelligence

1.6 Intelejensi Bisnis Real-time

Real-time Business Intelligence (RTBI) adalah proses penyampaian informasi intelijen bisnis (BI) atau informasi tentang [operasi bisnis] saat itu terjadi. *Real-time* berarti mendekati nol latensi dan akses ke informasi kapan pun diperlukan.

Kecepatan sistem pemrosesan saat ini telah memindahkan data warehousing klasik ke dunia real-time. Hasilnya adalah Intelejensi Bisnis *Real-time*. Transaksi bisnis yang terjadi diumpankan ke sistem BI *real-time* yang mempertahankan keadaan perusahaan saat ini. Sistem RTBI tidak hanya mendukung fungsi strategis klasik dari data warehousing untuk memperoleh informasi dan pengetahuan dari aktivitas perusahaan di masa lalu, tetapi juga menyediakan dukungan taktis real-time untuk mendorong tindakan perusahaan yang segera bereaksi terhadap peristiwa yang terjadi. Dengan demikian, ini menggantikan fungsi gudang data klasik dan integrasi aplikasi perusahaan (EAI). Pemrosesan yang digerakkan oleh peristiwa semacam itu adalah prinsip dasar intelijen bisnis *real-time*.

Dalam konteks ini, "*real-time*" berarti rentang dari milidetik hingga beberapa detik (5 detik) setelah peristiwa bisnis terjadi. Sementara BI umumnya menyajikan data historis untuk analisis manual, RTBI membandingkan peristiwa bisnis saat ini dengan pola historis untuk mendeteksi masalah atau peluang secara otomatis. Kemampuan analisis otomatis ini memungkinkan tindakan korektif untuk dimulai dan / atau aturan bisnis disesuaikan untuk mengoptimalkan proses bisnis.

RTBI adalah pendekatan di mana data hingga satu menit dianalisis, baik secara langsung dari sumber Operasional atau memasukkan transaksi bisnis ke dalam gudang data real-time dan sistem IB. RTBI menganalisis data *real-time*.

Intelijen bisnis *real-time* masuk akal untuk beberapa aplikasi tetapi tidak untuk yang lain - fakta yang perlu dipertimbangkan oleh organisasi saat mereka mempertimbangkan investasi dalam alat BI *real-time*. Kunci untuk memutuskan apakah strategi BI *real-time* akan membayar dividen adalah memahami kebutuhan bisnis dan menentukan apakah

pengguna akhir memerlukan akses langsung ke data untuk tujuan analitis, atau jika sesuatu yang kurang dari real-time cukup cepat.

Perkembangan RTBI

Dalam lingkungan yang kompetitif saat ini dengan harapan konsumen yang tinggi, keputusan yang didasarkan pada data terbaru yang tersedia untuk meningkatkan hubungan pelanggan, meningkatkan pendapatan, memaksimalkan efisiensi operasional, dan ya - bahkan menyelamatkan nyawa. Teknologi ini adalah kecerdasan bisnis waktu nyata. Sistem intelijen bisnis waktu nyata memberikan informasi yang diperlukan untuk secara strategis meningkatkan proses perusahaan serta untuk mengambil keuntungan taktis dari peristiwa yang terjadi.

Latensi

Semua sistem intelijen bisnis *real-time* memiliki beberapa latensi, tetapi tujuannya adalah untuk meminimalkan waktu dari peristiwa bisnis yang terjadi hingga tindakan korektif atau pemberitahuan yang dimulai. Analis Richard Hackathorn menjelaskan tiga jenis latensi:

- Latensi data; waktu yang dibutuhkan untuk mengumpulkan dan menyimpan data
- Latensi analisis; waktu yang dibutuhkan untuk menganalisis data dan mengubahnya menjadi informasi yang dapat ditindaklanjuti
- Latensi tindakan; waktu yang dibutuhkan untuk bereaksi terhadap informasi dan mengambil tindakan

Teknologi intelijen bisnis *real-time* dirancang untuk mengurangi ketiga latensi hingga mendekati nol, sedangkan kecerdasan bisnis tradisional hanya berupaya mengurangi latensi data dan tidak mengatasi latensi analisis atau latensi tindakan karena keduanya diatur oleh proses manual.

Beberapa pemerhati dan ahli tehknologi komputerisasi telah memperkenalkan konsep intelejensi bisnis *real-time* yang memberikan sara bahwa informasi harus disampaikan tepat sebelum diperlukan, dan tidak harus *real-time*.

Arsitektur Intelejensi Bisnis *Real-time*

Berbasis Peristiwa

Sistem *Business Intelligence real-time* didorong oleh peristiwa, dan dapat menggunakan teknik Pemrosesan Peristiwa Kompleks, Pemrosesan Aliran Peristiwa, dan Mashup (*hibrid aplikasi web*) untuk memungkinkan peristiwa dianalisis tanpa diubah dan disimpan terlebih dahulu dalam database. Teknik dalam memori ini memiliki keuntungan bahwa tingkat kejadian yang tinggi dapat dipantau, dan karena data tidak harus ditulis ke dalam basis data, latensi data dapat dikurangi menjadi milidetik.

Gudang Data

Pendekatan alternatif untuk arsitektur yang digerakkan oleh peristiwa adalah dengan meningkatkan siklus penyegaran dari gudang data yang ada untuk memperbarui data lebih sering. Sistem gudang data waktu nyata ini dapat mencapai pembaruan data hampir waktu nyata, di mana latensi data biasanya dalam kisaran dari menit hingga jam. Analisis data biasanya masih manual, sehingga total latensi berbeda secara signifikan dari pendekatan arsitektur berbasis peristiwa.

Teknologi Tanpa Server

Inovasi alternatif terbaru untuk arsitektur *even driven "real-time"* dan / atau gudang data "real-time" adalah Teknologi MSSO (*Multiple Source Simple Output*) yang menghilangkan kebutuhan untuk data warehouse dan server perantara karena ia dapat mengakses data langsung langsung dari sumbernya (bahkan dari berbagai sumber yang berbeda). Karena data langsung diakses secara langsung dengan cara yang tidak menggunakan server, ini memberikan potensi data nol-latensi, waktu-nyata dalam arti yang sebenarnya.

Process-aware

Ini terkadang dianggap sebagai bagian dari kecerdasan Operasional dan juga diidentifikasi dengan Pemantauan Aktivitas Bisnis. Ini memungkinkan seluruh proses (transaksi, langkah) untuk dipantau, metrik (latensi, rasio penyelesaian / gagal, dll.) Untuk dilihat, dibandingkan dengan data historis gudang, dan tren secara *real-time*. Implementasi lanjutan memungkinkan deteksi ambang, peringatan dan pemberian umpan balik ke sistem eksekusi proses itu sendiri, sehingga 'menutup loop'.

Pemantauan Kegiatan

Ini memungkinkan seluruh proses (transaksi, langkah-langkah) untuk dipantau, metrik (latensi, rasio penyelesaian / gagal, dll.) Dapat dilihat, dibandingkan dengan data historis yang disimpan, dan tren secara real-time. Implementasi lanjutan memungkinkan deteksi ambang, mengingatkan dan memberikan umpan balik ke sistem eksekusi proses itu sendiri, sehingga 'menutup loop'.

Teknologi yang Mendukung Analisis Real-time

Teknologi yang dapat didukung untuk memungkinkan intelijen bisnis waktu-nyata adalah visualisasi data, federasi data, integrasi informasi perusahaan, integrasi aplikasi perusahaan, dan arsitektur berorientasi layanan. Alat pemrosesan peristiwa yang kompleks dapat digunakan untuk menganalisis aliran data secara real time dan memicu tindakan otomatis atau mengingatkan pekerja akan pola dan tren.

Alat Gudang Data

Alat data warehouse adalah kombinasi produk perangkat keras dan perangkat lunak yang dirancang khusus untuk pemrosesan analitis. Dalam implementasi data warehouse, tugas-tugas yang melibatkan tuning, menambah atau mengedit struktur di sekitar data, migrasi data dari database lain, rekonsiliasi data dilakukan oleh DBA. Tugas lain untuk DBA adalah membuat database berfungsi dengan baik untuk sejumlah besar pengguna. Padahal dengan peralatan data warehouse, itu adalah tanggung jawab vendor atas desain fisik dan penyetelan perangkat lunak sesuai persyaratan perangkat keras. Paket alat data warehouse dilengkapi dengan sistem operasi, penyimpanan, DBMS, perangkat lunak, dan perangkat keras yang diperlukan. Jika diperlukan perangkat gudang data dapat dengan mudah diintegrasikan dengan alat lain.

Teknologi Seluler

Ada vendor yang sangat terbatas untuk menyediakan intelijen bisnis Seluler; MBI terintegrasi dengan arsitektur BI yang ada. MBI adalah paket yang menggunakan aplikasi BI yang ada sehingga orang dapat menggunakan ponsel mereka dan membuat keputusan secara real time.

Area Aplikasi

- Perdagangan algoritma
- Deteksi penipuan

- Pemantauan sistem
- Pemantauan kinerja aplikasi
- Pengelolaan hubungan pelanggan
- Sensing permintaan
- Manajemen harga dan hasil yang dinamis
- Validasi data
- Kecerdasan operasional dan manajemen risiko
- Pembayaran & pemantauan uang tunai
- Pemantauan keamanan data
- Optimalisasi rantai pasokan
- Analisis data jaringan RFID / sensor
- Alur kerja
- Optimalisasi pusat panggilan
- *Enterprise Mashup*s dan *Mashup Dashboards*
- Industri transportasi

Industri transportasi dapat diuntungkan dengan menggunakan analisis real-time. Untuk jaringan contoh kereta api. Bergantung pada hasil yang disediakan oleh analisis real-time, operator dapat membuat keputusan tentang jenis kereta apa yang dapat ia kirim di trek tergantung pada lalu lintas kereta dan komoditas yang dikirim.

Bab 2 : Analisis Komprehensif

Analisis lebih condong kepada sebuah kerangka berpikir sistematis guna menyelesaikan suatu masalah. Namun, dalam praktiknya, analisis juga dipakai dalam metode analisis pencarian data dalam suatu peristiwa. Analisis diterapkan dalam bisnis untuk meningkatkan kinerja mereka. Beberapa aspek yang dijelaskan dalam teks ini adalah analisis perangkat lunak, analisis tersemat, analisis pembelajaran, dan analisis media sosial. Bagian analisis menawarkan fokus wawasan, dengan mengingat masalah subjek yang kompleks.

2.1 Analisis Bisnis

Analisis bisnis adalah proses melihat dan menilai kekayaan data yang sudah dimiliki oleh perusahaan Anda dan menggunakannya untuk membuat keputusan berdasarkan data. Analisis bisnis lebih dari sekadar melihat angka untuk mengetahui apa yang terjadi, tetapi juga berupaya memberikan informasi tentang mengapa sesuatu terjadi, dan menyarankan langkah apa yang sebaiknya diambil. *Business Analytics* (BA) mengacu pada keterampilan, teknologi, praktik untuk eksplorasi berulang berkelanjutan dan investigasi kinerja bisnis masa lalu untuk mendapatkan wawasan dan mendorong perencanaan bisnis. Analisis bisnis berfokus pada pengembangan wawasan baru dan pemahaman kinerja bisnis berdasarkan data dan metode statistik. Sebaliknya, intelijen bisnis secara umum berfokus pada penggunaan serangkaian metrik yang konsisten untuk mengukur kinerja masa lalu dan memandu perencanaan bisnis, yang juga didasarkan pada data dan metode statistik.

Analisis bisnis menggunakan ekstensif analisis statistik, termasuk pemodelan penjelasan dan prediksi, dan manajemen berbasis fakta untuk mendorong pengambilan keputusan. Oleh karena itu terkait erat dengan ilmu manajemen. Analisis Bisnis dapat digunakan sebagai input untuk keputusan manusia atau dapat mendorong keputusan yang sepenuhnya otomatis. Intelejensi Bisnis bertanya, pelaporan, pemrosesan analisis online (OLAP), dan “peringatan.”

Dengan kata lain, permintaan, pelaporan, OLAP, dan alat peringatan dapat menjawab pertanyaan seperti apa yang terjadi, berapa banyak, seberapa sering, di mana masalahnya, dan tindakan apa yang diperlukan. Analisis bisnis dapat menjawab pertanyaan seperti mengapa ini terjadi, bagaimana jika tren ini berlanjut, apa yang akan terjadi selanjutnya (yaitu, memprediksi), apa yang terbaik yang dapat terjadi (yaitu, optimalkan).

Bank, seperti *Capital One*, menggunakan analisis data (atau analisis, seperti juga disebut dalam pengaturan bisnis), untuk membedakan antara pelanggan berdasarkan risiko kredit, penggunaan dan karakteristik lainnya dan kemudian mencocokkan karakteristik pelanggan dengan penawaran produk yang sesuai. Harrah, perusahaan game, menggunakan analisis dalam program loyalitas pelanggannya.

PT Gojek Indonesia kini sudah memiliki Lebih dari 200 ribu pengemudi (driver) yang tersebar di 15 kota dan nantinya akan terus bertambah, beroperasi di sepuluh kota-kota besar, seperti Jakarta, Bandung, Bali, Surabaya, Makasar, Yogyakarta, Medan, Palembang, Semarang, Balikpapan, Solo, Batam, Malang, Manado, Samarinda dan telah menuai prestasi sebagai Juara 1 dalam kompetisi bisnis *Global Entrepreneurship Program Indonesia* (GEPI) di Bali (Yuliantari, 2017). PT. Gojek merupakan salah satu perusahaan jasa transportasi di Indonesia yang saat ini telah mengalami perkembangan menjadi layanan *On demand Service* berbasis teknologi *Mobile Plattform online* yang menyediakan berbagai layanan mulai dari transportasi, logistik, pembayaran, layanan antar makanan dan berbagai layanan *On Demand* lainnya (Weking & Ndala, 2018). *On Demand Service* adalah sebuah pelayanan yang didasarkan atas sesuai dengan pemintaan dan pesanan dari pelanggan atau konsumen

2.2 Jenis-jenis Analisis

Analisis deskriptif

Menggali data dan menggunakan KPI untuk menunjukkan status bisnis Anda saat ini. Misalnya, informasi real-time tentang demografi, minat, atau perilaku pembelian pelanggan. Ini bisa berupa angka penjualan atau keuangan. Bisa juga berupa metrik sosial seperti jumlah suka pada Facebook, Tweet, atau pengikut yang Anda miliki. Analisis deskriptif tidak mencoba membentuk hubungan sebab akibat. Ini hanya berupa angka pasti.

Analisis prediktif

Jenis analisis ini selangkah lebih maju. Analisis ini mencoba memprediksi tindakan di masa mendatang berdasarkan data riwayat tren.

Berikut contohnya:
Gunakan informasi lampau untuk mencari tahu jenis produk yang mungkin diminati pelanggan berdasarkan jumlahnya terakhir kali, dan apakah mereka cenderung membelinya lagi. Jika Anda memiliki anggaran terbatas untuk kampanye pemasaran dan tidak dapat menyediakan diskon bagi semua orang, berdasarkan analisis deskriptif, analisis prediktif

dapat memaparkan informasi tentang pelanggan yang cenderung akan membeli produk Anda.

Analisis perspektif

Bentuk analisis bisnis ini dapat menunjukkan tindakan terbaik dalam situasi tertentu. Sementara analisis deskriptif menunjukkan hal yang telah terjadi, dan analisis prediktif mencoba memprakirakan apa yang akan terjadi selanjutnya, preskriptif menggunakan informasi tersebut untuk memberi potensi solusi berdasarkan situasi yang serupa (data tahun demi tahun, musiman, peluncuran produk). Misalnya, penjualan tiket untuk pertunjukan saat musim liburan menurun dari penjualan tahun kemarin. Analisis preskriptif dapat menyarankan perlunya menurunkan harga atau menambah peningkatan kinerja sebagai tindak lanjutnya.

Domain Dasar dalam Analisis

- Analisis perilaku
- Analisis Kelompok
- Analisis koleksi
- Pemodelan data kontekstual - mendukung penalaran manusia yang terjadi setelah melihat “dasbor eksekutif” atau analisis visual lainnya
- Analisis siber
- Optimalisasi Perusahaan
- Analisis layanan keuangan
- Analisis penipuan
- Analisis pemasaran
- Analisis harga
- Analisis penjualan eceran
- Analisis Risiko & Kredit
- Analisis Rantai Pasokan
- Analisis bakat
- Telekomunikasi
- Analisis transportasi

2.2.1 Sejarah

Analisis telah digunakan dalam bisnis sejak latihan manajemen diterapkan oleh Frederick Winslow Taylor pada akhir abad ke-19. Henry Ford mengukur waktu setiap komponen di jalur perakitan yang baru didirikannya. Tetapi analisis mulai mendapat lebih banyak

perhatian pada akhir 1960-an ketika komputer digunakan dalam sistem pendukung keputusan. Sejak itu, analisis telah berubah dan dibentuk dengan pengembangan sistem perencanaan sumber daya perusahaan (ERP), gudang data, dan sejumlah besar alat dan proses perangkat lunak lainnya. Di tahun-tahun berikutnya, analisis bisnis meledak dengan diperkenalkannya komputer. Perubahan ini telah membawa analisis ke tingkat yang baru dan membuat kemungkinan tidak terbatas. Sejauh analisis telah hadir dalam sejarah, dan apa bidang analitik saat ini, banyak orang tidak akan pernah berpikir bahwa analitik dimulai pada awal 1900-an dengan sang pencetus yaitu Henry Ford sendiri.

Kendala dalam Analisis Bisnis

Analisis bisnis bergantung pada volume data berkualitas tinggi yang memadai. Kesulitan dalam memastikan kualitas data adalah mengintegrasikan dan merekonsiliasi data di berbagai sistem, dan kemudian memutuskan subset data apa yang akan tersedia. Sebelumnya, analitik dianggap sebagai jenis metode setelah fakta untuk memperkirakan perilaku konsumen dengan memeriksa jumlah unit yang terjual pada kuartal terakhir atau tahun lalu. Jenis data warehousing ini membutuhkan lebih banyak ruang penyimpanan daripada kecepatannya.

Sekarang analisis bisnis menjadi alat yang dapat mempengaruhi hasil interaksi pelanggan. Ketika jenis pelanggan tertentu sedang mempertimbangkan pembelian, perusahaan yang mendukung analitik dapat mengubah promosi penjualan untuk menarik konsumen tersebut. Ini berarti ruang penyimpanan untuk semua data itu harus bereaksi sangat cepat untuk menyediakan data yang diperlukan secara *real-time*.

Daya Saing Analisis

Thomas Davenport, profesor teknologi informasi dan manajemen di *Babson College* berpendapat bahwa bisnis dapat mengoptimalkan kemampuan bisnis yang berbeda melalui analitik dan dengan demikian dapat bersaing dengan lebih baik. Dia mengidentifikasi karakteristik organisasi yang cenderung bersaing dalam analisis:

- Satu atau lebih eksekutif senior yang sangat menganjurkan pengambilan keputusan berdasarkan fakta dan, khususnya, analisis
- Penggunaan yang luas tidak hanya statistik deskriptif, tetapi juga pemodelan prediktif dan teknik pengoptimalan yang kompleks
- Penggunaan analitik secara substansial di berbagai fungsi atau proses bisnis
- Gerakan menuju pendekatan tingkat perusahaan untuk mengelola alat analisis, data, dan keterampilan dan kemampuan organisasi

2.3 Analitik

Analitik adalah penemuan, interpretasi, dan komunikasi pola bermakna dalam data. Sangat berharga di area yang kaya dengan informasi yang terekam, analitik mengandalkan aplikasi statistik, pemrograman komputer, dan riset operasi secara bersamaan untuk mengukur kinerja. Analitik sering kali lebih seperti visualisasi data untuk mengkomunikasikan wawasan. Organisasi dapat menerapkan analitik pada data bisnis untuk mendeskripsikan, memprediksi, dan meningkatkan kinerja bisnis. Secara khusus, area dalam analitik mencakup analitik prediktif, analitik preskriptif, manajemen keputusan perusahaan, analitik ritel, pengoptimalan toko dan unit penyimpanan stok, pengoptimalan pemasaran dan pemodelan bauran pemasaran, analitik web, ukuran dan pengoptimalan tenaga penjualan, pemodelan harga dan promosi, prediktif sains, analisis risiko kredit, dan analisis penipuan. Karena analitik dapat memerlukan komputasi ekstensif, algoritme dan perangkat lunak yang digunakan untuk analitik memanfaatkan metode terkini dalam ilmu komputer, statistik, dan matematika.

Analitik vs Analisis

Kebanyakan orang berpendapat bahwa Data analitik dan Data Analisis adalah dua hal yang sama. Dari situ terkadang beberapa orang menggunakannya secara bergantian. Secara teknis ini tidak benar. Sebenarnya ada perbedaan yang jelas antara keduanya. Jadi mari kita bahas perbedaan yang tidak begitu jelas antara istilah analisis dan analitik karena meskipun memiliki kesamaan kata-kata, namun memiliki pengertian berbeda.

Ketika memiliki kumpulan data yang besar dan berisi data dari beragam jenis yang berbeda. Agar menghindari risiko kesalahan atau agar tidak kewalahan dalam memahami data tersebut, kemudian data-data tersebut dipisah-pisah sehingga lebih mudah untuk mencerna potongan-potongan data dan mempelajarinya secara individu dan memeriksa bagaimana saling terkait satu dengan lainnya. Hal ini bisa dikatakan bahwa kita melakukan Analisis pada data yang diperoleh.

Namun satu hal penting yang perlu diingat adalah bahwa ketika melakukan analisis pada hal-hal yang telah terjadi di masa lalu. Misalnya seperti melakukan analisis untuk menjelaskan bagaimana akhir dari pencapaian target penjualan atau bagaimana historis penurunan curah hujan musim panas lalu. Semua ini berarti kita melakukan analisis untuk menjelaskan bagaimana dan atau mengapa sesuatu terjadi.

Analytics umumnya mengacu pada masa depan, alih-alih menjelaskan peristiwa masa lalu. Dengan kata lain adalah mengeksplorasi potensi masa depan. Analytics pada dasarnya adalah penerapan penalaran logis dan komputasi untuk bagian komponen yang diperoleh dalam analisis. Dalam melakukan kegiatan analitik ini, kita mencari pola dalam mengeksplorasi apa yang dapat dilakukan di masa depan. Analitik mempunyai dua jenis : Kualitatif dan Kuantitatif.

Analitik kualitatif biasanya menggunakan intuisi dan pengalaman Anda bersama dengan analisis untuk merencanakan langkah bisnis Anda berikutnya (yang seringnya digabungkan bersamaan dengan teknik analisis kuantitatif dengan cara menerapkan rumus dan algoritma ke angka yang telah Anda kumpulkan dari analisis Anda).

Misalnya, katakanlah seseorang memiliki toko pakaian online yang unggul dalam persaingan dan memiliki pemahaman yang baik tentang apa kebutuhan dan keinginan pelanggan. Pemilik usaha telah melakukan analisis yang sangat rinci dari artikel pakaian wanita dan merasa yakin tentang tren mode mana yang akan diikuti. Pemilik toko online dapat menggunakan intuisi ini untuk memutuskan gaya pakaian mana yang akan mulai dijual. Ini akan menjadi analisis kualitatif tetapi mungkin tidak tahu kapan harus memperkenalkan koleksi baru.

Dalam hal mengandalkan data penjualan sebelumnya dan data pengalaman pengguna, Pemilik dapat memperkirakan pada bulan apa yang terbaik untuk melakukannya dengan dasar perhitungan kuantitatif. Yang mana, analisis kuantitatif melibatkan angka dan perhitungan spesifik. Dalam hal ini, ia melakukan analisis kualitatif untuk menjelaskan bagaimana atau mengapa, serta melakukan analisis kuantitatif dengan data masa lalu untuk menjelaskan bagaimana penjualan menurun musim panas lalu untuk memperbaikinya di masa yang akan datang. Karena analitik dapat memerlukan komputasi ekstensif, algoritme dan perangkat lunak yang digunakan untuk analitik memanfaatkan metode terkini dalam ilmu komputer, statistik, dan matematika.

Contoh Penerapan

Optimasi Pemasaran

Pemasaran telah berkembang dari proses kreatif menjadi proses yang sangat didorong oleh data. Organisasi pemasaran menggunakan analitik untuk menentukan hasil kampanye atau upaya dan untuk memandu keputusan investasi dan penargetan konsumen. Studi demografis, segmentasi pelanggan, analisis konjoin, dan teknik lain memungkinkan

pemasar menggunakan pembelian konsumen, survei, dan data panel dalam jumlah besar untuk memahami dan mengomunikasikan strategi pemasaran.

Analisis web memungkinkan pemasar mengumpulkan informasi tingkat sesi tentang interaksi di situs web menggunakan operasi yang disebut sesiisasi. Google Analytics adalah contoh alat analitik gratis populer yang digunakan pemasar untuk tujuan ini. Interaksi tersebut memberikan sistem informasi analisis web dengan informasi yang diperlukan untuk melacak perujuk, kata kunci pencarian, mengidentifikasi alamat IP, dan melacak aktivitas pengunjung. Dengan informasi ini, pemasar dapat meningkatkan kampanye pemasaran, konten kreatif situs web, dan arsitektur informasi.

Teknik analisis yang sering digunakan dalam pemasaran meliputi pemodelan bauran pemasaran, analisis harga dan promosi, pengoptimalan tenaga penjualan, dan analisis pelanggan misalnya: segmentasi. Analisis web dan pengoptimalan situs web dan kampanye online sekarang sering bekerja sama dengan teknik analisis pemasaran yang lebih tradisional. Fokus pada media digital telah sedikit mengubah perbendaharaan kata sehingga marketing mix modelling sering disebut sebagai model atribusi dalam konteks digital atau *marketing mix modelling*. Alat dan teknik ini mendukung keputusan pemasaran strategis (seperti berapa banyak keseluruhan yang harus dibelanjakan untuk pemasaran, cara mengalokasikan anggaran ke seluruh portofolio merek dan bauran pemasaran) dan dukungan kampanye yang lebih taktis, dalam hal menargetkan pelanggan potensial terbaik dengan pesan optimal dalam media yang paling hemat biaya pada waktu yang ideal.

Analisis Portofolio Perbankan

Dalam manajemen strategi dan pemasaran portofolio digunakan untuk menunjukan sekumpulan produk, proyek, layanan jasa atau merk yang ditawarkan untuk dijual oleh suatu perusahaan. Setiap perusahaan selalu berupaya untuk meraih difersifikasi dan keseimbangan dalam portofolio produk yang ditawarkan. Perusahaan menggunakan metode optimisasi perbedaan makna guna untuk memaksimalkan keuntungan dengan cara menekan resiko.

Setiap perusahaan senantiasa berupaya untuk meraih difersifikasi dan keseimbangan dalam portfolio roduk yang ditawarkan. Keputusan tentang produk (barang atau jasa) merupakan keputusan strategik yang sangat penting karena mempengaruhi eksistensi perusahaan jangka panjang. Dampaknya mempengaruhi setiap fungsi dan tingkatan dalam organisasi bisnis.

Contohnya dalam usaha perbankan. Dalam hal ini, bank atau lembaga pemberi pinjaman memiliki kumpulan rekening dengan berbagai nilai dan risiko. Akun tersebut mungkin berbeda berdasarkan status sosial (kaya, kelas menengah, miskin, dll.) Dari pemiliknya, lokasi geografis, nilai bersihnya, dan banyak faktor lainnya. Pemberi pinjaman harus menyeimbangkan pengembalian pinjaman dengan risiko gagal bayar untuk setiap pinjaman. Pertanyaannya kemudian bagaimana mengevaluasi portofolio secara keseluruhan.

Pinjaman dengan risiko paling kecil mungkin diberikan kepada orang-orang yang sangat kaya, tetapi hanya ada sejumlah orang kaya yang sangat terbatas. Di sisi lain, ada banyak orang miskin yang dapat dipinjamkan, tetapi berisiko lebih besar. Beberapa keseimbangan harus dicapai yang memaksimalkan pengembalian dan meminimalkan risiko. Solusi analitik dapat menggabungkan analisis deret waktu dengan banyak masalah lain untuk membuat keputusan tentang kapan meminjamkan uang ke segmen peminjam yang berbeda ini, atau keputusan tentang suku bunga yang dibebankan kepada anggota segmen portofolio untuk menutupi kerugian di antara anggota di segmen itu.

Analisis Risiko

Model prediktif dalam industri perbankan dikembangkan untuk memberikan kepastian atas skor risiko bagi nasabah individu. Skor kredit dibangun untuk memprediksi perilaku kenakalan individu dan banyak digunakan untuk mengevaluasi kelayakan kredit setiap pemohon. Selanjutnya, analisis risiko dilakukan dalam dunia ilmiah dan industri asuransi. Ini juga banyak digunakan di lembaga keuangan seperti perusahaan Gateway Pembayaran Online untuk menganalisis apakah transaksi itu asli atau penipuan. Untuk tujuan ini mereka menggunakan riwayat transaksi pelanggan. Ini lebih umum digunakan dalam pembelian Kartu Kredit, ketika tiba-tiba terjadi lonjakan volume transaksi pelanggan, pelanggan mendapat panggilan konfirmasi jika transaksi diprakarsai olehnya. Ini membantu dalam mengurangi kerugian karena keadaan seperti itu.

2.3.3 Analisis Digital

Analisis digital adalah serangkaian aktivitas bisnis dan teknis yang menentukan, membuat, mengumpulkan, memverifikasi, atau mengubah data digital menjadi pelaporan, penelitian, analisis, rekomendasi, pengoptimalan, prediksi, dan otomatisasi. Ini juga termasuk SEO (Search Engine Optimization) di mana pencarian kata kunci dilacak dan data tersebut digunakan untuk tujuan pemasaran. Bahkan iklan spanduk dan klik berada di bawah analisis digital. Semua perusahaan pemasaran mengandalkan analitik digital untuk tugas pemasaran digital mereka, di mana MROI (Pengembalian Pemasaran atas Investasi) penting.

Analisis Keamanan

Kesalahan kerentanan dan keamanan perangkat lunak dapat dikurangi jika secure software development process (SSDP) diikuti, seperti memenuhi aspek keamanan pada tahap membangun perangkat lunak. Keamanan perangkat lunak sangat penting dalam perangkat lunak, sebagian besar keamanan perangkat lunak tidak ditangani sejak software development life cycle (SDLC). Sistem perangkat lunak yang aman akan memberikan tingkat kepercayaan yang tinggi dari user, user akan merasa nyaman dan aman ketika berhubungan dengan sistem yang sudah kita bangun.

Analisis keamanan mengacu pada solusi teknologi informasi (TI) yang mengumpulkan dan menganalisis peristiwa keamanan untuk menghadirkan kesadaran situasional dan memungkinkan staf TI untuk memahami dan menganalisis peristiwa yang menimbulkan risiko terbesar. Solusi di bidang ini termasuk

Analisis Kebutuhan Perangkat Lunak

Analisis kebutuhan perangkat lunak (*software requirements analysis*) merupakan aktivitas awal dari siklus hidup pengembangan perangkat lunak. Untuk proyek-proyek perangkat lunak yang besar, analisis kebutuhan dilaksanakan setelah tahap rekayasa sistem/ informasi dan *software project planning*.

Dengan mengadopsi pengertian tentang kebutuhan pada sub bab sebelumnya, analisis kebutuhan dapat diartikan sebagai berikut :

- Proses mempelajari kebutuhan pemakai untuk mendapatkan definisi kebutuhan sistem atau perangkat lunak
- Proses untuk menetapkan fungsi dan unjuk kerja perangkat lunak, menyatakan antarmuka perangkat lunak dengan elemen-elemen sistem lain, dan menentukan kendala yang harus dihadapi perangkat lunak.

Tujuan pelaksanaan analisis kebutuhan adalah :

- Memahami masalah secara menyeluruh (komprehensif) yang ada pada perangkat lunak yang akan dikembang seperti ruang lingkup produk perangkat lunak (*product space*) dan pemakai yang akan menggunakannya,
- Mendefinisikan apa yang harus dikerjakan oleh perangkat lunak untuk memenuhi keinginan pelanggan.

Tantangan

Dalam industri perangkat lunak analitik komersial, penekanan telah muncul pada penyelesaian tantangan menganalisis kumpulan data yang sangat besar dan kompleks, seringkali ketika data tersebut berada dalam keadaan perubahan yang konstan. Kumpulan data semacam itu biasanya disebut sebagai data besar. Jika dulu masalah yang ditimbulkan oleh data besar hanya ditemukan di komunitas ilmiah, saat ini data besar menjadi masalah bagi banyak bisnis yang mengoperasikan sistem transaksional secara online dan, akibatnya, mengumpulkan data dalam jumlah besar dengan cepat.

Analisis tipe data tidak terstruktur merupakan tantangan lain yang mendapatkan perhatian di industri. Data tidak terstruktur berbeda dari data terstruktur karena formatnya sangat bervariasi dan tidak dapat disimpan dalam database relasional tradisional tanpa upaya signifikan pada transformasi data. Sumber data tidak terstruktur, seperti email, konten dokumen pengolah kata, PDF, data geospasial, dll., Dengan cepat menjadi sumber intelijen bisnis yang relevan untuk bisnis, pemerintah, dan universitas. Misalnya, di Inggris, penemuan bahwa satu perusahaan secara ilegal menjual catatan dokter palsu untuk membantu orang menipu pemberi kerja dan perusahaan asuransi, merupakan peluang bagi perusahaan asuransi untuk meningkatkan kewaspadaan analisis data tidak terstruktur mereka. Institut *Global McKinsey* memperkirakan bahwa analisis data besar dapat menghemat sistem perawatan kesehatan Amerika $ 300 miliar per tahun dan sektor publik Eropa € 250 miliar.

Tantangan ini adalah inspirasi saat ini untuk banyak inovasi dalam sistem informasi analitik modern, melahirkan konsep analisis mesin yang relatif baru seperti pemrosesan acara yang kompleks, pencarian dan analisis teks lengkap, dan bahkan ide-ide baru dalam presentasi. Salah satu inovasi tersebut adalah pengenalan arsitektur mirip jaringan dalam analisis mesin, yang memungkinkan peningkatan kecepatan pemrosesan paralel secara masif dengan mendistribusikan beban kerja ke banyak komputer, semuanya dengan akses yang sama ke kumpulan data lengkap.

Analisis semakin banyak digunakan dalam pendidikan, terutama di tingkat distrik dan kantor pemerintah. Namun, kompleksitas ukuran kinerja siswa menghadirkan tantangan ketika pendidik mencoba memahami dan menggunakan analitik untuk melihat pola dalam kinerja siswa, memprediksi kemungkinan kelulusan, meningkatkan peluang keberhasilan siswa, dll. Misalnya, dalam studi yang melibatkan kabupaten yang terkenal dengan penggunaan datanya yang kuat , 48% guru kesulitan mengajukan pertanyaan berdasarkan data, 36%

tidak memahami data yang diberikan, dan 52% salah mengartikan data. Untuk mengatasi hal ini, beberapa alat analitik untuk pendidik mengikuti format data yang dijual bebas (label embedding, dokumentasi tambahan, dan sistem bantuan, dan membuat keputusan paket / tampilan dan konten utama) untuk meningkatkan pemahaman pendidik dan penggunaan analitik ditampilkan.

Satu tantangan lagi yang muncul adalah kebutuhan regulasi yang dinamis. Misalnya, dalam industri perbankan, Basel III dan kebutuhan kecukupan modal di masa depan cenderung membuat bank yang lebih kecil mengadopsi model risiko internal. Dalam kasus seperti itu, komputasi awan dan R sumber terbuka (bahasa pemrograman) dapat membantu bank yang lebih kecil untuk mengadopsi analitik risiko dan mendukung pemantauan tingkat cabang dengan menerapkan analitik prediktif.

Risiko

Risiko utama bagi masyarakat adalah diskriminasi seperti diskriminasi harga atau diskriminasi statistik. Proses analitis juga dapat mengakibatkan hasil diskriminatif yang mungkin melanggar undang-undang anti diskriminasi dan hak-hak sipil. Ada juga risiko bahwa pengembang dapat memperoleh keuntungan dari ide atau pekerjaan yang dilakukan oleh pengguna, seperti contoh ini: Pengguna dapat menulis ide baru di aplikasi pencatatan, yang kemudian dapat dikirim sebagai acara khusus, dan pengembang dapat memperoleh keuntungan dari ide-ide itu. Ini bisa terjadi karena kepemilikan konten biasanya tidak jelas secara hukum. Jika identitas pengguna tidak dilindungi, ada lebih banyak risiko; misalnya, risiko informasi pribadi tentang pengguna dipublikasikan di internet. Secara ekstrem, terdapat risiko bahwa pemerintah dapat mengumpulkan terlalu banyak informasi pribadi, karena pemerintah sekarang memberikan lebih banyak wewenang kepada diri mereka sendiri untuk mengakses informasi warga negara.

2.4 Analisis Perangkat Lunak

Analisis Perangkat Lunak mengacu pada analitik khusus untuk sistem perangkat lunak dan proses pengembangan perangkat lunak terkait. Ini bertujuan untuk mendeskripsikan, memprediksi, dan meningkatkan pengembangan, pemeliharaan, dan pengelolaan sistem perangkat lunak yang kompleks. Metode dan teknik analitik perangkat lunak biasanya bergantung pada pengumpulan, analisis, dan visualisasi informasi yang ditemukan dalam berbagai sumber data dalam lingkup sistem perangkat lunak dan proses pengembangan perangkat lunaknya --- analitik perangkat lunak "mengubahnya menjadi wawasan yang

dapat ditindaklanjuti untuk menginformasikan keputusan yang lebih baik terkait dengan perangkat lunak".

Analisis perangkat lunak merupakan komponen dasar diagnosis perangkat lunak yang umumnya bertujuan untuk menghasilkan temuan, kesimpulan, dan evaluasi tentang sistem perangkat lunak dan implementasinya, komposisi, perilaku, dan evolusinya. Analisis perangkat lunak sering menggunakan dan menggabungkan pendekatan dan teknik dari statistik, analisis prediksi, penggalian data, dan visualisasi ilmiah. Misalnya, analitik perangkat lunak dapat memetakan data melalui peta perangkat lunak yang memungkinkan eksplorasi interaktif.

Data yang sedang dieksplorasi dan dianalisis oleh Analisis Perangkat Lunak ada dalam siklus hidup perangkat lunak, termasuk kode sumber, spesifikasi persyaratan perangkat lunak, laporan bug, kasus pengujian, jejak / log eksekusi, dan umpan balik pengguna di dunia nyata, dll. Data memainkan peran penting dalam pengembangan perangkat lunak modern , karena yang tersembunyi dalam data adalah informasi dan wawasan tentang kualitas perangkat lunak dan layanan, pengalaman yang diterima pengguna perangkat lunak, serta dinamika pengembangan perangkat lunak.

Informasi berwawasan yang diperoleh oleh Software Analytics adalah informasi yang menyampaikan pemahaman atau pengetahuan yang bermakna dan berguna untuk melakukan tugas target. Biasanya informasi berwawasan tidak dapat dengan mudah diperoleh dengan penyelidikan langsung pada data mentah tanpa bantuan teknologi analitik. Informasi yang dapat ditindaklanjuti yang diperoleh oleh Software Analytics adalah informasi di mana praktisi perangkat lunak dapat menghasilkan solusi konkret (lebih baik daripada solusi yang ada jika ada) untuk menyelesaikan tugas target.

Analisis Perangkat Lunak berfokus pada trinitas sistem perangkat lunak, pengguna perangkat lunak, dan proses pengembangan perangkat lunak:

Sistem Perangkat Lunak

Bergantung pada skala dan kompleksitas, spektrum sistem perangkat lunak dapat menjangkau dari sistem operasi untuk perangkat hingga sistem jaringan besar yang terdiri dari ribuan server. Kualitas sistem seperti keandalan, kinerja dan keamanan, dll., Adalah kunci sukses sistem perangkat lunak modern. Seiring dengan peningkatan skala dan kompleksitas sistem, sejumlah besar data, misalnya, jejak dan log run-time,

dihasilkan; dan data menjadi sarana penting untuk memantau, menganalisis, memahami, dan meningkatkan kualitas sistem.

Pengguna Perangkat Lunak

Pengguna (hampir) selalu benar karena pada akhirnya mereka akan menggunakan perangkat lunak dan layanan dengan berbagai cara. Karena itu, penting untuk terus memberikan pengalaman terbaik kepada pengguna. Data penggunaan yang dikumpulkan dari dunia nyata mengungkapkan bagaimana pengguna berinteraksi dengan perangkat lunak dan layanan. Data ini sangat berharga bagi praktisi perangkat lunak untuk lebih memahami pelanggan mereka dan mendapatkan wawasan tentang cara meningkatkan pengalaman pengguna yang sesuai.

Proses Pengembangan Perangkat Lunak

Pengembangan perangkat lunak telah berevolusi dari bentuk tradisional menjadi menunjukkan karakteristik yang berbeda. Prosesnya lebih gesit dan insinyur lebih kolaboratif daripada sebelumnya. Analisis data pengembangan perangkat lunak menyediakan mekanisme yang kuat yang dapat dimanfaatkan oleh para praktisi perangkat lunak untuk mencapai produktivitas pengembangan yang lebih tinggi.
Secara umum, teknologi utama yang digunakan oleh Software Analytics meliputi teknologi analitis seperti pembelajaran mesin, penggalian data dan pengenalan pola, visualisasi informasi, serta komputasi & pemrosesan data skala besar.

Sejarah

Pada Mei 2009, *Software Analytics* pertama kali diciptakan dan diusulkan ketika Dr. Dongmei Zhang mendirikan *Software Analytics Group* (SA) di *Microsoft Research Asia* (MSRA). Istilah ini menjadi terkenal di komunitas penelitian rekayasa perangkat lunak setelah serangkaian tutorial dan pembicaraan tentang analitik perangkat lunak diberikan oleh Dr. Dongmei Zhang dan rekan-rekannya, bekerja sama dengan Profesor Tao Xie dari *North Carolina State University*, pada konferensi rekayasa perangkat lunak termasuk tutorial di Konferensi Internasional IEEE / ACM tentang Rekayasa Perangkat Lunak Otomatis (ASE 2011), ceramah di Lokakarya Internasional tentang Teknologi Pembelajaran Mesin dalam Rekayasa Perangkat Lunak (MALETS 2011), tutorial dan ceramah utama yang diberikan oleh Dr. Dongmei Zhang di I*EEE-CS Conference on Software Engineering Education and Training* (CSEE & T 2012), tutorial di I*nternational Conference on Software Engineering* (ICSE 2012) - Software Engineering in Practice Track, dan keynote talk

yang diberikan oleh Dr. Dongmei Zhang pada *Working Conference on Mining Software Repositories* (MSR 2012).

Pada November 2010, Analisis Pengembangan Perangkat Lunak (Analisis Perangkat Lunak dengan fokus pada Pengembangan Perangkat Lunak) diusulkan oleh Thomas Zimmermann dan rekan-rekannya di *Empirical Software Engineering Group* (ESE) di *Microsoft Research Redmond* dalam makalah FoSER 2010 mereka. Panel mangkuk ikan mas pada analitik pengembangan perangkat lunak diselenggarakan oleh Thomas Zimmermann dan Profesor Tim Menzies dari *West Virginia University* pada Konferensi Internasional tentang Rekayasa Perangkat Lunak (ICSE 2012), Rekayasa Perangkat Lunak dalam jalur Praktik.

Penyedia Analisis Perangkat Lunak

- CAST Softwar
- IBM Cognos Business Intelligence
- Kiuwan
- Microsoft Azure Application Insights
- Nalpeiron Software Analytics
- New Relic
- Squore
- Tableau Software
- Trackerbird Software Analytics

2.5 Analisis Tersemat

Analisis Tersemat (*Embedded analytics*) adalah teknologi yang dirancang untuk membuat analisis data dan intelijen bisnis lebih mudah diakses oleh semua jenis aplikasi atau pengguna.

Definisi

Menurut analis Gartner Kurt Schlegel, intelijen bisnis mengalami "kesengsaraan" pada tahun 2008 karena kurangnya integrasi antara data dan pengguna bisnis. Tujuan teknologi ini adalah untuk menjadi lebih luas dengan otonomi *real-time* dan layanan mandiri visualisasi atau kustomisasi data, sementara itu pembuat keputusan, pengguna bisnis atau bahkan pelanggan melakukan alur kerja dan tugas harian mereka sendiri.

Sejarah

Konsep pertama kali disebutkan oleh Howard Dresner, konsultan, penulis, mantan analis Gartner dan penemu istilah "intelijen bisnis". Konsolidasi kecerdasan bisnis "tidak berarti pasar BI telah mencapai kematangan" kata Howard Dresner ketika dia bekerja untuk *Hyperion Solutions*, sebuah perusahaan yang dibeli Oracle pada tahun 2007. Kemudian *Oracle* mulai menggunakan istilah "analitik tertanam" pada siaran pers mereka untuk *Oracle® Rapid Planning* tahun 2009. Gartner Group, sebuah perusahaan tempat Howard Dresner bekerja, akhirnya menambahkan istilah tersebut ke Glosarium IT mereka pada tanggal 5 November 2012. Jelas bahwa ini adalah teknologi arus utama ketika *Dresner Advisory Services* menerbitkan Studi Pasar Kecerdasan Bisnis Tersemat 2014 sebagai bagian dari Seri Riset *Wisdom of Crowds®*, termasuk 24 vendor.

Aplikasi dan Tools yang bisa digunakan :

- Actuate
- Dundas Data Visualization
- GoodData
- IBM
- icCube
- Logi Analytics
- Pentaho
- Qlik
- SAP
- SAS
- Tableau
- TIBCO
- Sisense

2.6 Analisis Pembelajaran

Analisis pembelajaran (*Learning Analytics*/LA) merupakan salah satu metode yang dapat digunakan untuk mengolah data yang diperoleh dalam pembelajaran. LA ini erat kaitannya dengan statistika, *data mining*, sistem informasi yang semula digunakan dalam bidang ekonomi dan industri, kemudian merambah ke bidang pendidikan. Seiring dengan perkembangan teknologi informasi dan penerapannya yang mulai meningkat dalam dunia pendidikan menjadikan data-data pembelajaran lebih mudah diperoleh dan peran LA menjadi lebih signifikan.

Learning analytics adalah pengukuran, pengumpulan, analisis, dan pelaporan data tentang peserta didik dan konteksnya, untuk tujuan memahami dan mengoptimalkan pembelajaran dan lingkungan tempat pembelajaran itu terjadi. Bidang terkait adalah penggalian data pendidikan. Untuk pengantar khalayak umum, lihat:

- Pengarahan Inisiatif Pembelajaran Educause
- Tinjauan Educause tentang Analisis Pembelajaran
- Dan “Ringkasan Kebijakan Analisis Pembelajaran” dari UNESCO (2012)

Pengertian Analisis Pembelajaran

Definisi dan tujuan analisis pembelajaran masih diperdebatkan. Satu definisi sebelumnya yang dibahas oleh komunitas menyarankan bahwa "Analisis pembelajaran adalah penggunaan data cerdas, data yang dihasilkan peserta, dan model analisis untuk menemukan informasi dan hubungan sosial untuk memprediksi dan metode penedekatan dalam pembelajaran berbasir *people learning*."

Tetapi definisi in mendapat sanggahan dari beberapa tokoh:

1. “Saya agak tidak setuju dengan definisi ini - ini berfungsi dengan baik sebagai konsep pengantar jika kita menggunakan analitik sebagai struktur pendukung untuk model pendidikan yang ada. Menurut saya, mempelajari analitik - pada implementasi yang canggih dan terintegrasi - dapat menghilangkan model kurikulum yang luar biasa ”. George Siemens, 2010.
2. “Dalam deskripsi pembelajaran analitik, kita berbicara tentang penggunaan data untuk“ memprediksi kesuksesan ”. Saya kesulitan dengan itu saat saya mempelajari database kami. Saya menyadari ada berbagai pandangan / tingkat kesuksesan. " Mike Sharkey 2010.

Pandangan yang lebih holistik daripada definisi belaka disediakan oleh kerangka analisis pembelajaran oleh Greller dan Drachsler (2012). Ini menggunakan analisis morfologi umum (GMA) untuk membagi domain menjadi enam "dimensi kritis".

Tinjauan sistematis tentang analisis pembelajaran dan konsep utamanya disediakan oleh Chatti et al. (2012) dan Chatti et al. (2014) melalui model referensi pembelajaran analitik berdasarkan empat dimensi yaitu data, lingkungan, konteks (apa?), Pemangku kepentingan (siapa?), Tujuan (mengapa?), Dan metode (bagaimana?).

Telah ditunjukkan bahwa ada kesadaran luas tentang analitik di seluruh lembaga pendidikan untuk berbagai pemangku kepentingan, tetapi cara 'analitik pembelajaran' didefinisikan dan diterapkan dapat bervariasi, termasuk:

1. bagi pelajar individu untuk merefleksikan pencapaian dan pola perilaku mereka dalam hubungannya dengan orang lain;
2. sebagai prediktor siswa yang membutuhkan dukungan dan perhatian ekstra;
3. untuk membantu guru dan staf pendukung merencanakan intervensi pendukung dengan individu dan kelompok;
4. untuk kelompok fungsional seperti tim kursus yang ingin meningkatkan kursus saat ini atau mengembangkan penawaran kurikulum baru; dan
5. untuk administrator kelembagaan yang mengambil keputusan tentang hal-hal seperti pemasaran dan rekrutmen atau efisiensi dan efektivitas. "

Dalam makalah pengarahan tersebut, Powell dan MacNeill selanjutnya menunjukkan bahwa beberapa motivasi dan implementasi analitik mungkin bertentangan dengan yang lain, misalnya menyoroti potensi konflik antara analitik untuk pelajar individu dan pemangku kepentingan organisasi.

Gašević, Dawson, dan Siemens berpendapat bahwa aspek komputasi dari analisis pembelajaran perlu dikaitkan dengan penelitian pendidikan yang ada jika bidang analisis pembelajaran ingin memenuhi janjinya untuk memahami dan mengoptimalkan pembelajaran.

Perbedaan Analisis Pembelajaran dan Data Mining Pendidikan

Membedakan data mining di bidang pendidikan (*Educational Data Mining*/EDM) dan Analytics Learning (LA) telah menjadi perhatian beberapa peneliti. George Siemens mengambil posisi bahwa penambangan data pendidikan mencakup analitik pembelajaran dan analitik akademis, yang sebelumnya ditujukan untuk pemerintah, lembaga pendanaan, dan administrator, bukan pelajar dan fakultas. Baepler dan Murdoch mendefinisikan analitik akademis sebagai area yang "... menggabungkan data kelembagaan pilihan, analisis statistik, dan pemodelan prediktif untuk menciptakan kecerdasan yang dapat digunakan oleh pelajar, instruktur, atau administrator untuk mengubah perilaku akademis". Mereka terus mencoba untuk membedakan penambangan data pendidikan dari analitik akademis berdasarkan apakah prosesnya didorong oleh hipotesis atau tidak, meskipun Brooks mempertanyakan apakah perbedaan ini ada dalam literatur. Brooks justru mengusulkan bahwa perbedaan yang lebih baik antara komunitas EDM dan LA ada di akar dari mana

setiap komunitas berasal, dengan kepenulisan di komunitas EDM didominasi oleh peneliti yang berasal dari paradigma bimbingan cerdas, dan peneliti anaytics belajar lebih fokus pada sistem pembelajaran perusahaan (misalnya mempelajari sistem manajemen konten).

Terlepas dari perbedaan antara LA dan EDM, kedua bidang ini memiliki pemahaman yang saling tumpang tindih baik dalam tujuan penyidik maupun dalam metode dan teknik yang digunakan dalam penyelidikan.

.

Sejarah

Konteks Analisis Pembelajaran

Dalam "*The State of Learning Analytics in 2012: A Review and Future Challenges*" Rebecca Ferguson melacak kemajuan analitik untuk pembelajaran sebagai pengembangan melalui:

1. Meningkatnya minat 'data besar' untuk intelijen bisnis
2. Maraknya pendidikan online yang berfokus pada Virtual Learning Environments (VLEs), Content Management Systems (CMSs), dan Management Information Systems (MIS) untuk pendidikan, yang memperlihatkan peningkatan data digital tentang latar belakang siswa (sering diadakan di MIS) dan data log pembelajaran (dari VLE). Perkembangan ini memberikan kesempatan untuk menerapkan teknik 'kecerdasan bisnis' untuk data pendidikan
3. Pertanyaan tentang pengoptimalan sistem untuk mendukung pembelajaran terutama diberikan pertanyaan tentang bagaimana kita dapat mengetahui apakah siswa terlibat / pemahaman jika kita tidak dapat melihatnya?
4. Meningkatkan fokus pada pembuktian kemajuan dan standar profesional untuk sistem akuntabilitas
5. Fokus ini mengarah pada pemegang saham guru dalam analitik - mengingat bahwa mereka terkait dengan sistem akuntabilitas
6. Dengan demikian penekanan yang meningkat ditempatkan pada kemampuan pedagogik dari pembelajaran analitik
7. Tekanan ini meningkat oleh keinginan ekonomi untuk meningkatkan keterlibatan dalam pendidikan online untuk penyampaian pendidikan berkualitas tinggi - terjangkau -

Asal Mula Munculnya Teknik dan Metode Analisis Pembelajaran

Dalam diskusi tentang sejarah analitik, Cooper memperolah teknik analisis pembelajran setelah meneliti sejumlah komunitas, teknik-teknik tersebut mencakup:

1. Statistik - yang merupakan sarana utama untuk menangani pengujian hipotesis. Untuk penetapan dan penerapan sebuah kondisi bagaimana suatu hipotesis dapat digunakan atau membantu melakukan sesuatu
2. Business Intelligence - yang memiliki kesamaan dengan analisis pembelajaran, meskipun secara historis ditargetkan untuk membuat produksi laporan lebih efisien dengan mengaktifkan akses data dan meringkas indikator kinerja.
3. Analisis web - alat seperti laporan analisis Google terhadap kunjungan halaman web dan referensi ke situs web, merek, dan istilah lain di internet. Semakin banyak 'butir halus' (*fine grain* : komponen yang lebih kecil di mana komponen yang lebih besar tersusun, layanan tingkat rendah) dari teknik ini dapat diadopsi dalam analisis pembelajaran untuk eksplorasi lintasan siswa melalui sumber belajar (kursus, materi, dll.).
4. Riset operasional - bertujuan untuk menyoroti optimasi desain untuk memaksimalkan tujuan melalui penggunaan model matematika dan metode statistik. Teknik-teknik tersebut diimplikasikan dalam analisis pembelajaran yang berusaha menciptakan model perilaku dunia nyata untuk aplikasi praktis.
5. Kecerdasan buatan dan Data mining - Teknik pembelajaran mesin yang dibangun di atas data mining dan metode AI mampu mendeteksi pola dalam data. Dalam mempelajari analitik, teknik semacam itu dapat digunakan untuk sistem bimbingan cerdas, klasifikasi siswa dengan cara yang lebih dinamis daripada faktor demografis sederhana, dan sumber daya seperti sistem 'kursus yang disarankan' yang meniru teknik penyaringan kolaboratif.
6. Analisis Jaringan Sosial - *Social Network Analysis* (SNA) menganalisis hubungan antara orang-orang dengan menjelajahi hubungan implisit (misalnya interaksi di forum) dan eksplisit (misalnya 'teman' atau 'pengikut') online dan offline. SNA dikembangkan dari karya sosiolog seperti Wellman dan Watts, dan ahli matematika seperti Barabasi dan Strogatz. Pekerjaan individu-individu ini telah memberi kita pemahaman yang baik tentang pola yang ditunjukkan jaringan (dunia kecil, hukum kekuatan), atribut koneksi (di awal 70-an, Granovetter mengeksplorasi koneksi dari perspektif kekuatan dan dampaknya pada informasi baru) , dan dimensi jaringan sosial (misalnya, geografi masih penting dalam dunia jaringan digital). Ini terutama digunakan untuk menjelajahi kelompok jaringan, mempengaruhi jaringan, keterlibatan dan pelepasan, dan telah digunakan untuk tujuan ini dalam mempelajari konteks analitik.
7. Visualisasi informasi - visualisasi adalah langkah penting dalam banyak analitik untuk memahami data yang disediakan - sehingga digunakan di sebagian besar teknik (termasuk yang di atas).

Sejarah Analisis Pembelajaran pada Perguruan Tinggi

Program pascasarjana pertama yang berfokus secara khusus pada analisis pembelajaran dibuat oleh Dr. Ryan Baker dan diluncurkan pada semester Musim Gugur 2015 di *Teachers College* - Universitas Columbia. Deskripsi program menyatakan bahwa “data tentang pembelajaran dan pelajar dihasilkan hari ini dalam skala yang belum pernah terjadi sebelumnya. Bidang analisis pembelajaran (LA) dan penambangan data pendidikan (EDM) telah muncul dengan tujuan mengubah data ini menjadi wawasan baru yang dapat bermanfaat bagi siswa, guru, dan administrator. Sebagai salah satu lembaga penelitian dan pengajaran terkemuka di dunia di bidang pendidikan, psikologi, dan kesehatan, kami bangga menawarkan kurikulum pascasarjana inovatif yang didedikasikan untuk meningkatkan pendidikan melalui teknologi dan analisis data."

Metode Analisis

Metode untuk mempelajari analisis meliputi:

- Analisis isi - terutama sumber daya yang dibuat siswa (seperti esai)
- Analisis Wacana Analisis wacana bertujuan untuk menangkap data yang berarti tentang interaksi siswa yang (tidak seperti 'analisis jaringan sosial') yang bertujuan untuk menjelajahi properti bahasa yang digunakan, bukan hanya jaringan interaksi, atau jumlah posting forum, dll.
- Analisis Pembelajaran Sosial yang bertujuan untuk mengeksplorasi peran interaksi sosial dalam pembelajaran, pentingnya jaringan pembelajaran, wacana yang digunakan untuk merasakan, dll.
- Analisis Disposisi yang berusaha untuk menangkap data mengenai disposisi siswa untuk pembelajaran mereka sendiri, dan hubungannya dengan pembelajaran mereka. Misalnya, pelajar yang "penasaran" mungkin lebih cenderung untuk mengajukan pertanyaan - dan data ini dapat diambil dan dianalisis untuk analitik pembelajaran.

Penggunaan Hasil Analisis

- Tujuan prediksi, misalnya untuk mengidentifikasi siswa yang 'berisiko' dalam hal drop out atau kegagalan kursus
- Personalisasi & Adaptasi, untuk memberi siswa jalur pembelajaran yang disesuaikan, atau materi penilaian
- Tujuan intervensi, memberikan informasi kepada pendidik untuk melakukan intervensi guna mendukung siswa
- Visualisasi informasi, biasanya dalam bentuk apa yang disebut dasbor pembelajaran yang memberikan gambaran umum data pembelajaran melalui alat visualisasi data

Perangkat Lunak

Banyak perangkat lunak yang saat ini digunakan untuk mempelajari fungsionalitas duplikat analitik dari perangkat lunak analisis web, tetapi menerapkannya untuk interaksi pelajar dengan konten. Alat analisis jaringan sosial biasanya digunakan untuk memetakan hubungan dan diskusi sosial. Beberapa contoh alat perangkat lunak analitik pembelajaran:

- *Student Success System*

  Student Success System menggunakan wawasan analitis untuk membuat prediksi tentang keberhasilan siswa dan tingkat risiko dalam pembelajaran. Analisis prediktif yang digunakan oleh *Student Success System* dapat disesuaikan dengan pendekatan instruksional setiap mata kuliah, memungkinkan untuk memantau keterlibatan siswa dan ekspektasi pencapaian untuk mata kuliah yang diajarkan. Dengan *Student Success System,* dapat menggunakan model prediktif untuk memantau dan merancang intervensi yang ditargetkan untuk siswa berisiko dalam pembelajaran aktif untuk meningkatkan keberhasilan, retensi, penyelesaian, dan tingkat kelulusan siswa.

- SNAPP

  SNAPP adalah perangkat lunak yang memungkinkan pengguna untuk memvisualisasikan jaringan interaksi yang dihasilkan dari postingan dan balasan forum diskusi. Visualisasi jaringan dari interaksi forum memberikan kesempatan bagi guru untuk dengan cepat mengidentifikasi pola perilaku pengguna - pada setiap tahap perkembangan kursus. SNAPP telah dikembangkan untuk mengekstrak semua interaksi pengguna dari berbagai sistem manajemen pembelajaran (LMS) komersial dan sumber terbuka seperti Papan tulis (termasuk WebCT sebelumnya), Moodle dan sekarang Sakai. SNAPP kompatibel untuk pengguna Mac dan PC dan beroperasi di Chrome dan Firefox.

  Sebagian besar data siswa yang dihasilkan dari Sistem Manajemen Pembelajaran (LMS) meliputi laporan jumlah sesi (*log-in*), waktu tunggu (berapa lama login berlangsung) dan jumlah download. Ini memberi tahu kita banyak hal tentang pengambilan konten dalam model transmisi pembelajaran dan pengajaran, tetapi tidak tentang bagaimana siswa berinteraksi satu sama lain dalam praktik yang lebih bersifat sosio-konstruktivis. Kegiatan forum diskusi merupakan indikator interaksi siswa yang baik dan ditangkap secara sistematis oleh sebagian besar LMS. SNAPP menggunakan informasi tentang siapa yang memposting dan membalas siapa, dan apa diskusi utama tentang, dan seberapa luas mereka, untuk menganalisis interaksi forum dan menampilkannya dalam Diagram Jaringan Sosial.

- LOCO-Analyst

LOCO-Analyst adalah alat pendidikan yang bertujuan untuk memberikan umpan balik kepada guru tentang aspek relevan dari proses pembelajaran yang berlangsung di lingkungan pembelajaran berbasis web, dan dengan demikian membantu mereka meningkatkan konten dan struktur kursus berbasis web mereka. LOCO-Analyst bertujuan untuk memberikan umpan balik kepada guru mengenai:

- semua jenis kegiatan yang dilakukan dan / atau diikuti oleh siswanya selama proses pembelajaran,
- • penggunaan dan kelengkapan konten pembelajaran yang telah mereka siapkan dan sebarkan di LCMS,
- • interaksi sosial kontekstual di antara siswa (yaitu, jejaring sosial) dalam lingkungan belajar virtual.

- *Student Activity Monitor*/SAM

  SAM mencakup seperangkat visualisasi kegiatan pelajar untuk meningkatkan kesadaran dan mendukung refleksi diri. Ini diimplementasikan sebagai widget dalam proyek ROLE

- BEESTAR INSIGHT

  Adalah Sistem *real time* yang secara otomatis mengumpulkan keterlibatan dan kehadiran siswa & menyediakan alat analitik dan dasbor untuk siswa, guru & manajemen. Beestar adalah organisasi pendidikan yang didirikan oleh guru dan orang tua Sugar Land.

  Melayani wilayah Houston sejak tahun 2003, Beestar mengkhususkan diri dalam mengembangkan matematika online dan latihan membaca untuk anak-anak sekolah dasar. Program setelah sekolah ini menumbuhkan minat siswa dan membantu mereka mencapai nilai akademik yang sangat baik. Sistem Beestar juga memfasilitasi para orang tua untuk berpartisipasi dalam proses belajar anaknya sehingga mereka dapat memahami dan mendampingi anaknya untuk berprestasi di sekolah.

- Solutionpath StREAM

  Sistem waktu nyata berbasis di Inggris Raya yang memanfaatkan model prediktif untuk menentukan semua aspek keterlibatan siswa menggunakan sumber terstruktur dan tidak terstruktur untuk semua peran kelembagaan. StREAM dibangun dengan fokus utama membantu setiap siswa untuk memahami dan mengelola perjalanan belajar mereka yang unik. Skor keterlibatan StREAM memberikan data yang dapat mereka gunakan untuk meninjau aktivitas mereka dan membuat keputusan berdasarkan pengetahuan ini yang dapat memengaruhi kesuksesan mereka di masa depan. Menawarkan skor keterlibatan harian dan menyoroti pola aktivitas, siswa Anda dapat menggunakan StREAM untuk lebih

memahami diri mereka sendiri dan perilaku mereka terkait dengan perjalanan belajar mereka.

Etika & Privasi

Etika pengumpulan data, analitik, pelaporan, dan akuntabilitas telah diangkat sebagai perhatian potensial untuk *Learning Analytics* (misalnya,), dengan masalah yang diangkat mengenai:

- Kepemilikan data
- Komunikasi seputar ruang lingkup dan peran *Learning Analytics*
- Peran penting dari umpan balik manusia dan koreksi kesalahan dalam sistem Learning Analytics
- Berbagi data antara sistem, organisasi, dan pemangku kepentingan
- Percaya pada klien data

Seperti yang ditunjukkan Kay, Kom dan Oppenheim, kisaran datanya luas, berpotensi berasal dari: "* Aktivitas yang direkam; catatan siswa, kehadiran, tugas, informasi peneliti (CRIS).

- Interaksi sistem; VLE, pencarian perpustakaan / repositori, transaksi kartu.
- Mekanisme umpan balik; survei, layanan pelanggan.
- Sistem eksternal yang menawarkan identifikasi yang andal seperti sektor dan layanan bersama serta jejaring sosial. "

Oleh karena itu, situasi hukum dan etika menjadi tantangan dan berbeda dari satu negara ke negara lain, meningkatkan implikasi untuk: "* Variasi data - prinsip pengumpulan, penyimpanan dan eksploitasi.

- Misi pendidikan - masalah yang mendasari manajemen pembelajaran, termasuk rekayasa sosial dan kinerja.
- Motivasi untuk pengembangan analitik - kebersamaan, kombinasi barang korporat, individu, dan umum.
- Harapan pelanggan - praktik bisnis yang efektif, harapan data sosial, pertimbangan budaya dari basis pelanggan global. * Kewajiban untuk bertindak - tugas kepedulian yang timbul dari pengetahuan dan konsekuensi tantangan dari manajemen kinerja siswa dan karyawan. "

Dalam beberapa kasus penting seperti bencana inBloom bahkan sistem fungsional penuh telah ditutup karena kurangnya kepercayaan dalam pengumpulan data oleh pemerintah,

pemangku kepentingan, dan kelompok hak-hak sipil. Sejak saat itu, komunitas Learning Analytics telah mempelajari kondisi hukum secara ekstensif dalam serangkaian lokakarya pakar tentang '*Ethics & Privacy 4 Learning Analytics*' yang merupakan penggunaan Learning Analytics tepercaya. Drachsler & Greller merilis daftar periksa 8 poin bernama DELICATE yang didasarkan pada studi intensif di area ini untuk mengungkap diskusi etika dan privasi seputar Analisis Pembelajaran:

1. **D**-*etermination*: Tentukan tujuan pembelajaran analitik untuk institusi Anda.
2. **E**-*xplain*: Mendefinisikan ruang lingkup pengumpulan dan penggunaan data.
3. **L**-*egitimate*: Jelaskan bagaimana Anda beroperasi sesuai kerangka hukum, mengacu pada undang-undang yang berlaku.
4. **I**-*nvolve*: Bicaralah dengan pemangku kepentingan dan berikan jaminan tentang distribusi dan penggunaan data.
5. **C**-*onsent*: Mintalah persetujuan melalui pertanyaan persetujuan yang jelas.
6. **A**-*nonymise*: Sedapat mungkin tidak mengidentifikasi individu
7. **T**-*echnical aspect*: Memantau siapa yang memiliki akses data, terutama di wilayah dengan tingkat pergantian staf yang tinggi.
8. **E**-*xternal partner*: Pastikan pihak eksternal memberikan standar keamanan data tertinggi

Ini menunjukkan cara untuk merancang dan memberikan privasi yang sesuai dengan *Learning Analytics* yang dapat bermanfaat bagi semua pemangku kepentingan.

Open Learning Analytics

Open Learning Analytics (OLA) adalah bidang yang muncul, yang menangani data pembelajaran yang dikumpulkan dari berbagai lingkungan dan konteks, dianalisis dengan berbagai metode analitik untuk memenuhi persyaratan pemangku kepentingan yang berbeda. OLA memperkenalkan seperangkat persyaratan dan implikasi untuk praktisi LA, pengembang, dan peneliti. Ini termasuk agregasi dan integrasi data, interoperabilitas, dapat digunakan kembali, modularitas, fleksibilitas, ekstensibilitas, kinerja, skalabilitas, kegunaan, privasi, dan personalisasi. Platform Analisis Pembelajaran Terbuka (OpenLAP) adalah kerangka kerja yang menangani masalah ini dan meletakkan dasar untuk ekosistem OLA.

OpenLAP menyediakan arsitektur OLA teknis terperinci dengan implementasi konkret dari semua komponennya, terintegrasi dengan mulus dalam platform. Ini mencakup berbagai pemangku kepentingan yang terkait melalui minat yang sama dalam analisis pembelajaran

tetapi dengan beragam kebutuhan dan tujuan, berbagai data yang berasal dari berbagai lingkungan dan konteks pembelajaran, serta berbagai infrastruktur dan metode yang memungkinkan untuk menarik nilai dari data untuk mendapatkan keuntungan. wawasan tentang proses pembelajaran.

OpenLAP mengikuti pendekatan yang berpusat pada pengguna untuk melibatkan pengguna akhir dalam definisi yang fleksibel dan pembuatan indikator yang dipersonalisasi secara dinamis. Indikator yang dihasilkan dijalankan dengan membuat kueri data, menerapkan filter, melakukan analisis, dan menghasilkan visualisasi untuk ditampilkan di sisi klien. Untuk memenuhi kebutuhan pengguna yang beragam, OpenLAP menyediakan arsitektur modular dan dapat diperluas yang memungkinkan integrasi yang mudah dari modul analitik baru, metode analitik, dan teknik visualisasi.

2.7 Analisis Prediktif

Analisis prediktif mencakup berbagai teknik statistik dari pemodelan prediktif, pembelajaran mesin, dan penggalian data yang menganalisis fakta terkini dan historis untuk membuat prediksi tentang peristiwa yang akan datang atau peristiwa yang tidak diketahui.

Dalam bisnis, model prediktif memanfaatkan pola yang ditemukan dalam data historis dan transaksional untuk mengidentifikasi risiko dan peluang. Model menangkap hubungan di antara banyak faktor untuk memungkinkan penilaian risiko atau potensi yang terkait dengan serangkaian kondisi tertentu, memandu pengambilan keputusan untuk calon transaksi.

Efek fungsional yang menentukan dari pendekatan teknis ini adalah bahwa analitik prediktif memberikan skor prediktif (probabilitas) untuk setiap individu (pelanggan, karyawan, pasien perawatan kesehatan, SKU produk, kendaraan, komponen, mesin, atau unit organisasi lainnya) untuk menentukan, menginformasikan , atau memengaruhi proses organisasi yang berkaitan dengan sejumlah besar individu, seperti dalam pemasaran, penilaian risiko kredit, deteksi penipuan, manufaktur, perawatan kesehatan, dan operasi pemerintah termasuk penegakan hukum.

Analisis prediktif digunakan dalam ilmu aktuaria, pemasaran, layanan keuangan, asuransi, telekomunikasi, ritel, perjalanan, perawatan kesehatan, perlindungan anak, farmasi, perencanaan kapasitas, dan bidang lainnya.

Salah satu aplikasi yang paling terkenal adalah penilaian kredit, yang digunakan di seluruh layanan keuangan. Model penilaian memproses riwayat kredit pelanggan, pengajuan pinjaman, data pelanggan, dll., Untuk menentukan peringkat individu berdasarkan kemungkinan mereka melakukan pembayaran kredit di masa depan tepat waktu

Definisi

Analisis prediktif adalah area penambangan data yang berhubungan dengan penggalian informasi dari data dan menggunakannya untuk memprediksi tren dan pola perilaku. Seringkali peristiwa menarik yang tidak diketahui terjadi di masa depan, tetapi analitik prediktif dapat diterapkan ke semua jenis hal yang tidak diketahui apakah itu di masa lalu, sekarang atau masa depan. Misalnya, mengidentifikasi tersangka setelah kejahatan dilakukan, atau penipuan kartu kredit saat terjadi. Inti dari analitik prediktif bergantung pada menangkap hubungan antara variabel penjelas dan variabel yang diprediksi dari kejadian masa lalu, dan memanfaatkannya untuk memprediksi hasil yang tidak diketahui. Namun, penting untuk dicatat bahwa keakuratan dan kegunaan hasil akan sangat bergantung pada tingkat analisis data dan kualitas asumsi.

Analisis prediktif sering didefinisikan sebagai prediksi pada tingkat perincian yang lebih rinci, yaitu menghasilkan skor prediktif (probabilitas) untuk setiap elemen organisasi. Ini membedakannya dari peramalan. Misalnya, “Analisis prediktif — Teknologi yang belajar dari pengalaman (data) untuk memprediksi perilaku individu di masa depan untuk mendorong keputusan yang lebih baik.” Dalam sistem industri masa depan, nilai analitik prediktif akan memprediksi dan mencegah potensi masalah untuk mencapai kerusakan mendekati nol dan selanjutnya diintegrasikan ke dalam analitik preskriptif untuk pengoptimalan keputusan. Selanjutnya data hasil konversi dapat digunakan untuk perbaikan siklus hidup produk closed-loop yang merupakan visi *Industrial Internet Consortium*.

Jenis

Umumnya, istilah analisis prediktif digunakan untuk memodelkan prediktif, "penilaian" data dengan model prediktif, dan perkiraan. Namun, orang semakin banyak menggunakan istilah ini untuk merujuk pada disiplin analitik terkait, seperti pemodelan deskriptif dan pemodelan atau pengoptimalan keputusan. Disiplin ini juga melibatkan analisis data yang ketat, dan banyak digunakan dalam bisnis untuk segmentasi dan pengambilan keputusan, tetapi memiliki tujuan yang berbeda dan teknik statistik yang mendasari mereka bervariasi.

Model Prediktif

Model prediktif adalah model hubungan antara kinerja spesifik suatu unit dalam sampel dan satu atau lebih atribut atau fitur unit yang diketahui. Tujuan dari model ini adalah untuk menilai kemungkinan bahwa unit serupa dalam sampel yang berbeda akan menunjukkan kinerja tertentu. Kategori ini mencakup model di banyak area, seperti pemasaran, di mana mereka mencari pola data halus untuk menjawab pertanyaan tentang kinerja pelanggan, atau model deteksi penipuan. Model prediktif sering melakukan perhitungan selama transaksi langsung, misalnya, untuk mengevaluasi risiko atau peluang pelanggan atau transaksi tertentu, untuk memandu keputusan. Dengan kemajuan dalam kecepatan komputasi, sistem pemodelan agen individu telah mampu mensimulasikan perilaku atau reaksi manusia terhadap rangsangan atau skenario yang diberikan.

Unit sampel yang tersedia dengan atribut yang diketahui dan performa yang diketahui disebut sebagai "sampel pelatihan". Unit dalam sampel lain, dengan atribut yang diketahui tetapi performanya tidak diketahui, disebut sebagai unit "sampel di luar [pelatihan]". Sampel yang keluar tidak memiliki hubungan kronologis dengan unit sampel pelatihan. Misalnya, sampel pelatihan mungkin terdiri dari atribut sastra dari tulisan-tulisan oleh penulis Victoria, dengan atribusi yang diketahui, dan unit sampel di luar mungkin ditemukan tulisan baru dengan penulis yang tidak diketahui; model prediktif dapat membantu dalam menghubungkan sebuah karya dengan penulis yang dikenal. Contoh lain diberikan oleh analisis percikan darah pada simulasi TKP dimana out of sample unit adalah pola percikan darah yang sebenarnya dari TKP. Unit sampel yang keluar mungkin berasal dari waktu yang sama dengan unit pelatihan, dari waktu sebelumnya, atau dari waktu mendatang.

Model Deskriptif

Model deskriptif mengukur hubungan dalam data dengan cara yang sering digunakan untuk mengklasifikasikan pelanggan atau prospek ke dalam kelompok. Tidak seperti model prediktif yang berfokus pada prediksi perilaku pelanggan tunggal (seperti risiko kredit), model deskriptif mengidentifikasi banyak hubungan yang berbeda antara pelanggan atau produk. Model deskriptif tidak mengurutkan pelanggan berdasarkan kemungkinan mereka mengambil tindakan tertentu seperti yang dilakukan model prediktif. Sebaliknya, model deskriptif dapat digunakan, misalnya, untuk mengkategorikan pelanggan berdasarkan preferensi produk dan tahap kehidupan mereka. Alat pemodelan deskriptif dapat digunakan untuk mengembangkan model lebih lanjut yang dapat mensimulasikan sejumlah besar agen individual dan membuat prediksi.

Model Keputusan

Model keputusan menggambarkan hubungan antara semua elemen keputusan — data yang diketahui (termasuk hasil model prediktif), keputusan, dan hasil ramalan keputusan — untuk memprediksi hasil keputusan yang melibatkan banyak variabel. Model ini dapat digunakan dalam pengoptimalan, memaksimalkan hasil tertentu sambil meminimalkan yang lain. Model keputusan umumnya digunakan untuk mengembangkan logika keputusan atau seperangkat aturan bisnis yang akan menghasilkan tindakan yang diinginkan untuk setiap pelanggan atau keadaan.

Aplikasi

Meskipun analisis prediktif dapat digunakan di banyak aplikasi, kami menguraikan beberapa contoh di mana analisis prediktif telah menunjukkan dampak positif dalam beberapa tahun terakhir.

Customer Relationship Management (CRM)

Analisis CRM atau Manajemen Hubungan Pelanggan adalah aplikasi komersial yang sering digunakan untuk analisis prediktif. CRM (Customer Relationship Management) adalah strategi bisnis yang memadukan proses, manusia dan teknologi. Membantu menarik prospek penjualan, mengkonfersi mereka menjadi pelanggan, dan mempertahankan pelanggan yang sudah ada, pelanggan yang puas dan loyal.

Tujuan dari CRM adalah untuk mengetahui sebanyak mungkin tentang bagaimana kebutuhan dan perilaku pelanggan, untuk selanjutnya memberikan sebuah pelayanan yang optimal dan mempertahankan hubungan yang sudah ada, karena kunci sukses dari bisnis sangat tergantung seberapa jauh kita tahu tentang pelanggan dan memenuhi kebutuhan mereka. Sulit bagi sebuah perusahaan untuk mencapai dan mempertahankan kepemimpinan dan profitabilitas tanpa melakukan fokus secara berkesinambungan yang dapat dilakukan pada CRM.
CRM menjangkau banyak bidang dalam organisasi, termasuk : Penjualan, Layanan Pelanggan dan Pemasaran

Health Care

Healt care merupakan aplikasi yang digunakan dalam proses prediktif analisis tentang perawatan kesehatan. Aplikasi ini akan menentukan informasi mengenai pasien yang berisiko mengalami kondisi tertentu, seperti asma, diabetes, dan penyakit seumur hidup

lainnya. Sistem aplikasi ini akan mendukung keputusan klinis yang menggabungkan prediktif analitik untuk bis mengambil keputusan medis pada titik perawatan.

Sistem Pengambil Keputusan Klinis

Sistem Pengambil Keputusan Klinis (*Clinical Decision Support System* / CDSS) merupakan suatu sistem elektronik yang didesain untuk membantu klinisi atau tenaga medis dalam mengambil keputusan klinik. Pada penggunaan CDSS yang berbasis elektronik memiliki beberapa keunggulan dan kemudahan, jika dibandingkan dengan non-elektronik, apalagi jika sudah terintegrasi dengan rekam kesehatan elektronik (Services Human & Investigator 2012).

Ada beberapa keunggulan komputerisasi dengan CDSS, diantaranya adalah kapasitas penyimpanan *knowledge based* dan kecepatan menganalisa sebuah kasus, serta dalam memberikan rekomendasi kepada klinisi dalam bentuk alert atau peringatan (Lee et al. 2014). Pada Umumnya CDSS elektronik mengkombinasikan karakteristik klinis dan kondisi pasien, dengan basis pengetahuan elektronik (*computerized knowledge base*), yang kemudian secara otomatis menghasilkan rekomendasi-rekomendasi untuk bahan pertimbangan klinisi, baik dokter, perawat, bidan dan tenaga kesehatan lain, yang kemudian dapat membantu dalam menentukan diagnosis dan pemberian tindakan medis lainnya (Parshutin & Kirshners 2013).

CDSS merupakan media elektronik yang digunakan untuk menentukan diagnosis, interpretasi klinis, pemberitahuan (alerting), pengingat (reminder), analisis prediktif dengan sebuah aplikasi, yang terhubung dengan data (Aynes & Aplan 2001). Definisi lain mengatakan bahwa CDSS menyediakan informasi bagi tenaga medis, pasien atau individu atau populasi tertentu, untuk menghasilkan proses kesehatan yang lebih cepat, lebih efisien, lebih baik baik bagi layanan kesehatan individual maupun bagi kesehatan suatu populasi (Sheikhtaheri et al. 2012).

Dari definisi diatas, dapat disimpulkan bahwa CDSS memiliki tujuan utama untuk mendukung berbagai macam fungsi klinis, seperti: dokumentasi dan pengkodean klinis, mengatur kompleksitas klinis, menyimpan dan memelihara database pasien, melakukan tracking order pasien, monitoring dan tindak lanjut kesehatan, maupun tindakan pencegahan suatu penyakit (Main et al. 2010).

Collection Analytics

Perusahaan penagih hutang sekarang beralih ke *speech analytics* - proses menganalisa konten relevan dari panggilan yang direkam - untuk membantu mereka mengurangi kenakalan dan mengurangi kerugian yang memungkinkan bisnis untuk memaksimalkan pemulihan piutang mereka. Alat bantu analitik koleksi untuk memahami preferensi pelanggan dan pola perilaku, yang pada gilirannya membantu mengembangkan strategi pengumpulan yang lebih baik.

Strategi pengumpulan sangat dibutuhkan untuk meningkatkan produktivitas. Tidaklah mungkin untuk menyewa agen (yang membutuhkan uang) untuk tetap melakukan panggilan penagihan dari daftar pembayaran yang jatuh tempo. Strategi pengumpulan membantu untuk menentukan akun mana yang memiliki kemungkinan kerugian lebih tinggi, mengkategorikan berbagai jenis pelanggan, dan memprioritaskan dan menargetkan pelanggan.

Cross-Sell

Seringkali organisasi perusahaan mengumpulkan dan memelihara data yang berlimpah (misalnya catatan pelanggan, transaksi penjualan) karena memanfaatkan hubungan tersembunyi dalam data dapat memberikan keunggulan kompetitif. Untuk organisasi yang menawarkan beberapa produk, analitik prediktif dapat membantu menganalisis pengeluaran pelanggan, penggunaan, dan perilaku lainnya, yang mengarah ke penjualan silang yang efisien, atau menjual produk tambahan ke pelanggan saat ini. Ini secara langsung mengarah pada profitabilitas yang lebih tinggi per pelanggan dan hubungan pelanggan yang lebih kuat. *Cross sell* berfungsi untuk menganalisis pengeluaran yang terjadi pada pelanggan, penggunaan dan perilaku lainnya, Dan mengarah pada penjualan silang yang efisien. Aplikasi ini juga bisa dijadikan sebagai media untuk menjual produk tambahan kepada pelanggan tersebut.

Retensi pelanggan

Dengan jumlah layanan yang bersaing yang tersedia, bisnis perlu memfokuskan upaya untuk mempertahankan kepuasan pelanggan yang berkelanjutan, menghargai loyalitas konsumen, dan meminimalkan pengurangan pelanggan. Selain itu, sedikit peningkatan retensi pelanggan telah terbukti meningkatkan keuntungan secara tidak proporsional. Satu studi menyimpulkan bahwa peningkatan 5% dalam tingkat retensi pelanggan akan meningkatkan keuntungan sebesar 25% hingga 95%. Bisnis cenderung menanggapi pengurangan pelanggan secara reaktif, bertindak hanya setelah pelanggan memulai

proses untuk menghentikan layanan. Pada tahap ini, peluang untuk mengubah keputusan pelanggan hampir nol. Penerapan analitik prediktif yang tepat dapat mengarah pada strategi retensi yang lebih proaktif. Dengan pemeriksaan rutin terhadap penggunaan layanan pelanggan sebelumnya, kinerja layanan, pengeluaran dan pola perilaku lainnya, model prediktif dapat menentukan kemungkinan pelanggan menghentikan layanan dalam waktu dekat. Intervensi dengan penawaran yang menguntungkan dapat meningkatkan kemungkinan mempertahankan pelanggan. Atrisi diam-diam, perilaku pelanggan yang perlahan tapi pasti mengurangi penggunaan, adalah masalah lain yang dihadapi banyak perusahaan. Analisis prediktif juga dapat memprediksi perilaku ini, sehingga perusahaan dapat mengambil tindakan yang tepat untuk meningkatkan aktivitas pelanggan.

Direct Marketing

Saat memasarkan produk dan layanan konsumen, ada tantangan untuk mengikuti produk dan perilaku konsumen yang bersaing. Selain mengidentifikasi prospek, analitik prediktif juga dapat membantu mengidentifikasi kombinasi paling efektif dari versi produk, materi pemasaran, saluran komunikasi, dan waktu yang harus digunakan untuk menargetkan konsumen tertentu. Tujuan analitik prediktif biasanya untuk menurunkan biaya per pesanan atau biaya per tindakan. Aplikasi ini berfungsi untuk melakukan identifikasi terhadap kombinasi yang paling efektif dari versi produk, soluran komunikasi, materi pemasaran, dan waktu yang dipergunakan untuk menargetkan pelanggan tertentu.

Fraud Detection

Aplikasi prediktif analisis yang berfungsi untuk menemukan aplikasi kredit yang tidak akurat, Dan adanya transaksi penipuan yang terjadi saat online maupun offline, klaim asuransi palsu dan pencurian data identitas.

Penipuan adalah masalah besar bagi banyak bisnis dan dapat dari berbagai jenis: aplikasi kredit yang tidak akurat, transaksi penipuan (baik offline maupun online), pencurian identitas dan klaim asuransi palsu. Masalah-masalah ini mengganggu perusahaan dari semua ukuran di banyak industri. Beberapa contoh yang mungkin menjadi korban adalah penerbit kartu kredit, perusahaan asuransi, pedagang eceran, produsen, pemasok bisnis-ke-bisnis, dan bahkan penyedia layanan. Model prediktif dapat membantu menyingkirkan "hal-hal buruk" dan mengurangi risiko bisnis terhadap penipuan.

Pemodelan prediktif juga dapat digunakan untuk mengidentifikasi kandidat penipuan berisiko tinggi dalam bisnis atau sektor publik. Mark Nigrini mengembangkan metode

penilaian risiko untuk mengidentifikasi target audit. Dia menjelaskan penggunaan pendekatan ini untuk mendeteksi penipuan dalam laporan penjualan franchisee dari rantai makanan cepat saji internasional. Setiap lokasi dinilai menggunakan 10 prediktor. 10 skor tersebut kemudian dibobot untuk memberikan satu skor risiko keseluruhan akhir untuk setiap lokasi. Pendekatan penilaian yang sama juga digunakan untuk mengidentifikasi rekening cek kiting berisiko tinggi, agen perjalanan yang berpotensi menipu, dan vendor yang meragukan. Model yang cukup kompleks digunakan untuk mengidentifikasi laporan bulanan yang curang yang dikirimkan oleh pengontrol divisi. *Internal Revenue Service (*IRS) Amerika Serikat juga menggunakan analitik prediktif untuk menambang pengembalian pajak dan mengidentifikasi penipuan pajak.

Kemajuan terbaru dalam teknologi juga telah memperkenalkan analisis perilaku prediktif untuk deteksi penipuan web. Jenis solusi ini menggunakan heuristik untuk mempelajari perilaku pengguna web normal dan mendeteksi anomali yang menunjukkan upaya penipuan.

Portofolio, Prediksi Tingkat Produk atau Ekonomi

Seringkali fokus analisis bukanlah konsumen tetapi produk, portofolio, perusahaan, industri atau bahkan ekonomi. Misalnya, pengecer mungkin tertarik untuk memprediksi permintaan tingkat toko untuk tujuan manajemen inventaris. Atau Dewan Federal Reserve mungkin tertarik untuk memprediksi tingkat pengangguran untuk tahun depan. Jenis masalah ini dapat diatasi dengan analitik prediktif menggunakan teknik deret waktu. Mereka juga dapat ditangani melalui pendekatan pembelajaran mesin yang mengubah deret waktu asli menjadi ruang vektor fitur, di mana algoritme pembelajaran menemukan pola yang memiliki kekuatan prediksi.

Manajemen Risiko Proyek/ *Project Risk Management*

Aplikasi ini untuk memprediksi portofolio terbaik yang berguna untuk memaksimalkan pengembalian dalam model penentuan harga aset modal. Bukan hanya itu *Risk management* juga berfungsi sebagai penilaian resiko probabilistik untuk menghasilkan prediksi yang akurat. Manajemen Risiko Proyek adalah proses sistematis seperti mengidentifikasi, merencanakan, menganalisis dan merespon setiap risiko yang muncul selama siklus hidup suatu proyek agar proyek tersebut tetap berada pada jalur yang terkendali hingga tercapainya tujuan dari proyek tersebut. Pada umumnya, tujuan utama dari manajemen risiko proyek adalah mencegah atau meminimalisasi akibat dari suatu kejadian yang tidak terduga dengan mempersiapkan rencana untuk keadaan yang tidak pasti yang

berkaitan dengan risiko tersebut. Manajemen risiko tidak hanya reaktif namun sudah menjadi keharusan pada proses perencanaannya untuk mencari tahu risiko apa saja yang mungkin terjadi selama pengerjaan proyek dan bagaimana tindakan yang diperlukan untuk mengendalikan risiko tersebut jika benar-benar terjadi nantinya.

Underwriting/Penjaminan

Banyak bisnis harus memperhitungkan paparan risiko karena layanan mereka yang berbeda dan menentukan biaya yang diperlukan untuk menutup risiko. Sebagai contoh, penyedia asuransi mobil perlu secara akurat menentukan jumlah premi yang akan dibebankan untuk menutup setiap mobil dan sopir. Perusahaan keuangan perlu menilai potensi dan kemampuan peminjam untuk membayar sebelum memberikan pinjaman. Untuk penyedia asuransi kesehatan, analisis prediktif dapat menganalisis beberapa tahun dari data klaim medis masa lalu, serta lab, apotek, dan catatan lainnya jika tersedia, untuk memprediksi seberapa mahal pendaftar mungkin berada di masa depan. Analisis prediktif dapat membantu menanggung jumlah ini dengan memprediksi kemungkinan penyakit, default, kebangkrutan, dll. Analisis prediktif dapat merampingkan proses akuisisi pelanggan dengan memprediksi perilaku risiko masa depan seorang pelanggan menggunakan data level aplikasi. Analisis prediktif dalam bentuk skor kredit telah mengurangi jumlah waktu yang diperlukan untuk persetujuan pinjaman, terutama di Indonesia pasar hipotek di mana keputusan peminjaman sekarang dibuat dalam hitungan jam, bukan hari atau bahkan berminggu-minggu. Analisis prediktif yang tepat dapat menghasilkan keputusan penetapan harga yang tepat, yang dapat membantu mengurangi risiko gagal bayar di masa mendatang.

Teknologi dan Pengaruh Big Data

Big data adalah kumpulan kumpulan data yang begitu besar dan kompleks sehingga menjadi canggung untuk digunakan dengan menggunakan alat manajemen database tradisional. Volume, variasi, dan kecepatan big data telah menghadirkan tantangan di seluruh papan untuk penangkapan, penyimpanan, pencarian, berbagi, analisis, dan visualisasi. Contoh sumber data besar termasuk log web, RFID, data sensor, jaringan sosial, pengindeksan pencarian Internet, catatan detail panggilan, pengawasan militer, dan data kompleks dalam ilmu astronomi, biogeokimia, genomik, dan atmosfer. Big Data adalah inti dari layanan analitik paling prediktif yang ditawarkan oleh organisasi TI. Berkat kemajuan teknologi dalam perangkat keras komputer — CPU yang lebih cepat, memori yang lebih murah, dan arsitektur MPP — dan teknologi baru seperti *Hadoop*, *MapReduce*, serta dalam database dan analitik teks untuk memproses data besar,

sekarang dimungkinkan untuk mengumpulkan, menganalisis, dan menambang sejumlah besar data terstruktur dan tidak terstruktur untuk wawasan baru. Algoritme prediktif juga dapat dijalankan pada data streaming. Saat ini, menjelajahi big data dan menggunakan analitik prediktif berada dalam jangkauan lebih banyak organisasi daripada sebelumnya dan metode baru yang mampu menangani kumpulan data tersebut diusulkan.

Teknik Analisis

Pendekatan dan teknik yang digunakan untuk melakukan analitik prediktif secara luas dapat dikelompokkan ke dalam teknik regresi dan teknik pembelajaran mesin.

Model regresi adalah treknik andalan analitik prediktif. Fokusnya terletak pada pembentukan persamaan matematika sebagai model untuk mewakili interaksi antara variabel yang berbeda dalam pertimbangan. Analisis regresi adalah metode untuk mengembangkan sebuah model (persamaan) yang menjelaskan hubungan di antara beberapa variabel. Output dari analisis regresi adalah sebuah persamaan regresi.

Pada model regresi variabel dibedakan menjadi dua bagian, yaitu variabel respons (response) atau biasa juga disebut variabel bergantung (dependent variable) serta variabel explanatory atau biasa juga disebut variabel penduga (predictor variable) atau disebut juga variabel bebas (independent variable).

Perlu dicatat, dalam analisis regresi sangat penting untuk membedakan sebuah variabel apakah termasuk dalam variabel dependen (tergantung) atau variabel independen (bebas). Misalnya, apabila ingin membuat sebuah model untuk menggambarkan hubungan antara gaji dengan pengalaman kerja maka bisa ditentukan gaji adalah variabel dependen dan pengalaman kerja adalah variabel independen.

Bergantung pada situasinya, ada berbagai macam model yang dapat diterapkan saat melakukan analitik prediktif. Beberapa di antaranya dibahas secara singkat di bawah ini.

Model Regresi Linier

Model regresi linier menganalisis hubungan antara respon atau variabel dependen dan sekumpulan variabel independen atau prediktor. Hubungan ini dinyatakan sebagai persamaan yang memprediksi variabel respon sebagai fungsi linier dari parameter. Parameter ini disesuaikan sehingga ukuran kecocokan dioptimalkan. Sebagian besar

upaya dalam penyesuaian model difokuskan pada meminimalkan ukuran residual, serta memastikan bahwa itu didistribusikan secara acak sehubungan dengan prediksi model.

Tujuan dari regresi adalah untuk memilih parameter model sehingga meminimalkan jumlah residual kuadrat. Ini disebut estimasi kuadrat terkecil (*ordinary least squares* - OLS) dan menghasilkan estimasi tidak bias linier terbaik (*best linear unbiased estimates* - BLUE) dari parameter jika dan hanya jika asumsi Gauss-Markov terpenuhi. BLUE adalah singkatan dari *Best, Linear, Unbiased Estimator*. *Best* artinya memiliki varians yang paling minimum diantara nilai varians alternatif setiap model yang ada. *Linear* artinya linier dalam variabel acak (Y). *Unbiased* artinya tidak bias atau nilai harapan dari *estimator* sama atau mendekati nilai parameter yang sebenarnya.

Setelah model diestimasi, kami akan tertarik untuk mengetahui apakah variabel prediktor termasuk dalam model — yaitu. apakah estimasi kontribusi masing-masing variabel dapat diandalkan? Untuk melakukan ini, kita dapat memeriksa signifikansi statistik dari koefisien model yang dapat diukur menggunakan statistik-t. Jumlah ini untuk menguji apakah koefisiennya berbeda secara signifikan dari nol. Seberapa baik model memprediksi variabel dependen berdasarkan nilai variabel independen dapat dinilai dengan menggunakan statistik R^2. Ini mengukur kekuatan prediksi model yaitu proporsi variasi total dalam variabel dependen yang "dijelaskan" (diperhitungkan) oleh variasi dalam variabel independen.

Model Pilihan Diskrit

Regresi multivariasi (di atas) umumnya digunakan ketika variabel respon kontinu dan memiliki rentang tak terbatas. Seringkali variabel respon mungkin tidak kontinu melainkan diskrit. Meskipun secara matematis layak untuk menerapkan regresi multivariat untuk variabel dependen berurutan diskrit, beberapa asumsi di balik teori regresi linier multivariat tidak lagi berlaku, dan ada teknik lain seperti model pilihan diskrit yang lebih cocok untuk jenis analisis ini. Jika variabel dependennya diskrit, beberapa metode unggulan tersebut adalah regresi logistik, multinomial logit dan model probit. Model regresi logistik dan probit digunakan ketika variabel dependen adalah biner.

Regresi logistik

Dalam pengaturan klasifikasi, menetapkan probabilitas hasil untuk observasi dapat dicapai melalui penggunaan model logistik, yang pada dasarnya adalah metode yang mengubah informasi tentang variabel dependen biner menjadi variabel kontinu tak terbatas dan memperkirakan model multivariat reguler. Uji Wald dan likelihood-ratio digunakan untuk

menguji signifikansi statistik dari setiap koefisien b dalam model. Pengujian yang menilai kesesuaian model klasifikasi adalah "persentase yang diprediksi dengan benar".

Regresi Logistik Multinomial

Perpanjangan model logit biner ke kasus di mana variabel dependen memiliki lebih dari 2 kategori adalah model logit multinomial. Dalam kasus seperti itu, menciutkan data menjadi dua kategori mungkin tidak masuk akal atau dapat menyebabkan hilangnya kekayaan data. Model logit multinomial adalah teknik yang sesuai dalam kasus ini, terutama jika kategori variabel dependen tidak diurutkan (untuk contoh warna seperti merah, biru, hijau). Beberapa penulis telah memperluas regresi multinomial untuk memasukkan metode pemilihan / kepentingan fitur seperti logit multinomial acak.

Regresi Probit

Model probit menawarkan alternatif untuk regresi logistik untuk memodelkan variabel dependen kategoris. Meskipun hasilnya cenderung serupa, distribusi yang mendasarinya berbeda. Model Probit populer dalam ilmu sosial seperti ekonomi. Cara yang baik untuk memahami perbedaan utama antara model probit dan logit adalah dengan mengasumsikan bahwa variabel dependen didorong oleh variabel laten z, yang merupakan jumlah dari kombinasi linear variabel penjelas dan suku noise acak.

Kami tidak mengamati z melainkan mengamati y yang mengambil nilai 0 (ketika z <0) atau 1 (sebaliknya). Dalam model logit kami mengasumsikan bahwa suku gangguan acak mengikuti distribusi logistik dengan mean nol. Dalam model probit kami mengasumsikan bahwa ia mengikuti distribusi normal dengan mean nol. Perhatikan bahwa dalam ilmu sosial (misalnya ekonomi), probit sering digunakan untuk memodelkan situasi di mana variabel y yang diamati bersifat kontinu tetapi mengambil nilai antara 0 dan 1.

Logit Versus Probit

Model probit sudah ada lebih lama dari model logit. Mereka berperilaku serupa, hanya saja distribusi logistik cenderung berekor lebih datar. Salah satu alasan mengapa model logit dirumuskan adalah model probit sulit secara komputasi karena persyaratan integral yang menghitung secara numerik. Namun komputasi modern telah membuat perhitungan ini cukup sederhana. Koefisien yang diperoleh dari model logit dan probit cukup dekat. Namun, rasio peluang lebih mudah diinterpretasikan dalam model logit.
Alasan praktis untuk memilih model probit daripada model logistik adalah:

- Ada keyakinan kuat bahwa distribusi yang mendasarinya adalah normal

- Peristiwa sebenarnya bukanlah hasil biner (misalnya, status kebangkrutan) tetapi proporsi (misalnya, proporsi populasi pada tingkat hutang yang berbeda).

Model Time Series

Model *Time Series* (Model deret waktu) digunakan untuk memprediksi atau meramalkan perilaku variabel di masa depan. Model ini memperhitungkan fakta bahwa titik data yang diambil dari waktu ke waktu mungkin memiliki struktur internal (seperti autokorelasi, tren, atau variasi musiman) yang harus diperhitungkan. Akibatnya, teknik regresi standar tidak dapat diterapkan pada data deret waktu dan metodologi telah dikembangkan untuk menguraikan komponen tren, musiman, dan siklus dari rangkaian tersebut. Memodelkan jalur dinamis variabel dapat meningkatkan perkiraan karena komponen rangkaian yang dapat diprediksi dapat diproyeksikan ke masa depan.

Model deret waktu memperkirakan persamaan perbedaan yang mengandung komponen stokastik. Dua bentuk yang umum digunakan dari model ini adalah model autoregressive (AR) dan model rata-rata bergerak (MA). Metodologi Box-Jenkins (1976) dikembangkan oleh George Box dan G.M. Jenkins menggabungkan model AR dan MA untuk menghasilkan model ARMA (autoregressive moving average) yang merupakan landasan analisis deret waktu stasioner. ARIMA (model rata-rata bergerak terintegrasi autoregressive) di sisi lain digunakan untuk menggambarkan deret waktu non-stasioner. Box dan Jenkins menyarankan untuk membedakan deret waktu non-stasioner untuk mendapatkan deret stasioner yang dapat digunakan model ARMA. Rangkaian waktu non-stasioner memiliki tren yang jelas dan tidak memiliki mean atau varians jangka panjang yang konstan.

Box dan Jenkins mengusulkan metodologi tiga tahap yang meliputi: identifikasi model, estimasi dan validasi. Tahap identifikasi melibatkan identifikasi apakah rangkaian tersebut diam atau tidak dan adanya kemusiman dengan memeriksa plot rangkaian, fungsi autokorelasi dan autokorelasi parsial. Pada tahap estimasi, model diestimasi menggunakan non linier time series atau prosedur estimasi kemungkinan maksimum. Terakhir, tahap validasi melibatkan pemeriksaan diagnostik seperti merencanakan residual untuk mendeteksi pencilan dan bukti kecocokan model.

Dalam beberapa tahun terakhir model deret waktu telah menjadi lebih canggih dan mencoba untuk memodelkan heteroskedastisitas bersyarat dengan model seperti ARCH (heteroskedastisitas bersyarat autoregresif) dan GARCH (heteroskedastisitas bersyarat autoregresif umum) yang sering digunakan untuk deret waktu keuangan. Selain itu

model time series juga digunakan untuk memahami keterkaitan antar variabel ekonomi yang direpresentasikan oleh sistem persamaan menggunakan model VAR (vector auto regression) dan struktural VAR.

Analisis Survival

Analisis survival, yang dikenal juga dengan nama *Time-to-Event Analysis* atau *Analysis of Failure Time Data*, adalah studi tentang model dan metode untuk menganalisis data tentang waktu kehidupan, waktu tunggu, atau secara umum waktu sampai dengan terjadinya suatu peristiwa spesifik. Teknik-teknik ini terutama dikembangkan dalam ilmu medis dan biologi, tetapi mereka juga banyak digunakan dalam ilmu sosial seperti ekonomi, serta dalam rekayasa (keandalan dan analisis waktu kegagalan).

Penyensoran dan non-normalitas yang merupakan karakteristik data survival menimbulkan kesulitan ketika mencoba menganalisis data dengan menggunakan model statistik konvensional seperti regresi linier berganda. Distribusi normal, sebagai distribusi simetris, mengambil nilai positif dan juga negatif, tetapi durasi pada dasarnya tidak boleh negatif dan oleh karena itu normalitas tidak dapat diasumsikan ketika berhadapan dengan data durasi / kelangsungan hidup. Oleh karena itu asumsi normalitas model regresi dilanggar. Asumsinya adalah jika data tidak disensor maka akan mewakili populasi yang diminati. Dalam analisis kelangsungan hidup, pengamatan yang disensor muncul setiap kali variabel dependen yang diminati mewakili waktu ke peristiwa terminal, dan durasi penelitian dibatasi dalam waktu.

Konsep penting dalam analisis kelangsungan hidup adalah tingkat bahaya, yang didefinisikan sebagai probabilitas bahwa peristiwa akan terjadi pada waktu t tergantung pada bertahan sampai waktu t. Konsep lain yang terkait dengan tingkat bahaya adalah fungsi kelangsungan hidup yang dapat didefinisikan sebagai probabilitas bertahan sampai waktu t.

Sebagian besar model mencoba memodelkan tingkat bahaya dengan memilih distribusi yang mendasari tergantung pada bentuk fungsi bahaya. Distribusi yang fungsi bahayanya miring ke atas dikatakan memiliki ketergantungan durasi positif, bahaya yang menurun menunjukkan ketergantungan durasi negatif sedangkan bahaya konstan adalah proses tanpa memori yang biasanya dicirikan oleh distribusi eksponensial. Beberapa pilihan distribusi dalam model survival adalah: F, gamma, Weibull, log normal, inverse normal, exponential, dll. Semua distribusi ini untuk variabel acak non-negatif.

Model durasi dapat berupa parametrik, non-parametrik atau semi-parametrik. Beberapa model yang umum digunakan adalah model hazard proporsional Kaplan-Meier dan Cox (non parametrik).

Pohon Klasifikasi dan Regresi (CART)

Classification and Regression Trees (CART) adalah salah satu metode atau algoritma dari teknik pohon keputusan. CART adalah suatu metode statistik nonparametrik yang dapat menggambarkan hubungan antara variabel respon (variabel dependen) dengan satu atau lebih variabel prediktor (variabel independen). Menurut Breiman dkk (1993), apabila variabel respon berbentuk kontinu maka metode yang digunakan adalah metode regresi pohon (*regression trees*), sedangkan apabila variabel respon memiliki skala kategorik maka metode yang digunakan adalah metode klasifikasi pohon (*classification trees*).

G*lobally-optimal classification tree analysis* (GO-CTA) (juga disebut analisis diskriminan optimal hierarkis) adalah generalisasi analisis diskriminan optimal yang dapat digunakan untuk mengidentifikasi model statistik yang memiliki akurasi maksimum untuk memprediksi nilai variabel dependen kategoris untuk dataset yang terdiri dari variabel kategori dan variabel kontinu. Keluaran HODA adalah pohon non-ortogonal yang menggabungkan variabel kategori dan titik potong untuk variabel kontinu yang menghasilkan akurasi prediksi maksimum, penilaian tingkat kesalahan Tipe I yang tepat, dan evaluasi potensi generalisasi silang dari model statistik. Analisis diskriminan optimal hierarkis dapat dianggap sebagai generalisasi analisis diskriminan linier Fisher. Analisis diskriminan optimal merupakan alternatif dari ANOVA (analisis varian) dan analisis regresi, yang mencoba untuk mengekspresikan satu variabel dependen sebagai kombinasi linier dari fitur atau pengukuran lain. Namun, ANOVA dan analisis regresi memberikan variabel dependen yaitu variabel numerik, sedangkan analisis diskriminan optimal hierarkis memberikan variabel dependen yaitu variabel kelas.

Pohon klasifikasi dan regresi (CART) adalah teknik pembelajaran pohon keputusan non-parametrik yang menghasilkan pohon klasifikasi atau regresi, bergantung pada apakah variabel dependen masing-masing bersifat kategorikal atau numerik. *Decision tress* (pohon keputusan) termasuk dalam model klasifikasi yang membagi data menjadi himpunan bagian berdasarkan kategori variabel masukan. Model ini dapat membantu kita bagaimana seseorang mengambil keputusan. Decision trees membagi data menjadi kelompok yang paling berbeda. Pohon keputusan dibentuk oleh kumpulan aturan berdasarkan variabel dalam kumpulan data pemodelan:

- Aturan berdasarkan nilai variabel dipilih untuk mendapatkan pemisahan terbaik untuk membedakan observasi berdasarkan variabel dependen
- Setelah aturan dipilih dan memisahkan node menjadi dua, proses yang sama diterapkan untuk masing-masing
- Node "anak" (yaitu prosedur rekursif)
- Pemisahan berhenti ketika CART mendeteksi tidak ada penguatan lebih lanjut yang dapat dibuat, atau beberapa aturan penghentian yang telah ditetapkan terpenuhi. (Alternatifnya, data dibagi sebanyak mungkin dan kemudian pohonnya dipangkas nanti.)

Setiap cabang pohon berakhir di simpul terminal. Setiap observasi jatuh ke dalam satu dan tepat satu node terminal, dan setiap node terminal secara unik ditentukan oleh seperangkat aturan.
Metode yang sangat populer untuk analitik prediktif adalah Random Forest yang diciptakan oleh Leo Breiman.

Multivariate Adaptive Regression Splines

Multivariate adaptive regression splines (MARS) adalah teknik non-parametrik yang membangun model fleksibel dengan menyesuaikan regresi linier sepotong-sepotong. Konsep penting yang terkait dengan splines regresi adalah sebuah simpul. Simpul adalah di mana satu model regresi lokal memberi jalan ke yang lain dan dengan demikian merupakan titik persimpangan antara dua splines.

Dalam splines regresi multivariat dan adaptif, fungsi basis adalah alat yang digunakan untuk menggeneralisasi pencarian knot. Fungsi dasar adalah sekumpulan fungsi yang digunakan untuk merepresentasikan informasi yang terkandung dalam satu atau lebih variabel. Model Splines Regresi Multivariasi dan Adaptif hampir selalu membuat fungsi basis berpasangan.

Pendekatan spline regresi multivariasi dan adaptif sengaja menutupi model dan kemudian memangkasnya untuk mendapatkan model yang optimal. Algoritme ini sangat intensif secara komputasi dan dalam praktiknya kami diminta untuk menentukan batas atas jumlah fungsi basis.

Teknik Machine Learning

Machine learning adalah cabang aplikasi dari Artificial Intelligence (Kecerdasan Buatan) yang focus pada pengembangan sebuah sistem yang mampu belajar "sendiri" tanpa harus berulang kali di program oleh manusia. Aplikasi Machine learning membutuhkan Data sebagai bahan belajar (training) sebelum mengeluarkan output. Aplikasi sejenis ini juga biasanya berada dalam domain spesifik alias tidak bisa diterapkan secara general untuk semua permasalahan.

Saat ini, karena mencakup sejumlah metode statistik canggih untuk regresi dan klasifikasi, ia menemukan aplikasi di berbagai bidang termasuk diagnostik medis, deteksi penipuan kartu kredit, pengenalan wajah dan ucapan, serta analisis pasar saham. Dalam aplikasi tertentu, cukup memprediksi variabel dependen secara langsung tanpa berfokus pada hubungan yang mendasarinya antar variabel. Dalam kasus lain, hubungan yang mendasarinya bisa sangat kompleks dan bentuk matematika dari ketergantungannya tidak diketahui. Untuk kasus seperti itu, teknik pembelajaran mesin meniru kognisi manusia dan belajar dari contoh pelatihan untuk memprediksi peristiwa di masa depan.

Diskusi singkat tentang beberapa metode yang biasa digunakan untuk analitik prediktif disediakan di bawah ini. Sebuah studi mendetail tentang pembelajaran mesin dapat ditemukan di Mitchell (1997).

Jaringan Saraf Tiruan (*Neural Networks*)

Jaringan saraf adalah teknik pemodelan nonlinier canggih yang mampu memodelkan fungsi yang kompleks. Mereka dapat diterapkan untuk masalah prediksi, klasifikasi atau kontrol dalam spektrum yang luas dari bidang-bidang seperti keuangan, psikologi kognitif / ilmu saraf, kedokteran, teknik, dan fisika. Jaringan saraf digunakan ketika sifat pasti hubungan antara input dan output tidak diketahui. Fitur utama dari jaringan saraf adalah mereka mempelajari hubungan antara input dan output melalui pelatihan. Ada tiga jenis pelatihan dalam jaringan saraf yang digunakan oleh jaringan yang berbeda, pelatihan yang diawasi dan tidak diawasi, pembelajaran penguatan, dengan diawasi menjadi yang paling umum.

Beberapa contoh teknik pelatihan jaringan saraf adalah backpropagation, quick propagation, conjugate gradient descent, projection operator, Delta-Bar-Delta, dll. Beberapa arsitektur jaringan tanpa pengawasan adalah perceptrons multilayer, Kohonen networks, Hopfield networks, dll.

Multilayer Perceptron (MLP)

Multilayer perceptron (MLP) terdiri dari input dan output layer dengan satu atau lebih lapisan tersembunyi dari node nonlinear-aktivasi atau node sigmoid. Ini ditentukan oleh vektor bobot dan perlu untuk menyesuaikan bobot jaringan. Propagasi mundur menggunakan penurunan gradien untuk meminimalkan kesalahan kuadrat antara nilai keluaran jaringan dan nilai yang diinginkan untuk keluaran tersebut. Bobot disesuaikan dengan proses iteratif dari atribut yang berulang. Perubahan kecil dalam bobot untuk mendapatkan nilai yang diinginkan dilakukan dengan proses yang disebut pelatihan jaring dan dilakukan oleh set pelatihan (aturan pembelajaran).

Radial Basis Functions

Jaringan syaraf tiruan radial basis function (JST RBF) dikenal sebagai salah satu jaringan syaraf yang memiliki tiga lapis bersifat feedforward yang dapat memecahkan masalah klasifikasi atau pengenalan pola. *Radial Basis Function*sl (RBF) adalah fungsi yang telah dibangun di dalamnya kriteria jarak sehubungan dengan pusat. Fungsi tersebut dapat digunakan dengan sangat efisien untuk interpolasi dan untuk memperlancar data.

RBF telah diterapkan di area jaringan saraf di mana mereka digunakan sebagai pengganti fungsi transfer sigmoidal. Jaringan tersebut memiliki 3 lapisan, lapisan masukan, lapisan tersembunyi dengan non-linieritas RBF dan lapisan keluaran linier. Pilihan paling populer untuk non-linearitas adalah Gaussian. Jaringan RBF memiliki keuntungan karena tidak dikunci ke dalam minimum lokal seperti halnya jaringan umpan-maju seperti perceptron multilayer.

Support Vector Machines

Support Vector Machines (SVM) digunakan untuk mendeteksi dan mengeksploitasi pola kompleks dalam data dengan mengelompokkan, mengklasifikasikan, dan memberi peringkat data. Mereka adalah mesin pembelajaran yang digunakan untuk melakukan klasifikasi biner dan estimasi regresi. Mereka biasanya menggunakan metode berbasis kernel untuk menerapkan teknik klasifikasi linier pada masalah klasifikasi non-linier. Ada beberapa tipe SVM seperti linear, polynomial, sigmoid dll.

Naïve Bayes

Naïve Bayes berdasarkan aturan probabilitas bersyarat Bayes digunakan untuk melakukan tugas klasifikasi. Naïve Bayes mengasumsikan prediktor independen secara statistik yang menjadikannya alat klasifikasi efektif yang mudah ditafsirkan. Ini paling baik digunakan

ketika dihadapkan pada masalah 'kutukan dimensi' yaitu ketika jumlah prediktor sangat tinggi.

K-nearest neighbors

K-nearest neighbors atau KNN adalah algoritma yang berfungsi untuk melakukan klasifikasi suatu data berdasarkan data pembelajaran (train data sets), yang diambil dari k tetangga terdekatnya (nearest neighbors). Dengan k merupakan banyaknya tetangga terdekat. Algoritma *K-nearest neighbors* (KNN) milik kelas metode statistik pengenalan pola. Metode ini tidak memaksakan apriori asumsi tentang distribusi dari mana sampel pemodelan diambil. Ini melibatkan serangkaian pelatihan dengan nilai-nilai positif dan negatif. Sampel baru diklasifikasikan dengan menghitung jarak ke kasus pelatihan tetangga terdekat. *K-nearest neighbor*s melakukan klasifikasi dengan proyeksi data pembelajaran pada ruang berdimensi banyak. Ruang ini dibagi menjadi bagian-bagian yang merepresentasikan kriteria data pembelajaran. Setiap data pembelajaran direpresentasikan menjadi titik-titik c pada ruang dimensi banyak.

Tanda titik tersebut akan menentukan klasifikasi sampel. Dalam pengklasifikasi k-terdekat tetangga, titik k terdekat dipertimbangkan dan tanda mayoritas digunakan untuk mengklasifikasikan sampel. Kinerja algoritma kNN dipengaruhi oleh tiga faktor utama: (1) ukuran jarak yang digunakan untuk mencari tetangga terdekat; (2) aturan keputusan yang digunakan untuk mendapatkan klasifikasi dari k-terdekat tetangga; dan (3) jumlah tetangga yang digunakan untuk mengklasifikasikan sampel baru. Hal ini dapat dibuktikan bahwa, tidak seperti metode lain, metode ini secara universal konvergen asimtotik, yaitu: sebagai ukuran set pelatihan meningkat, jika pengamatan independen dan terdistribusi identik (iid), terlepas dari distribusi dari mana sampel diambil , kelas yang diprediksi akan bertemu dengan tugas kelas yang meminimalkan kesalahan klasifikasi.

Analisis Pemodelan Prediktif Geospasial

Secara konseptual, pemodelan prediktif geospasial berakar pada prinsip bahwa kemunculan peristiwa yang dimodelkan terbatas dalam distribusi. Kemunculan peristiwa tidak seragam atau acak dalam distribusi — ada faktor lingkungan spasial (infrastruktur, sosiokultural, topografi, dll.) Yang membatasi dan mempengaruhi di mana lokasi peristiwa terjadi. Pemodelan prediktif geospasial berupaya untuk menggambarkan kendala dan pengaruh tersebut dengan menghubungkan kejadian lokasi geospasial historis secara spasial dengan faktor lingkungan yang mewakili kendala dan pengaruh tersebut. Pemodelan

prediktif geospasial adalah proses untuk menganalisis peristiwa melalui filter geografis untuk membuat pernyataan kemungkinan terjadinya atau kemunculan peristiwa.

Alat Pendukung

Secara historis, menggunakan alat analitik prediktif — serta memahami hasil yang mereka berikan — membutuhkan keterampilan tingkat lanjut. Namun, alat analitik prediktif modern tidak lagi terbatas pada spesialis TI. Karena semakin banyak organisasi mengadopsi analitik prediktif ke dalam proses pengambilan keputusan dan mengintegrasikannya ke dalam operasi mereka, mereka menciptakan pergeseran pasar menuju pengguna bisnis sebagai konsumen utama informasi. Pengguna bisnis menginginkan alat yang dapat mereka gunakan sendiri. Vendor merespons dengan membuat perangkat lunak baru yang menghilangkan kerumitan matematika, menyediakan antarmuka grafis yang ramah pengguna dan / atau membuat jalan pintas yang dapat, misalnya, mengenali jenis data yang tersedia dan menyarankan model prediksi yang sesuai. Alat analitik prediktif telah menjadi cukup canggih untuk menyajikan dan membedah masalah data secara memadai, sehingga setiap pekerja informasi yang paham data dapat menggunakannya untuk menganalisis data dan mengambil hasil yang bermakna dan berguna. Misalnya, alat modern menyajikan temuan dengan menggunakan bagan, grafik, dan skor sederhana yang menunjukkan kemungkinan hasil yang mungkin.

Ada banyak alat yang tersedia di pasar yang membantu pelaksanaan analitik prediktif. Mulai dari yang membutuhkan kecanggihan pengguna yang sangat sedikit hingga yang dirancang untuk praktisi ahli. Perbedaan antara alat-alat ini sering kali terletak pada tingkat penyesuaian dan pengangkatan data berat yang diperbolehkan.

Alat analisis prediktif *open source* yang terkenal meliputi:

- Apache Mahout
- GNU Octave
- KNIME
- OpenNN
- Orange
- R
- scikit-learn
- Weka

Alat analisis prediktif komersial yang terkenal meliputi:

- Alpine Data Labs
- Alteryx
- Angoss KnowledgeSTUDIO
- BIRT Analytics
- IBM SPSS Statistics and IBM SPSS Modeler
- KXEN Modeler
- Mathematica
- MATLAB
- Minitab
- LabVIEW
- Neural Designer
- Oracle Advanced Analytics
- Pervasive
- Predixion Software
- RapidMiner
- RCASE
- Revolution Analytics
- SAP HANA and SAP BusinessObjects Predictive Analytics
- SAS and SAS Enterprise Miner
- STATA
- Statgraphics
- STATISTICA
- TeleRetail
- TIBCO

Selain paket perangkat lunak ini, alat khusus juga telah dikembangkan untuk aplikasi industri. Misalnya, Watchdog Agent Toolbox telah dikembangkan dan dioptimalkan untuk analisis prediktif dalam aplikasi prognostik dan manajemen kesehatan dan tersedia untuk MATLAB dan LabVIEW.

Paket perangkat lunak analitik prediktif komersial paling populer menurut Survei Analisis Rexer untuk 2013 adalah IBM SPSS Modeler, SAS Enterprise Miner, dan Dell Statistica.

Kritikan Lebih Lanjut

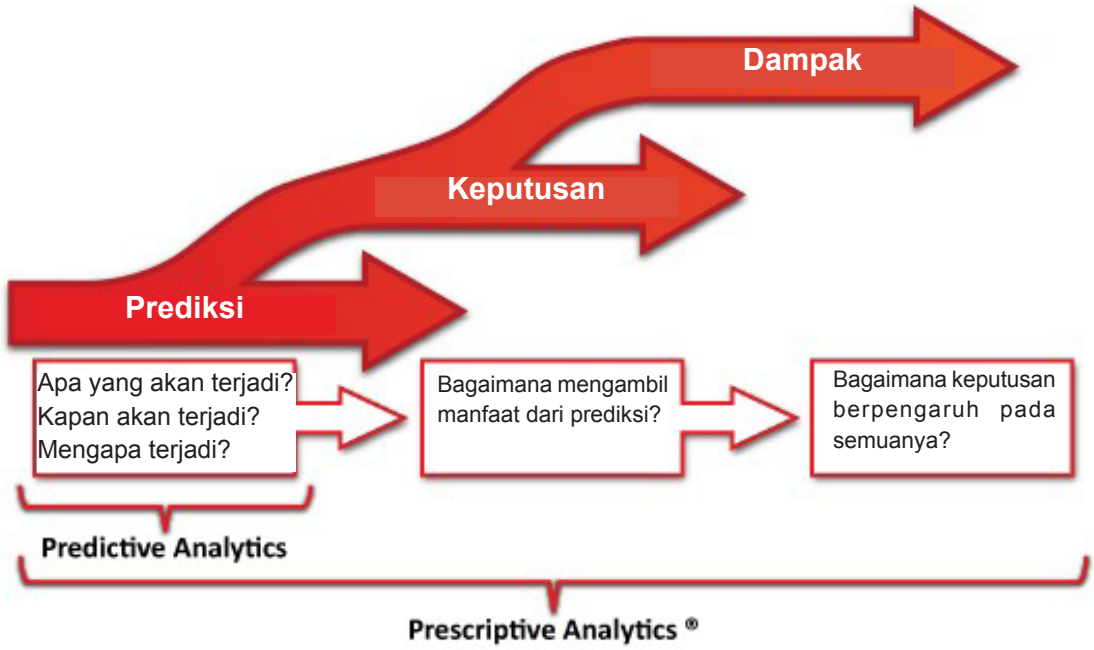
Ada banyak orang yang meragukan kemampuan komputer dan algoritme untuk memprediksi masa depan, termasuk Gary King, seorang profesor dari Universitas Harvard dan direktur *Institute for Quantitative Social Science*. Orang-orang dipengaruhi oleh lingkungannya dengan cara yang tidak terhitung banyaknya. Mencoba memahami apa yang akan dilakukan orang selanjutnya mengasumsikan bahwa semua variabel yang berpengaruh dapat diketahui dan diukur secara akurat. “Lingkungan orang berubah lebih cepat daripada yang mereka lakukan sendiri. Segala sesuatu mulai dari cuaca hingga hubungan mereka dengan ibu mereka dapat mengubah cara orang berpikir dan bertindak. Semua variabel tersebut tidak dapat diprediksi. Bagaimana mereka akan berdampak pada seseorang bahkan lebih sulit diprediksi. Jika ditempatkan dalam situasi yang sama persis besok, mereka mungkin membuat keputusan yang sama sekali berbeda. Artinya, prediksi statistik hanya valid dalam kondisi laboratorium yang steril, yang tiba-tiba tidak berguna seperti sebelumnya. "

2.8 Analisis Preskriptif

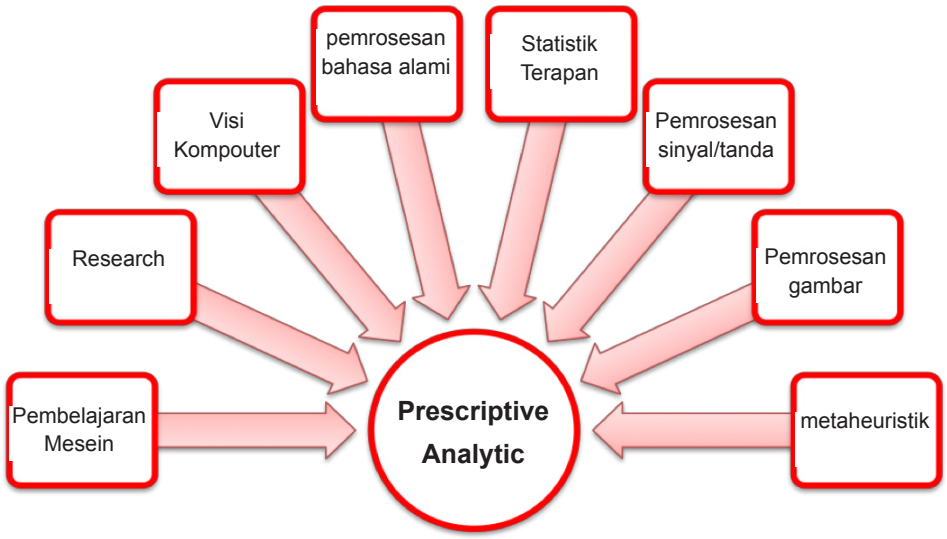
Analisis preskriptif adalah tahap ketiga dan terakhir dari analitik (BA) yang juga mencakup analitik deskriptif dan prediktif. Disebut sebagai "batas akhir kemampuan analitik," analitik preskriptif memerlukan penerapan ilmu matematika dan komputasi menyarankan pilihan keputusan untuk memanfaatkan hasil analitik deskriptif dan prediktif. Tahap pertama dari analitik bisnis adalah analitik deskriptif, yang masih menyumbang sebagian besar dari semua analitik bisnis saat ini. Analisis deskriptif melihat kinerja masa lalu dan memahami kinerja itu dengan menambang data historis untuk mencari alasan di balik keberhasilan atau kegagalan masa lalu. Sebagian besar pelaporan manajemen - seperti penjualan, pemasaran, operasi, dan keuangan - menggunakan jenis analisis post-mortem.

Fase selanjutnya adalah analitik prediktif. Analisis prediktif menjawab pertanyaan apa yang mungkin terjadi. Ini adalah saat data historis digabungkan dengan aturan, algoritme, dan terkadang data eksternal untuk menentukan kemungkinan hasil di masa mendatang dari suatu peristiwa atau kemungkinan terjadinya situasi. Fase terakhir adalah analitik preskriptif, yang melampaui prediksi hasil di masa depan dengan juga menyarankan tindakan untuk mendapatkan keuntungan dari prediksi dan menunjukkan implikasi dari setiap opsi keputusan. Analisis preskriptif tidak hanya mengantisipasi apa yang akan terjadi dan kapan itu akan terjadi, tetapi juga mengapa itu akan terjadi. Selanjutnya, analitik preskriptif menyarankan opsi keputusan tentang bagaimana memanfaatkan peluang masa

depan atau mengurangi risiko di masa depan dan menunjukkan implikasi dari setiap opsi keputusan. Analisis preskriptif dapat terus mengambil data baru untuk memprediksi ulang dan meresepkan ulang, sehingga secara otomatis meningkatkan akurasi prediksi dan menentukan pilihan keputusan yang lebih baik.



Analisis Preskriptif *melampaui analitik prediktif dengan menentukan tindakan yang diperlukan untuk mencapai hasil yang diprediksi, dan efek yang saling terkait dari setiap keputusan*

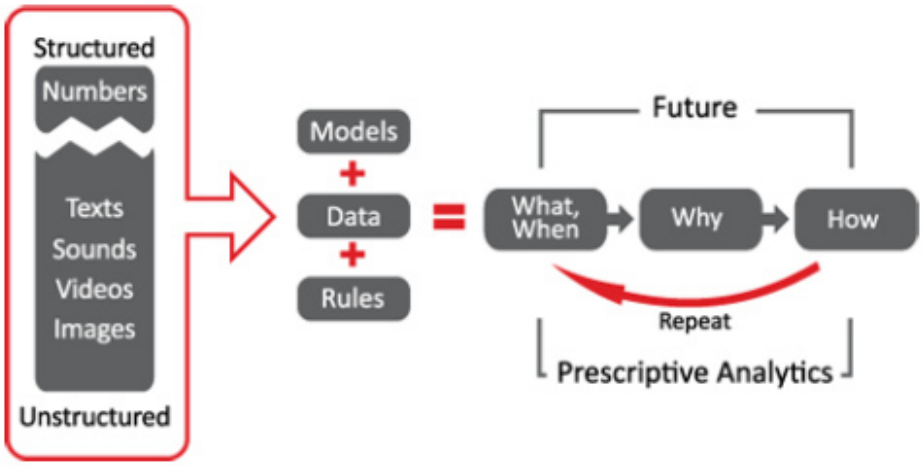


Disiplin ilmiah *yang termasuk dalam Analisis Preskriptif*

Analisis preskriptif menyerap data hibrid, kombinasi dari data terstruktur (angka, kategori) dan tidak terstruktur (video, gambar, suara, teks), dan aturan bisnis untuk memprediksi apa yang akan terjadi di depan dan untuk menentukan cara memanfaatkan masa depan yang diprediksi ini tanpa mengorbankan yang lain. prioritas. Ketiga fase analitik dapat dilakukan melalui layanan atau teknologi profesional atau kombinasi keduanya. Untuk menskalakan, teknologi analitik preskriptif perlu adaptif untuk memperhitungkan peningkatan volume, kecepatan, dan variasi data yang mungkin dihasilkan oleh sebagian besar proses kritis misi dan lingkungannya. Salah satu kritik terhadap analitik preskriptif adalah bahwa perbedaannya dari analitik prediktif tidak terdefinisi dengan baik dan karenanya tidak dipahami.

Sejarah

Sementara istilah Analisis Preskriptif, yang pertama kali diciptakan oleh IBM dan kemudian diberi merek dagang oleh Ayata, konsep yang mendasarinya telah ada selama ratusan tahun. Teknologi di balik analitik preskriptif secara sinergis menggabungkan data hybrid, aturan bisnis dengan model matematika dan model komputasi. Input data ke analitik preskriptif dapat berasal dari berbagai sumber: internal, seperti di dalam perusahaan; dan eksternal, juga dikenal sebagai data lingkungan. Data dapat terstruktur, yang meliputi angka dan kategori, serta data tidak terstruktur, seperti teks, gambar, suara, dan video. Data tidak terstruktur berbeda dari data terstruktur karena formatnya sangat bervariasi dan tidak dapat disimpan dalam database relasional tradisional tanpa upaya signifikan pada transformasi data. Lebih dari 80% data dunia saat ini tidak terstruktur, menurut IBM.



Analisis Preskriptif *menggabungkan data terstruktur dan tidak terstruktur, serta menggunakan* kombinasi teknik *analitik tingkat lanjut dan disiplin ilmu untuk memprediksi,* meresepkan, dan *beradaptasi.*

Selain variasi jenis data dan volume data yang terus bertambah, data yang masuk juga dapat berkembang sehubungan dengan kecepatan, yaitu, lebih banyak data yang dihasilkan dengan kecepatan yang lebih cepat atau variabel. Aturan bisnis menentukan proses bisnis dan mencakup batasan tujuan, preferensi, kebijakan, praktik terbaik, dan batasan. Model matematika dan model komputasi adalah teknik yang berasal dari ilmu matematika, ilmu komputer, dan disiplin terkait seperti statistik terapan, pembelajaran mesin, penelitian operasi, pemrosesan bahasa alami, visi komputer, pengenalan pola, pemrosesan gambar, pengenalan ucapan, dan pemrosesan sinyal. Penerapan yang benar dari semua metode ini dan verifikasi hasil mereka menyiratkan kebutuhan akan sumber daya dalam skala besar termasuk manusia, komputasi dan temporal untuk setiap proyek Analitik Preskriptif. Untuk menghemat pengeluaran puluhan orang, mesin berkinerja tinggi dan kerja berminggu-minggu seseorang harus mempertimbangkan pengurangan sumber daya dan oleh karena itu pengurangan akurasi atau keandalan hasilnya. Rute yang disukai adalah pengurangan yang menghasilkan hasil probabilistik dalam batas yang dapat diterima.

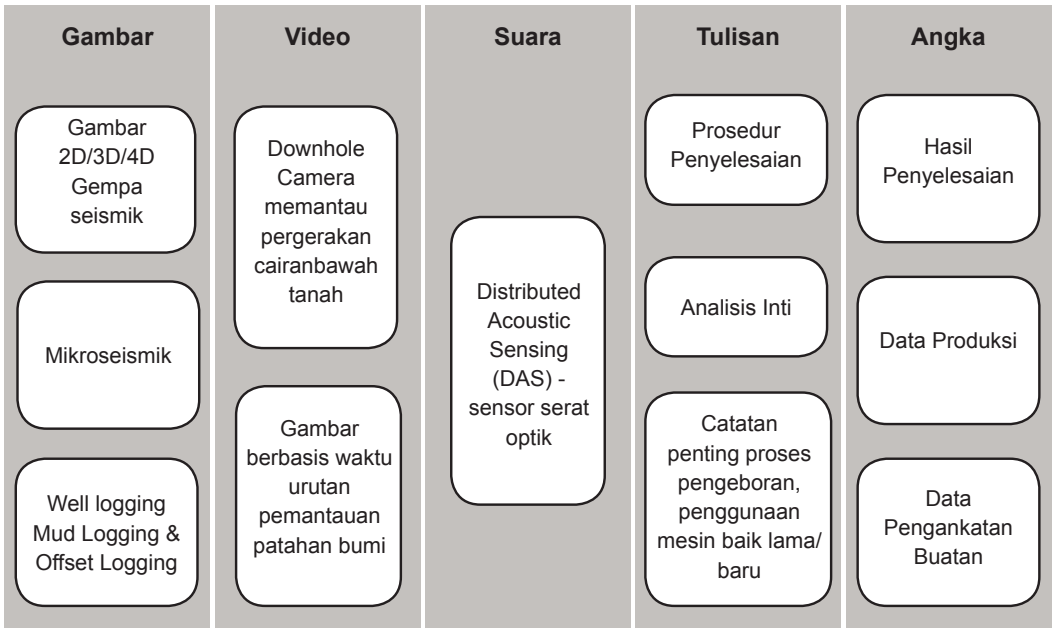
Aplikasi dalam Minyak dan Gas

Perencanaan	• Variabel reservoir, pengeboran, penyelesaian, dan produksi mana yang memiliki pengaruh terbesar pada produksi? • Seberapa dekat sebaiknya jarak antar sumur (minyak)? Apakah kita memiliki tahap yang saling tumpang tindih? Apakah jarak antar sumur terkendali? • Apakah urutan penanganan dan / atau produksi sumur yang berdekatan penting? Mengapa?
Produksi	• Tahapan dan cluster mana yang diperlakukan secara efektif? Diperlakukan seperti yang diharapkan? Mengapa? • Tahapan mana yang menghasilkan? Memproduksi seperti yang diharapkan? Mana yang bukan? Mengapa? • Bagaimana seharusnya sebuah sumur diproduksi untuk memaksimalkan nilai masa pakainya?
Pemulihan Sekunder EOR	• Kapan pengangkatan buatan harus diperkenalkan dalam siklus hidup sumur untuk memaksimalkan EOR (Enhanced Oil Recovery)? Apakah EOR menghasilkan tingkat pemulihan yang lebih tinggi atau apakah pemulihan hanya dipercepat? • Berapa ROI inkremental dari EOR? Di mana gunanya hasil yang semakin berkurang?

Pertanyaan Kunci *Jawaban perangkat lunak Analisis Preskriptif untuk produsen minyak dan gas*

Energi adalah industri terbesar di dunia (dengan estimasi $ 6 triliun). Proses dan keputusan terkait eksplorasi, pengembangan, dan produksi minyak dan gas alam menghasilkan data dalam jumlah besar. Banyak jenis data yang ditangkap digunakan untuk membuat model dan gambar struktur dan lapisan bumi 5.000 - 35.000 kaki di bawah permukaan dan untuk menggambarkan aktivitas di sekitar sumur itu sendiri, seperti karakteristik pengendapan, kinerja mesin, laju aliran minyak, suhu dan tekanan reservoir. . Perangkat lunak analitik preskriptif dapat membantu menemukan dan memproduksi hidrokarbon dengan mengambil data seismik, data log sumur, data produksi, dan kumpulan data terkait lainnya untuk menentukan resep spesifik tentang bagaimana dan di mana mengebor, menyelesaikan, dan memproduksi sumur untuk mengoptimalkan pemulihan , meminimalkan biaya, dan mengurangi jejak lingkungan.

Pengembangan Sumber Daya Tidak Konvensional



Contoh kumpulan *data terstruktur dan tidak terstruktur yang dihasilkan dan oleh perusahaan* minyak dan *gas serta ekosistem penyedia layanannya yang dapat dianalisis bersama* \ menggunakan perangkat *lunak Analisis Deskriptif*

Dengan nilai produk akhir yang ditentukan oleh ekonomi komoditas global, dasar persaingan bagi operator di E&P hulu adalah kemampuan untuk menyebarkan modal secara efektif untuk mencari dan mengekstrak sumber daya secara lebih efisien, efektif, dapat diprediksi, dan aman daripada rekan-rekan mereka. Dalam permainan sumber daya yang tidak konvensional, efisiensi dan efektivitas operasional berkurang oleh inkonsistensi

reservoir, dan pengambilan keputusan terganggu oleh tingkat ketidakpastian yang tinggi. Tantangan ini memanifestasikan dirinya dalam bentuk faktor pemulihan yang rendah dan variasi kinerja yang luas.

Perangkat lunak Preskriptif Analytics dapat secara akurat memprediksi produksi dan menentukan konfigurasi optimal variabel pengeboran, penyelesaian, dan produksi yang dapat dikontrol dengan memodelkan banyak variabel internal dan eksternal secara bersamaan, terlepas dari sumber, struktur, ukuran, atau formatnya. Perangkat lunak analitik preskriptif juga dapat memberikan opsi keputusan dan menunjukkan dampak dari setiap opsi keputusan sehingga manajer operasi dapat secara proaktif mengambil tindakan yang sesuai, tepat waktu, untuk menjamin kinerja eksplorasi dan produksi di masa depan, dan memaksimalkan nilai ekonomi aset di setiap titik selama kursus. dari masa pakai mereka..

Perawatan Peralatan Ladang Minyak

Dalam bidang perawatan peralatan ladang minyak, Analisis Preskriptif dapat mengoptimalkan konfigurasi, mengantisipasi dan mencegah waktu henti yang tidak direncanakan, mengoptimalkan penjadwalan lapangan, dan meningkatkan perencanaan perawatan. Menurut General Electric, ada lebih dari 130.000 pompa submersible elektrik (ESP) yang dipasang di seluruh dunia, terhitung 60% dari produksi minyak dunia. Analisis Preskriptif telah diterapkan untuk memprediksi kapan dan mengapa ESP akan gagal, dan merekomendasikan tindakan yang diperlukan untuk mencegah kegagalan tersebut.

Di bidang Kesehatan, Keselamatan, dan Lingkungan, analitik preskriptif dapat memprediksi dan mencegah insiden yang dapat menyebabkan kerugian reputasi dan finansial bagi perusahaan minyak dan gas.

Penetapan Harga

Penetapan harga adalah area fokus lainnya. Harga gas alam berfluktuasi secara dramatis tergantung pada pasokan, permintaan, ekonometrika, geopolitik, dan kondisi cuaca. Produsen gas, perusahaan transmisi pipa, dan perusahaan utilitas memiliki minat yang besar untuk memprediksi harga gas secara lebih akurat sehingga mereka dapat mengunci persyaratan yang menguntungkan sambil melindungi nilai risiko penurunan. Perangkat lunak analitik preskriptif dapat secara akurat memprediksi harga dengan memodelkan variabel internal dan eksternal secara bersamaan dan juga memberikan opsi keputusan dan menunjukkan dampak dari setiap opsi keputusan.

Aplikasi dalam Kesehatan

Banyak faktor yang mendorong penyedia layanan kesehatan untuk secara dramatis meningkatkan proses bisnis dan operasi saat industri perawatan kesehatan Amerika Serikat memulai migrasi yang diperlukan dari sistem berbasis volume layanan berbayar ke sistem berbasis nilai dan biaya untuk kinerja. Analisis preskriptif memainkan peran kunci untuk membantu meningkatkan kinerja di sejumlah area yang melibatkan berbagai pemangku kepentingan: pembayar, penyedia, dan perusahaan farmasi.

Analisis preskriptif dapat membantu penyedia meningkatkan efektivitas pemberian perawatan klinis mereka kepada populasi yang mereka kelola dan dalam prosesnya mencapai kepuasan dan retensi pasien yang lebih baik. Penyedia layanan dapat melakukan manajemen kesehatan populasi yang lebih baik dengan mengidentifikasi model intervensi yang sesuai untuk populasi bertingkat risiko yang menggabungkan data dari tahap-tahap perawatan di fasilitas dan *telehealth* berbasis rumah.

Analitik preskriptif juga dapat menguntungkan penyedia layanan kesehatan dalam perencanaan kapasitas mereka dengan menggunakan analitik untuk memanfaatkan data operasional dan penggunaan yang dikombinasikan dengan data faktor eksternal seperti data ekonomi, tren demografis populasi dan tren kesehatan populasi, untuk merencanakan investasi modal masa depan yang lebih akurat seperti yang baru. pemanfaatan fasilitas dan peralatan serta memahami trade-off antara menambahkan tempat tidur tambahan dan memperluas fasilitas yang ada versus membangun yang baru.

Analisis preskriptif dapat membantu perusahaan farmasi untuk mempercepat pengembangan obat mereka dengan mengidentifikasi kelompok pasien yang paling sesuai untuk uji klinis di seluruh dunia - pasien yang diharapkan patuh dan tidak akan keluar dari uji coba karena komplikasi. Analytics dapat memberi tahu perusahaan berapa banyak waktu dan uang yang dapat mereka hemat jika mereka memilih satu kelompok pasien di negara tertentu vs. yang lain.

Dalam negosiasi penyedia-pembayar, penyedia dapat meningkatkan posisi negosiasi mereka dengan perusahaan asuransi kesehatan dengan mengembangkan pemahaman yang kuat tentang pemanfaatan layanan di masa depan. Dengan memprediksi pemanfaatan secara akurat, penyedia juga dapat mengalokasikan personel dengan lebih baik.

2.9 Analisis Media Sosial

Analisis Media Sosial sebagai bagian dari analisis sosial adalah proses pengumpulan data dari percakapan pemangku kepentingan di media digital dan diproses menjadi wawasan terstruktur yang mengarah pada keputusan bisnis yang lebih didorong oleh informasi dan peningkatan sentralitas pelanggan untuk merek dan bisnis.

Analisis media sosial juga dapat disebut sebagai mendengarkan media sosial, pemantauan media sosial atau kecerdasan media sosial. Sumber media digital untuk analitik media sosial meliputi saluran media sosial, blog, forum, situs berbagi gambar, situs berbagi video, agregator, iklan baris, keluhan, Q&A, ulasan, Wikipedia, dan lain-lain.

Analisis media sosial adalah praktik agnostik industri dan biasanya digunakan dalam berbagai pendekatan pada keputusan bisnis, pemasaran, layanan pelanggan, manajemen reputasi, penjualan, dan lainnya. Ada berbagai alat yang menawarkan analisis media sosial, bervariasi dari tingkat kebutuhan bisnis. Logika di balik algoritme yang dirancang untuk alat ini adalah seleksi, pemrosesan awal data, transformasi, penambangan, dan evaluasi pola tersembunyi.

Agar proses analisis media sosial yang lengkap berhasil, penting untuk menentukan indikator kinerja utama (KPI) untuk mengevaluasi data secara objektif. Analisis media sosial penting ketika seseorang perlu memahami pola yang tersembunyi dalam sejumlah besar data sosial yang terkait dengan merek tertentu. Homofili digunakan sebagai bagian dari analitik, ini adalah kecenderungan bahwa kontak antara orang yang mirip terjadi pada tingkat yang lebih tinggi daripada di antara orang yang berbeda. Menurut penelitian, dua pengguna yang mengikuti secara timbal balik berbagi minat topik dengan menambang ribuan tautan mereka. Semua ini digunakan untuk mengambil keputusan bisnis besar di sektor media sosial.

Keberhasilan alat pemantauan media sosial (SMM) dapat bervariasi dari satu perusahaan ke perusahaan lainnya. Menurut Soleman dan Cohard (2016), selain faktor teknis yang terkait dengan pemantauan media sosial (SMM) (kualitas sumber, fungsionalitas, kualitas alat), organisasi juga harus mempertimbangkan kebutuhan akan kapabilitas baru, manusia, manajerial, dan organisasi. keterampilan untuk memanfaatkan alat SMM mereka.

2.10 Analisis Perilaku

Analisis perilaku (*Behavior Analytics*) adalah kemajuan terbaru dalam analisis bisnis yang mengungkapkan wawasan baru tentang perilaku konsumen di platform eCommerce, game online, aplikasi web dan seluler, dan IoT. Peningkatan cepat dalam volume data peristiwa mentah yang dihasilkan oleh dunia digital memungkinkan metode yang melampaui analisis khas oleh demografi dan metrik tradisional lainnya yang memberi tahu kita orang seperti apa yang mengambil tindakan di masa lalu. Analisis perilaku berfokus pada pemahaman tentang bagaimana konsumen bertindak dan mengapa, memungkinkan prediksi yang akurat tentang bagaimana mereka cenderung bertindak di masa depan. Ini memungkinkan pemasar untuk membuat penawaran yang tepat ke segmen konsumen yang tepat pada waktu yang tepat.

Analisis perilaku menggunakan volume besar data peristiwa pengguna mentah yang diambil selama sesi saat konsumen menggunakan aplikasi, game, atau situs web, termasuk data lalu lintas seperti jalur navigasi, klik, interaksi media sosial, keputusan pembelian, dan daya respons pemasaran. Selain itu, data peristiwa dapat menyertakan metrik iklan seperti waktu klik menuju konversi, serta perbandingan antara metrik lain seperti nilai uang pesanan dan jumlah waktu yang dihabiskan di situs. Poin data ini kemudian dikumpulkan dan dianalisis, baik dengan melihat perkembangan sesi dari saat pengguna pertama kali memasuki platform hingga penjualan dilakukan, atau produk lain yang dibeli atau dilihat pengguna sebelum pembelian ini. Analisis perilaku memungkinkan tindakan dan tren di masa depan untuk diprediksi berdasarkan pengumpulan data tersebut.

Sementara analisis bisnis memiliki fokus yang lebih luas pada siapa, apa, di mana dan kapan intelijen bisnis, analisis perilaku mempersempit ruang lingkup itu, memungkinkan seseorang untuk mengambil poin data yang tampaknya tidak terkait untuk mengekstrapolasi, memprediksi dan menentukan kesalahan dan tren masa depan. Dibutuhkan pandangan data yang lebih holistik dan manusiawi, menghubungkan poin data individu untuk memberi tahu kita tidak hanya apa yang terjadi, tetapi juga bagaimana dan mengapa hal itu terjadi.

Contoh dan *Real World Applications*

Data menunjukkan bahwa sebagian besar pengguna yang menggunakan platform eCommerce tertentu menemukannya dengan menelusuri "makanan Thailand" di Google. Setelah membuka halaman beranda, kebanyakan orang menghabiskan waktu di halaman

"Makanan Asia" dan kemudian keluar tanpa melakukan pemesanan. Melihat setiap peristiwa ini sebagai titik data terpisah tidak mewakili apa yang sebenarnya terjadi dan mengapa orang tidak melakukan pembelian. Namun, melihat titik data ini sebagai representasi dari perilaku pengguna secara keseluruhan memungkinkan seseorang untuk menginterpolasi bagaimana dan mengapa pengguna bertindak dalam kasus khusus ini.

Representasi Visual *Peristiwa yang memuat Analisis Perilaku*

Analisis perilaku melihat semua lalu lintas situs dan tampilan halaman sebagai garis waktu peristiwa terhubung yang tidak menghasilkan pesanan. Karena sebagian besar pengguna keluar setelah melihat halaman "Makanan Asia", mungkin ada keterputusan antara apa yang mereka cari di Google dan apa yang ditampilkan halaman "Makanan Asia". Mengetahui hal ini, sekilas melihat halaman "Makanan Asia" mengungkapkan bahwa itu tidak menampilkan makanan Thailand secara mencolok dan oleh karena itu orang-orang tidak mengira itu benar-benar ditawarkan, meskipun demikian.

Analisis perilaku menjadi semakin populer di lingkungan yang bersifat komersial. Amazon.com adalah pemimpin dalam menggunakan analisis perilaku untuk merekomendasikan produk tambahan yang kemungkinan besar dibeli pelanggan berdasarkan pola pembelian mereka sebelumnya di situs. Analisis perilaku juga digunakan oleh Target untuk menyarankan produk kepada pelanggan di toko ritel mereka, sementara kampanye politik menggunakannya untuk menentukan bagaimana calon pemilih harus didekati. Selain aplikasi ritel dan politik, analisis perilaku juga digunakan oleh bank dan perusahaan manufaktur untuk memprioritaskan prospek yang dihasilkan oleh situs web mereka. Analisis perilaku juga memungkinkan pengembang untuk mengelola pengguna dalam permainan online dan aplikasi web.

Jenis Aplikasi BA

- E-niaga dan ritel - Rekomendasi produk dan memprediksi tren penjualan di masa depan
- Game online - Memprediksi tren penggunaan, beban, dan preferensi pengguna dalam rilis mendatang
- Pengembangan aplikasi - Menentukan bagaimana pengguna menggunakan aplikasi untuk memprediksi penggunaan dan preferensi di masa mendatang.
- Analisis kelompok/*cohort analysis* - Mengelompokkan pengguna ke dalam kelompok serupa untuk mendapatkan pemahaman yang lebih terfokus tentang perilaku mereka.
- Keamanan - Mendeteksi kredensial yang disusupi dan ancaman orang dalam dengan menemukan perilaku yang tidak wajar.
- Saran/*suggestions* - Orang yang menyukai ini juga menyukai ...
- Penyajian konten yang relevan berdasarkan perilaku pengguna.

Komponen Analisis Perilaku

Solusi analitik perilaku yang ideal akan mencakup:

- Pengambilan data peristiwa mentah dalam jumlah besar secara real-time di semua perangkat digital yang relevan dan aplikasi yang digunakan selama sesi
- Agregasi otomatis data peristiwa mentah menjadi kumpulan data yang relevan untuk akses cepat, pemfilteran, dan analisis
- Kemampuan untuk menanyakan data dengan cara yang tidak terbatas, memungkinkan pengguna untuk mengajukan pertanyaan bisnis apa pun
- Pustaka yang luas dari fungsi analisis bawaan seperti analisis kelompok, jalur, dan corong
- Komponen visualisasi

Jenis-Jenis Analisis Perilaku

Analisis Jalur (*Path Analysis*)

Analisis jalur, adalah analisis jalur, yang menggambarkan rangkaian peristiwa berturut-turut yang dilakukan oleh pengguna atau kelompok tertentu selama periode waktu tertentu saat menggunakan situs web, game online, atau platform eCommerce. Sebagai bagian dari analisis perilaku, analisis jalur adalah cara untuk memahami perilaku pengguna untuk mendapatkan wawasan yang dapat ditindaklanjuti ke dalam data. Analisis jalur memberikan

gambaran visual dari setiap peristiwa yang dilakukan pengguna atau kelompok sebagai bagian dari jalur selama jangka waktu tertentu.

Meskipun memungkinkan untuk melacak jalur pengguna melalui situs, dan bahkan menunjukkan jalur tersebut sebagai representasi visual, pertanyaan sebenarnya adalah bagaimana mendapatkan wawasan yang dapat ditindaklanjuti ini. Jika analisis jalur hanya mengeluarkan grafik "cantik", meskipun mungkin terlihat bagus, ini tidak memberikan sesuatu yang konkret untuk ditindaklanjuti.

Studi Kasus

Untuk mendapatkan hasil maksimal dari analisis jalur, langkah pertama adalah menentukan apa yang perlu dianalisis dan apa tujuan analisis. Sebuah perusahaan mungkin mencoba mencari tahu mengapa situs mereka berjalan lambat, apakah jenis pengguna tertentu tertarik dengan halaman atau produk tertentu, atau apakah antarmuka pengguna mereka disiapkan dengan cara yang logis.

Sekarang tujuan telah ditetapkan, ada beberapa cara untuk melakukan analisis. Jika sebagian besar dari kelompok tertentu, orang-orang yang berusia antara 18-25 tahun, masuk ke game online, membuat profil, dan kemudian menghabiskan 10 menit berikutnya untuk menjelajahi halaman menu, mungkin antarmuka penggunanya tidak logis. Dengan melihat sekelompok pengguna ini mengikuti jalur yang mereka lakukan, pengembang akan dapat menganalisis data dan menyadari bahwa setelah membuat profil, tombol "mainkan game" tidak muncul. Dengan demikian, analisis jalur dapat memberikan data yang dapat ditindaklanjuti bagi perusahaan untuk bertindak dan memperbaiki kesalahan.

Di eCommerce, analisis jalur dapat membantu menyesuaikan pengalaman berbelanja untuk setiap pengguna. Dengan melihat produk apa yang dilihat oleh pelanggan lain dalam kelompok tertentu sebelum membeli satu, perusahaan dapat menyarankan "barang yang mungkin juga Anda sukai" kepada pelanggan berikutnya dan meningkatkan peluang mereka melakukan pembelian. Selain itu, analisis jalur dapat membantu memecahkan masalah kinerja pada platform. Misalnya, sebuah perusahaan melihat sebuah jalur dan menyadari bahwa situs mereka berhenti beroperasi setelah kombinasi peristiwa tertentu. Dengan menganalisis jalur dan perkembangan peristiwa yang menyebabkan kesalahan, perusahaan dapat menentukan kesalahan dan memperbaikinya.

Perkembangan Terkini

Secara historis, analisis jalur termasuk dalam kategori luas analisis situs web, dan hanya terkait dengan analisis jalur melalui situs web. Analisis jalur dalam analisis situs web adalah proses menentukan urutan halaman yang dikunjungi dalam sesi pengunjung sebelum beberapa peristiwa yang diinginkan, seperti pengunjung membeli item atau meminta buletin. Urutan tepat halaman yang dikunjungi mungkin atau mungkin tidak penting dan mungkin atau mungkin tidak ditentukan. Dalam praktiknya, analisis ini dilakukan secara agregat, memberi peringkat jalur (urutan halaman) yang dikunjungi sebelum acara yang diinginkan, dengan menurunkan frekuensi penggunaan. Idenya adalah untuk menentukan fitur situs web apa yang mendorong hasil yang diinginkan. "Analisis kejatuhan", bagian dari analisis jalur, melihat "lubang hitam" di situs, atau jalur yang paling sering mengarah ke jalan buntu, jalur atau fitur yang membingungkan atau kehilangan calon pelanggan.

Dengan munculnya data besar bersama dengan aplikasi berbasis web, game online, dan platform eCommerce, analisis jalur telah mencakup lebih dari sekadar analisis jalur web. Memahami bagaimana pengguna bergerak melalui aplikasi, game, atau platform web lainnya adalah bagian dari analisis jalur modern.

Memahami Pengunjung Media

Di dunia nyata ketika Anda mengunjungi toko, rak dan produk tidak ditempatkan secara acak. Pemilik toko dengan hati-hati menganalisis pengunjung dan jalur yang mereka lalui melalui toko, terutama saat mereka memilih atau membeli produk. Selanjutnya pemilik toko akan menyusun ulang rak dan produk untuk mengoptimalkan penjualan dengan meletakkan semuanya dalam urutan yang paling logis bagi pengunjung. Di supermarket, hal ini biasanya menghasilkan rak anggur di samping berbagai kue, keripik, kacang, dll. Hanya karena orang minum anggur dan makan kacang dengannya.

Di sebagian besar situs web, ada logika yang sama yang dapat diterapkan. Pengunjung yang memiliki pertanyaan tentang suatu produk akan masuk ke bagian informasi produk atau dukungan di situs web. Dari sana mereka membuat langkah logis ke halaman pertanyaan yang sering diajukan jika mereka memiliki pertanyaan tertentu. Pemilik situs web juga ingin menganalisis perilaku pengunjung. Misalnya, jika situs web menawarkan produk untuk dijual, pemilik ingin mengonversi sebanyak mungkin pengunjung menjadi pembelian yang telah diselesaikan. Jika ada formulir pendaftaran dengan banyak halaman, pemilik situs web ingin memandu pengunjung ke halaman pendaftaran terakhir.

Analisis jalur menjawab pertanyaan umum seperti:

- Ke mana tujuan sebagian besar pengunjung setelah mereka memasuki beranda saya?
- Apakah ada hubungan pengunjung yang kuat antara produk A dan produk B di situs web saya ?. Pertanyaan yang tidak dapat dijawab oleh klik halaman dan statistik pengunjung unik.

Corong dan Alur Sasaran

Google Analytics menyediakan fungsi jalur dengan Corong dan Alur Sasaran (*Funnels and Goals*). Jalur halaman situs web yang telah ditentukan sebelumnya ditentukan dan setiap pengunjung yang berjalan di jalur tersebut adalah sebuah sasaran (*a goal*). Pendekatan ini sangat membantu saat menganalisis berapa banyak pengunjung mencapai halaman tujuan tertentu, yang disebut analisis titik akhir.

Menggunakan Maps

Jalur yang dilalui pengunjung di situs web dapat mengarah ke jalur unik yang tak terbatas. Akibatnya, tidak ada gunanya menganalisis setiap jalur, tetapi mencari jalur terkuat. Jalur terkuat ini biasanya ditampilkan dalam peta grafis atau dalam teks seperti: Halaman A -> Halaman B -> Halaman D -> Keluar.

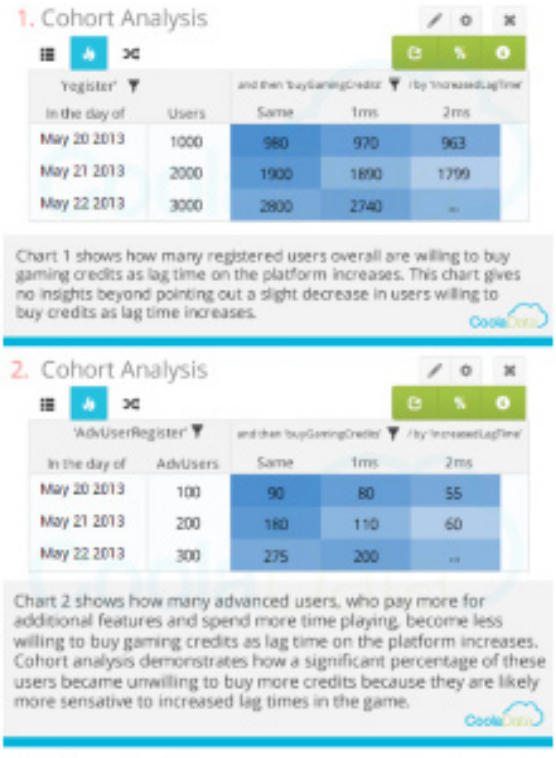
Analisis Kelompok (Cohort Analysis)

Analisis kelompok adalah bagian dari analisis perilaku yang mengambil data dari kumpulan data tertentu (misalnya, platform eCommerce, aplikasi web, atau game online) dan alih-alih melihat semua pengguna sebagai satu unit, ia memecahnya menjadi kelompok terkait untuk dianalisis. Kelompok atau kelompok terkait ini, biasanya berbagi karakteristik atau pengalaman yang sama dalam rentang waktu yang ditentukan. Analisis kelompok memungkinkan perusahaan untuk "melihat pola dengan jelas di seluruh siklus hidup pelanggan (atau pengguna), daripada memotong semua pelanggan secara membabi buta tanpa memperhitungkan siklus alami yang dialami pelanggan." Dengan melihat pola waktu ini, perusahaan dapat menyesuaikan dan menyesuaikan layanannya dengan kelompok tertentu tersebut. Meskipun analisis kohort terkadang dikaitkan dengan studi kohort, mereka berbeda dan tidak boleh dipandang sebagai satu dan sama. Analisis kelompok telah datang untuk mendeskripsikan secara spesifik analisis kelompok yang berkaitan dengan data besar dan analisis bisnis, sementara studi kelompok adalah istilah umum yang lebih umum yang menggambarkan jenis studi di mana data dipecah menjadi beberapa kelompok serupa.

Contoh Kasus

Tujuan dari alat analitik bisnis adalah untuk menganalisis dan menyajikan informasi yang dapat ditindaklanjuti. Agar perusahaan dapat bertindak berdasarkan informasi tersebut, informasi tersebut harus relevan dengan situasi yang dihadapi. Basis data yang penuh dengan ribuan atau bahkan jutaan entri dari semua data pengguna membuatnya sulit untuk mendapatkan data yang dapat ditindaklanjuti, karena data tersebut mencakup banyak kategori dan periode waktu yang berbeda. Analisis kelompok yang dapat ditindaklanjuti memungkinkan kemampuan untuk menelusuri ke pengguna dari setiap kelompok tertentu untuk mendapatkan pemahaman yang lebih baik tentang perilaku mereka, seperti jika pengguna check out, dan berapa banyak mereka membayar. Dalam analisis kohor, "setiap grup baru [kohor] memberikan kesempatan untuk memulai dengan sekumpulan pengguna baru", memungkinkan perusahaan untuk melihat hanya data yang relevan dengan kueri saat ini dan menindaklanjutinya.

Dalam e-Commerce, perusahaan mungkin hanya tertarik pada pelanggan yang mendaftar dalam dua minggu terakhir dan melakukan pembelian, yang merupakan contoh kelompok tertentu. Pengembang perangkat lunak mungkin hanya peduli tentang data dari pengguna yang mendaftar setelah peningkatan tertentu, atau yang menggunakan fitur tertentu dari platform.



Contoh analisis kelompok pemain game di platform tertentu: Pemain ahli, kelompok 1, akan lebih peduli tentang fitur-fitur canggih dan waktu jeda dibandingkan dengan pendaftaran baru, kelompok 2. Dengan dua kelompok ini ditentukan, dan analisis dijalankan, permainan perusahaan akan disajikan dengan representasi visual dari data yang spesifik untuk dua kelompok. Kemudian dapat dilihat bahwa sedikit kelambatan dalam waktu muat telah menyebabkan hilangnya pendapatan yang signifikan dari pemain tingkat lanjut, sementara pendaftar baru bahkan tidak menyadari adanya kelambatan tersebut. Seandainya perusahaan hanya melihat laporan pendapatan keseluruhannya untuk semua pelanggan, perusahaan tidak akan dapat melihat perbedaan antara kedua kelompok ini. Analisis kelompok memungkinkan perusahaan untuk mengetahui pola dan tren dan membuat perubahan yang diperlukan untuk membuat pemain yang sudah mahir dan pemain baru senang.

Analisis Kelompok yang Dapat Ditindaklanjuti

“Metrik yang dapat ditindaklanjuti adalah metrik yang menghubungkan tindakan spesifik dan berulang dengan hasil yang diamati [seperti pendaftaran pengguna, atau pembayaran]. Kebalikan dari metrik yang dapat ditindaklanjuti adalah metrik yang sia-sia (seperti klik web atau jumlah unduhan) yang hanya berfungsi untuk mendokumentasikan keadaan produk saat ini tetapi tidak menawarkan wawasan tentang bagaimana kami sampai di sini atau apa yang harus dilakukan selanjutnya. ”

Tanpa analitik yang dapat ditindaklanjuti, informasi yang disajikan mungkin tidak memiliki aplikasi praktis apa pun, karena satu-satunya poin data mewakili metrik kesombongan yang tidak diterjemahkan ke dalam hasil tertentu. Meskipun berguna bagi perusahaan untuk mengetahui berapa banyak orang yang ada di situs mereka, metrik itu sendiri tidak berguna. Agar dapat ditindaklanjuti, perlu menghubungkan "tindakan berulang untuk [sebuah] hasil yang diamati".

Melakukan Analisis Kelompok

Untuk melakukan analisis kelompok yang tepat, ada empat tahapan utama:

- Tentukan pertanyaan apa yang ingin Anda jawab. Inti dari analisis ini adalah untuk menghasilkan informasi yang dapat ditindaklanjuti untuk bertindak guna meningkatkan bisnis, produk, pengalaman pengguna, omset, dll. Untuk memastikan hal itu terjadi, pertanyaan yang tepat harus diajukan. Dalam contoh game di atas, perusahaan tidak yakin mengapa mereka kehilangan pendapatan karena jeda waktu meningkat, meskipun pengguna masih mendaftar dan bermain game.
- Tentukan metrik yang dapat membantu Anda menjawab pertanyaan. Analisis kelompok yang tepat membutuhkan identifikasi peristiwa, seperti pengguna yang check-out, dan properti tertentu, seperti berapa banyak pengguna membayar. Contoh permainan mengukur kesediaan pelanggan untuk membeli kredit permainan berdasarkan berapa lama jeda waktu yang ada di situs.
- Tentukan kelompok spesifik yang relevan. Dalam membuat kelompok, seseorang harus menganalisis semua pengguna dan menargetkan mereka atau melakukan kontribusi atribut untuk menemukan perbedaan yang relevan di antara mereka masing-masing, pada akhirnya untuk menemukan dan menjelaskan perilaku mereka sebagai kelompok tertentu. Contoh di atas membagi pengguna menjadi pengguna "dasar" dan "lanjutan" karena setiap grup berbeda dalam tindakan, sensitivitas struktur harga, dan tingkat penggunaan.
- Lakukan analisis kohort. Analisis di atas dilakukan dengan menggunakan visualisasi data yang memungkinkan perusahaan game menyadari bahwa pendapatan mereka turun karena pengguna tingkat lanjut yang membayar lebih tinggi tidak menggunakan sistem karena jeda waktu meningkat. Karena pengguna tingkat lanjut merupakan bagian besar dari pendapatan perusahaan, pendaftaran pengguna dasar tambahan tidak menutupi kerugian finansial karena kehilangan pengguna tingkat lanjut. Untuk mengatasinya, perusahaan meningkatkan waktu jeda mereka dan mulai melayani lebih banyak pengguna tingkat lanjut.

Bab 3 : Tinjauan tentang Data Mining

Proses memahami pola yang ditemukan dalam kumpulan *Big data* dikenal sebagai *data mining* (penambangan data). Beberapa aspek *data mining* yang telah dijelaskan pada bagian berikut ini adalah pembelajaran aturan asosiasi, analisis cluster, analisis regresi, peringkasan otomatis dan contoh data mining. Bab tentang *data mining* menawarkan fokus yang mendalam, dengan mengingat materi pelajaran yang kompleks.

3.1 Pengertian Data Mining

Banyak sekali definisi mengenai apa itu data mining. *Data mining* merupakan sebuah 'tool' yang memungkinkan para pengguna untuk mengakses data dengan jumlah yang besar dengan cepat. Pengertian yang lebih khusus dari data mining, yaitu suatu alat dan aplikasi menggunakan analisis statistik pada data. *Data mining* adalah suatu proses ekstraksi atau penggalian data dan informasi yang besar, yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting.

Data Mining adalah subbidang interdisipliner ilmu komputer. Ini adalah proses komputasi untuk menemukan pola dalam kumpulan data besar yang melibatkan metode gabungan antara kecerdasan buatan, pembelajaran mesin, statistik, dan sistem database. Tujuan keseluruhan dari proses *data mining* adalah untuk mengekstrak informasi dari kumpulan data dan mengubahnya menjadi struktur yang dapat dimengerti untuk digunakan lebih lanjut. Selain tahap analisis mentah, ini melibatkan aspek database dan manajemen data, pra-pemrosesan data, pertimbangan model dan inferensi, metrik ketertarikan, pertimbangan kompleksitas, pasca-pemrosesan struktur yang ditemukan, visualisasi, dan pembaruan online. *Data Mining* adalah langkah analisis dari proses "penemuan pengetahuan dalam database", atau "*knowledge discovery in databases*" / KDD.

Istilah ini keliru, karena tujuannya adalah ekstraksi pola dan pengetahuan dari sejumlah besar data, bukan ekstraksi (mining) data itu sendiri. Ini juga merupakan kata kunci dan sering diterapkan pada segala bentuk pemrosesan data atau informasi skala besar (pengumpulan, ekstraksi, pergudangan, analisis, dan statistik) serta aplikasi sistem pendukung keputusan komputer, termasuk kecerdasan buatan, pembelajaran mesin, dan intelijen bisnis.

Tugas *Data Mining* yang sebenarnya adalah analisis otomatis atau semi-otomatis dari sejumlah besar data untuk mengekstrak pola yang sebelumnya tidak diketahui dan menarik seperti kelompok catatan data (analisis kluster), catatan yang tidak biasa (deteksi anomali), dan ketergantungan (penambangan aturan asosiasi) . Ini biasanya melibatkan penggunaan teknik database seperti indeks spasial. Pola ini kemudian dapat dilihat sebagai semacam ringkasan data masukan, dan dapat digunakan dalam analisis lebih lanjut atau, misalnya, dalam pembelajaran mesin dan analitik prediktif. Misalnya, langkah *Data Mining* mungkin mengidentifikasi beberapa grup dalam data, yang kemudian dapat digunakan untuk mendapatkan hasil prediksi yang lebih akurat oleh sistem pendukung keputusan. Baik pengumpulan data, persiapan data, maupun interpretasi dan pelaporan hasil bukanlah bagian dari langkah penggalian data, tetapi termasuk dalam keseluruhan proses KDD sebagai langkah tambahan. Istilah terkait pengerukan data, penangkapan data, dan pengintaian data mengacu pada penggunaan metode *Data Mining* untuk mengambil sampel bagian dari kumpulan data populasi yang lebih besar yang (atau mungkin) terlalu kecil untuk kesimpulan statistik yang dapat diandalkan tentang validitas pola ditemukan. Metode ini, bagaimanapun, dapat digunakan dalam membuat hipotesis baru untuk diuji terhadap populasi data yang lebih besar.

Berikut ini beberapa definisi data mining dari beberapa sumber (Larose, 2005):

- *Data mining* adalah proses menemukan sesuatu yang bermakna dari suatu korelasi baru, pola dan tren yang ada dengan cara memilah-milah data berukuran besar yang disimpan dalam repositori, menggunakan teknologi pengenalan pola serta teknik matematika dan statistik.
- *Data mining* adalah analisis pengamatan database untuk menemukan hubungan yang tidak terduga dan untuk meringkas data dengan cara atau metode baru yang dapat dimengerti dan bermanfaat kepada pemilik data.
- *Data mining*merupakan bidang ilmu interdisipliner yang menyatukan teknik pembelajaran dari mesin (machine learning), pengenalan pola (pattern recognition), statistik, database, dan visualisasi untuk mengatasi masalah ekstraksi informasi dari basis data yang besar.
- *Data mining* diartikan sebagai suatu proses ekstraksi informasi berguna dan potensial dari sekumpulan data yang terdapat secara implisit dalam suatu basis data

Pengertian secara Etimologi

Pada tahun 1960-an, ahli statistik menggunakan istilah seperti "*Data Fishing*" atau "*Data Dredging*" untuk merujuk pada apa yang mereka anggap sebagai praktik buruk dalam menganalisis data tanpa hipotesis a-priori. Istilah "*Data Mining*" muncul sekitar tahun 1990 dalam komunitas database. Untuk waktu yang singkat di tahun 1980-an, frase "*Databae Mining*", digunakan, tetapi karena itu adalah merek dagang oleh HNC, sebuah perusahaan yang berbasis di San Diego, untuk memasang *Database Mining Workstation* mereka; akibatnya peneliti beralih ke "*Data Mining*". Istilah lain yang digunakan termasuk Data *Archaeology, Information Harvesting, Information Discovery, Knowledge Extraction,* dll. Dalam bahasa Indonesia digunakan istilah Penggalian Data atau Penambangan Data. Gregory Piatetsky-Shapiro menciptakan istilah "*Knowledge Discovery in Database*" untuk lokakarya pertama dengan topik yang sama (KDD-1989) dan istilah ini menjadi lebih populer di AI dan Komunitas Pembelajaran Mesin (*Machine Learning*). Namun, istilah *data mining* menjadi lebih populer di komunitas bisnis dan pers.

Pada awal tahun 2000-an, perusahaan web mulai melihat kekuatan dari data mining, dan praktiknya benar-benar berjalan bahkan sampai sekarang ini. Sementara frase “data mining” sejak saat itu telah dikalahkan oleh kata kunci lain seperti “data analytics atau analisis data” “big data” dan “machine learning” prosesnya tetap menjadi bagian integral dari praktik bisnis. Bahkan, wajar untuk mengatakan bahwa data mining telah menjadi bagian de facto dari menjalankan bisnis modern seperti di tahun 2020 sekarang ini.

Sejarah Data Mining

Pada 1990-an, istilah "Data Mining" diperkenalkan, tetapi data mining adalah evolusi dari sektor dengan sejarah yang luas. Teknik awal untuk mengidentifikasi pola dalam data termasuk teorema Bayes (1700-an), dan evolusi regresi (1800-an). Generasi dan kekuatan ilmu komputer yang terus berkembang telah meningkatkan pengumpulan, penyimpanan, dan manipulasi data karena kumpulan data memiliki ukuran dan tingkat kompleksitas yang luas. Investigasi data langsung secara eksplisit telah ditingkatkan secara progresif dengan pemrosesan data tidak langsung dan otomatis, dan penemuan ilmu komputer lainnya seperti jaringan saraf, pengelompokan, algoritme genetika (1950-an), pohon keputusan (1960-an), dan mesin vektor pendukung (1990-an).

Asal data mining ditelusuri kembali ke tiga garis keluarga: Statistik klasik, Kecerdasan buatan, dan Pembelajaran mesin.

Statistik klasik:

Statistik adalah dasar dari sebagian besar teknologi dimana data mining dibangun, seperti analisis regresi, deviasi standar, distribusi standar, varian standar, analisis diskriminatif, analisis cluster, dan interval trust. Semua ini digunakan untuk menganalisis data dan koneksi data.

Kecerdasan buatan:

AI atau Artificial intelligence didasarkan pada heuristik dan bukan statistik. Ia mencoba menerapkan pemikiran manusia seperti pemrosesan ke masalah statistik. Konsep AI tertentu diadopsi oleh beberapa produk komersial kelas atas, seperti modul pengoptimalan kueri untuk sistem *Relational Database Management System* (RDBMS).

Pembelajaran mesin:

Pembelajaran mesin atau disebut sebagai *Machine Learning* adalah kombinasi dari statistik dan AI. *Machine Learning* dapat dianggap sebagai evolusi AI karena menggabungkan heuristik AI dengan analisis statistik yang kompleks. Pembelajaran mesin mencoba untuk memungkinkan program komputer mengetahui tentang data yang mereka pelajari sehingga program membuat keputusan yang berbeda berdasarkan karakteristik data yang diperiksa. Pembelajaran mesin menggunakan statistik untuk konsep dasar dan menambahkan lebih banyak heuristik dan algoritma AI untuk mencapai targetnya.

Proses dan tahapan Data Mining

Knowledge Discovery in Databases (KDD) umumnya didefinisikan dengan tahapan:
(1) Seleksi
(2) Pra-pemrosesan
(3) Transformasi
(4) Penambangan Data
(5) Interpretasi / Evaluasi.

Pada tahun 1996, sekelompok perusahaan yang termasuk teradata dan NCR memimpin proyek untuk menstandarisasi dan memformalkan metodologi penambangan data. Pekerjaan mereka menghasilkan proses standar lintas-industri untuk penambangan data atau *Cross-Industry Standard Process for Data Mining* (CRISP-DM).
Standar terbuka ini memecah proses data mining menjadi 6 (enam) fase:

1. Business Understanding (pemahaman bisnis)
2. Data Understanding (pemahaman data)
3. Data Preparation (persiapan data)
4. Modelling (pemodelan)

5. Evaluation (evaluasi) dan;
6. Deployment (penyebaran)

atau proses yang disederhanakan seperti (1) pra-pemrosesan, (2) penggalian data, dan (3) validasi hasil.

Pra-pemrosesan

Sebelum algoritma data mining dapat digunakan, kumpulan data target harus dirakit. Karena data mining hanya dapat mengungkap pola yang benar-benar ada dalam data, kumpulan data target harus cukup besar untuk memuat pola-pola ini sambil tetap cukup ringkas untuk ditambang dalam batas waktu yang dapat diterima. Sumber data yang umum adalah data mart atau data warehouse. Pra-pemrosesan penting untuk menganalisis kumpulan data multivariat sebelum penambangan data. Set target kemudian dibersihkan. Pembersihan data menghilangkan pengamatan yang mengandung kebisingan dan data yang hilang.

Penambangan Data

Penambangan data melibatkan enam kelas tugas umum:

a. Deteksi anomali (Deteksi outlier / perubahan / deviasi) - Identifikasi yang tidak biasa catatan data, yang mungkin menarik atau kesalahan data yang memerlukan penyelidikan lebih lanjut.
b. Pembelajaran aturan asosiasi (Model ketergantungan) - Mencari hubungan antar variabel. Misalnya, supermarket mungkin mengumpulkan data tentang kebiasaan pembelian pelanggan. Dengan menggunakan pembelajaran aturan asosiasi, supermarket dapat menentukan produk mana yang sering dibeli bersama dan menggunakan informasi ini untuk tujuan pemasaran. Ini kadang-kadang disebut sebagai analisis keranjang pasar.
c. *Clustering* - adalah tugas untuk menemukan grup dan struktur dalam data yang dalam beberapa cara atau “serupa” lainnya, tanpa menggunakan struktur yang diketahui dalam data.
d. *Clasification* - adalah tugas menggeneralisasi struktur yang diketahui untuk diterapkan ke data baru. Misalnya, program email mungkin mencoba untuk mengklasifikasikan email sebagai "sah" atau sebagai "spam".
e. Regresi - mencoba menemukan fungsi yang memodelkan data dengan kesalahan paling kecil. Regresi atau *regression* adalah alat statistik yang lebih maju yang umum dalam analitik prediktif. Ini dapat membantu pengembang media sosial dan aplikasi

smartphone meningkatkan keterlibatan atau yang biasanya lebih dikenal dengan istilah *engagement.*

f. Peringkasan (*Summarization*) - memberikan representasi yang lebih ringkas dari kumpulan data, termasuk visualisasi dan pembuatan laporan. Sebagai contoh misalnya, Anda dapat menggunakan ringkasan untuk membuat grafik atau menghitung rata-rata dari set data yang diberikan. *Summarization* adalah salah satu bentuk penambangan data yang paling akrab dan dapat dengan mudah digunakan.

Validasi Hasil

Data mining dapat disalahgunakan secara tidak sengaja, dan kemudian dapat memberikan hasil yang tampak signifikan; tetapi yang tidak benar-benar memprediksi perilaku masa depan dan tidak dapat direproduksi pada sampel data baru dan tidak banyak digunakan. Seringkali ini hasil dari penyelidikan terlalu banyak hipotesis dan tidak melakukan pengujian hipotesis statistik yang tepat. Versi sederhana dari masalah ini dalam pembelajaran mesin dikenal sebagai overfitting, tetapi masalah yang sama dapat muncul pada fase proses yang berbeda dan dengan demikian pemisahan latihan / pengujian - jika berlaku sama sekali - mungkin tidak cukup untuk mencegah hal ini terjadi.

Langkah terakhir dari penemuan pengetahuan dari data adalah untuk memverifikasi bahwa pola yang dihasilkan oleh algoritma data mining terjadi pada kumpulan data yang lebih luas. Tidak semua pola yang ditemukan oleh algoritma data mining valid. Biasanya algoritme data mining menemukan pola dalam set pelatihan yang tidak ada dalam set data umum. Ini disebut overfitting. Untuk mengatasi hal ini, evaluasi menggunakan serangkaian data uji yang tidak dilatihkan algoritme data mining. Pola yang dipelajari diterapkan ke set pengujian ini, dan keluaran yang dihasilkan dibandingkan dengan keluaran yang diinginkan. Misalnya, algoritme penambangan data yang mencoba membedakan "spam" dari email "sah" akan dilatih pada sekumpulan pelatihan contoh email. Setelah dilatih, pola yang dipelajari akan diterapkan ke set pengujian email yang belum pernah dilatih sebelumnya. Keakuratan pola kemudian dapat diukur dari berapa banyak email yang mereka klasifikasikan dengan benar. Sejumlah metode statistik dapat digunakan untuk mengevaluasi algoritme, seperti kurva ROC.

Jika pola yang dipelajari tidak memenuhi standar yang diinginkan, maka perlu dilakukan evaluasi ulang dan perubahan tahap pra-pemrosesan dan data mining. Jika pola yang dipelajari memenuhi standar yang diinginkan, maka langkah terakhir adalah menafsirkan pola yang dipelajari dan mengubahnya menjadi pengetahuan.

Standarisasi Data Mining

Ada beberapa upaya untuk menentukan standar untuk proses data mining, misalnya Standar Proses Lintas Industri Eropa 1999 untuk Data Mining (CRISP-DM 1.0) dan standar Java Data Mining 2004 (JDM 1.0). Pengembangan penerus proses ini (CRISP-DM 2.0 dan JDM 2.0) aktif pada tahun 2006, tetapi terhenti sejak itu. JDM 2.0 ditarik tanpa mencapai draf akhir.

Untuk menukar model yang diekstrak - khususnya untuk digunakan dalam analitik prediktif - standar kuncinya adalah Predictive Model Markup Language (PMML), yang merupakan bahasa berbasis XML yang dikembangkan oleh Data Mining Group (DMG) dan didukung sebagai format pertukaran oleh banyak orang. aplikasi data mining. Seperti namanya, ini hanya mencakup model prediksi, tugas penambangan data tertentu yang sangat penting untuk aplikasi bisnis. Namun, ekstensi untuk menutupi (misalnya) pengelompokan subruang telah diusulkan secara independen dari DMG.

Penerapan Data Mining

Dalam bidang apasaja data mining dapat diterapkan? Berikut beberapa contoh bidang penerapan data mining:

Telekomunikasi

Sebuah perusahaan telekomunikasi menerapkan data mining untuk melihat dari jutaan transaksi yang masuk, transaksi mana sajakah yang masih harus ditangani secara manual.

Keuangan

Financial Crimes Enforcement Network di Amerika Serikat baru-baru ini menggunakan data mining untuk menambang triliyunan dari berbagai subyek seperti property, rekening bank dan transaksi keuangan lainnya untuk mendeteksi transaksi-transaksi keuangan yang mencurigakan (Seperti money laundry).

Asuransi

Australian Health Insurance Commision menggunakan data mining untuk mengidentifikasi layanan lesehatan yang sebenarnya tidak perlu tetapi tetap dilakukan oleh peserta asuransi.

Olahraga

IBM Advanced Scout menggunakan data mining untuk menganalisis statistik permainan NBA (jumlah shots blocked, assists dan fouls) dalam rangka mencapai keunggulan bersaing (competitive advantage) untuk tim New York Knicks dan Miami Heat.

Masalah Etika dan Privasi

Meskipun istilah "penambangan data" itu sendiri tidak memiliki implikasi etis, istilah ini sering dikaitkan dengan penambangan informasi dalam kaitannya dengan perilaku masyarakat (etis dan lainnya).

Cara-cara data mining dapat digunakan dalam beberapa kasus dan konteks menimbulkan pertanyaan tentang privasi, legalitas, dan etika. Secara khusus, data mining pemerintah atau komersial untuk tujuan keamanan nasional atau penegakan hukum, seperti dalam *Total Information Awareness Program* atau di ADVISE, telah mengangkat masalah privasi.

Penambangan data membutuhkan persiapan data yang dapat mengungkap informasi atau pola yang dapat membahayakan kerahasiaan dan kewajiban privasi. Cara umum agar hal ini terjadi adalah melalui agregasi data. Agregasi data melibatkan penggabungan data bersama-sama (mungkin dari berbagai sumber) dengan cara yang memfasilitasi analisis (tetapi itu juga dapat membuat identifikasi pribadi, data tingkat individu dapat disimpulkan atau terlihat jelas). Ini bukan data mining semata, tetapi hasil dari persiapan data sebelum - dan untuk tujuan - analisis. Ancaman terhadap privasi individu mulai berlaku ketika data, setelah dikumpulkan, menyebabkan penambang data, atau siapa pun yang memiliki akses ke kumpulan data yang baru dikompilasi, dapat mengidentifikasi individu tertentu, terutama ketika datanya awalnya anonim.

Direkomendasikan agar seseorang mengetahui hal-hal berikut sebelum data dikumpulkan:

- tujuan pengumpulan data dan setiap proyek data mining (yang diketahui);
- bagaimana data akan digunakan;
- siapa yang akan bisa menambang data dan menggunakan data dan turunannya;
- status keamanan seputar akses ke data;
- bagaimana data yang dikumpulkan dapat diperbarui.

Data juga dapat dimodifikasi sehingga menjadi anonim, sehingga individu tidak dapat langsung diidentifikasi. Namun, bahkan kumpulan data yang "tidak diidentifikasi" / "dianonimkan" berpotensi berisi cukup informasi untuk memungkinkan identifikasi individu, seperti yang terjadi ketika jurnalis dapat menemukan beberapa individu berdasarkan kumpulan riwayat pencarian yang secara tidak sengaja dirilis oleh AOL.

Pengungkapan informasi yang dapat diidentifikasi secara tidak sengaja yang mengarah ke penyedia melanggar Praktik Informasi yang Adil. Kecerobohan ini dapat menyebabkan kerugian finansial, emosional, atau fisik bagi individu yang ditunjukkan. Dalam satu contoh pelanggaran privasi, pelanggan Walgreens mengajukan gugatan terhadap perusahaan pada tahun 2011 karena menjual informasi resep ke perusahaan penambangan data yang pada gilirannya memberikan data tersebut kepada perusahaan farmasi.

Aturan dan Undang-Undang Perlindungan Privasi di Eropa

Eropa memiliki undang-undang privasi yang cukup kuat, dan upaya sedang dilakukan untuk lebih memperkuat hak-hak konsumen. Namun, U.S.-E.U. Prinsip Safe Harbor saat ini secara efektif mengekspos pengguna Eropa ke eksploitasi privasi oleh perusahaan AS. Sebagai konsekuensi dari pengungkapan pengawasan Global Edward Snowden, telah terjadi peningkatan diskusi untuk mencabut perjanjian ini, karena khususnya data akan diungkapkan sepenuhnya ke Badan Keamanan Nasional, dan upaya untuk mencapai kesepakatan telah gagal.

Aturan dan Undang-Undang Perlindungan Privasi di Amerika Serikat

Di Amerika Serikat, masalah privasi telah ditangani oleh Kongres AS melalui pengesahan kontrol regulasi seperti Health Insurance Portability and Accountability Act (HIPAA). HIPAA mengharuskan individu untuk memberikan "persetujuan yang diinformasikan" mereka mengenai informasi yang mereka berikan dan penggunaan yang dimaksudkan saat ini dan di masa depan. Menurut sebuah artikel di Biotech Business Week ',“ '[i] n practice, HIPAA mungkin tidak menawarkan perlindungan yang lebih besar dari peraturan yang telah lama ada di arena penelitian, 'kata AAHC. Lebih penting lagi, tujuan aturan perlindungan melalui persetujuan yang diinformasikan dirusak oleh kompleksitas formulir persetujuan yang diperlukan dari pasien dan peserta, yang mendekati tingkat ketidakmampuan untuk memahami rata-rata individu. " Ini menggarisbawahi pentingnya anonimitas data dalam agregasi data dan praktik penambangan.

Undang-undang privasi informasi A.S. seperti HIPAA dan Undang-Undang Hak Pendidikan dan Privasi Keluarga (FERPA) hanya berlaku untuk area tertentu yang ditangani oleh masing-masing undang-undang tersebut. Penggunaan data mining oleh sebagian besar bisnis di A.S. tidak dikendalikan oleh undang-undang apa pun.

Aturan dan Undang-Undang Perlindungan Privasi di Indonesia

Dalam melindungi data pribadi pengguna internet, pemerintah Indonesia menggunakan beberapa instrumen hukum yang masing-masing berdiri sendiri.

Mereka adalah UU tentang Informasi dan Transaksi Elektronik (ITE), Peraturan Pemerintah Nomor 71 Tahun 2019 tentang Penyelenggaraan Sistem dan Transaksi Elektronik, Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016 tentang Perlindungan Data Pribadi dalam Sistem Elektronik.

Hukum Sektoral terkait Big Data & Privasi Data :

- Peraturan Bank Indonesia Nomor 7/15 / PBI / 2007 tentang Penerapan Manajemen Risiko dalam Pemanfaatan Teknologi Informasi oleh Bank.
- Undang-Undang Nomor 36 Tahun 2009 tentang Kesehatan (UU Kesehatan), yang meliputi sektor kesehatan.
- Peraturan Otoritas Jasa Keuangan Nomor 1 / POJK.07 / 2013 tentang Perlindungan Jasa Keuangan Sektor Consumer. Peraturan ini berlaku untuk sektor jasa keuangan, seperti Bank Umum, Bursa Efek, konsultan investasi, perusahaan asuransi, perusahaan bagunan.

Dasar Hukum Hak Cipta

Hukum Hak Cipta di Eropa

Karena kurangnya fleksibilitas dalam undang-undang hak cipta dan basis data Eropa, penambangan karya berhak cipta seperti penambangan web tanpa izin dari pemilik hak cipta tidak sah. Jika basis data adalah data murni di Eropa, kemungkinan besar tidak ada hak cipta, tetapi hak basis data mungkin ada sehingga penambangan data tunduk pada peraturan oleh *Database Directive*. Atas rekomendasi peninjauan Hargreaves, hal ini menyebabkan pemerintah Inggris Raya mengubah undang-undang hak ciptanya pada tahun 2014 untuk mengizinkan penambangan konten sebagai batasan dan pengecualian. Hanya negara kedua di dunia yang melakukannya setelah Jepang, yang memperkenalkan pengecualian pada 2009 untuk data mining. Namun, karena pembatasan Petunjuk Hak Cipta (*Copyright Directive*), pengecualian Inggris Raya hanya mengizinkan penambangan konten untuk tujuan non-komersial. Undang-undang hak cipta Inggris juga tidak mengizinkan ketentuan ini untuk ditimpa oleh syarat dan ketentuan kontrak. Komisi Eropa memfasilitasi diskusi pemangku kepentingan tentang penggalian teks dan data pada tahun 2013, dengan judul *Licences for Europe*. Fokus pada solusi untuk masalah hukum

ini menjadi perizinan dan bukan batasan dan pengecualian menyebabkan perwakilan universitas, peneliti, perpustakaan, kelompok masyarakat sipil dan penerbit akses terbuka untuk meninggalkan dialog pemangku kepentingan pada Mei 2013.

Hukum Hak Cipta di Amerika Serikat

Berbeda dengan Eropa, sifat fleksibel dari undang-undang hak cipta AS, dan khususnya penggunaan wajar berarti bahwa penambangan konten di Amerika, serta negara-negara penggunaan wajar lainnya seperti Israel, Taiwan, dan Korea Selatan dipandang legal. Karena penambangan konten bersifat transformatif, yaitu tidak menggantikan karya asli, itu dipandang sah menurut penggunaan wajar. Misalnya, sebagai bagian dari penyelesaian Buku Google, hakim ketua dalam kasus tersebut memutuskan bahwa proyek digitalisasi Google atas buku-buku dengan hak cipta adalah sah, sebagian karena penggunaan transformatif yang ditampilkan oleh proyek digitalisasi - salah satunya adalah penggalian teks dan data.

Hukum Hak Cipta di Indonesia

Hak Cipta diatur dalam UU No. 19 Tahun 2002 tentang Hak Cipta. Hak cipta adalah hak eksklusif bagi pencipta atau penerima hak untuk mengumumkan atau memperbanyak ciptaannya atau memberi izin untuk itu dengan tidak mengurangi pembatasan-pembatasan menurut perundang-undangan yang berlaku (pasal 1 angka 1 UU Hak Cipta).

Dalam pasal 72 ayat (3) UU Hak Cipta disebutkan bahwa “Barangsiapa dengan sengaja dan tanpa hak memperbanyak penggunaan untuk kepentingan komersial suatu Program Komputer dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/atau denda paling banyak Rp 500.000.000,00 (lima ratus juta rupiah).”

Suatu tindakan pelanggaran program komputer terjadi apabila dipenuhi unsur-unsur berikut:

1. Melakukan perbanyakan perangkat lunak (menggandakan atau menyalin program komputer dalam bentuk source code atau pun program aplikasinya);
2. Perbanyakan perangkat lunak dilakukan dengan sengaja dan tanpa hak (artinya tidak memiliki hak ciptaan atau lisensi hak cipta untuk menggunakan atau memperbanyak perangkat lunak);
3. Perbanyakan perangkat lunak dilakukan untuk kepentingan komersial (kepentingan komersial diterjemahkan secara praktek adalah perangkat lunak tersebut digunakan untuk kepentingan komersial, diperjualbelikan, disewakan atau cara-cara lain yang menguntungkan pelaku perbanyakan secara komersial).

Jadi, dapat dilihat bahwa yang dilindungi hak cipta mengenai software adalah program komputernya. Penerjemahan dari hasil pemrosesan data oleh software tersebut tidak bisa dikatakan melanggar hak cipta terhadap software tersebut, karena yang dilindungi oleh hak cipta adalah penggunaan program komputernya, bukan hasil dari program komputer tersebut.

Perangkat lunak Data Mining

Perangkat Lunak dan Aplikasi Penambangan Data Gratis

Aplikasi berikut tersedia di bawah lisensi gratis / sumber terbuka. Akses publik ke kode sumber aplikasi juga tersedia.

- Carrot2: Teks dan kerangka kerja pengelompokan hasil penelusuran.
- Chemicalize.org: Penambang struktur kimia dan mesin pencari web.
- ELKI: Sebuah proyek penelitian universitas dengan analisis cluster lanjutan dan metode deteksi outlier yang ditulis dalam bahasa Java.
- GATE: alat pemrosesan bahasa alami dan rekayasa bahasa.
- KNIME: The Konstanz Information Miner, kerangka kerja analisis data yang ramah pengguna dan komprehensif.
- Analisis Online Masif (MOA): penambangan aliran data besar waktu nyata dengan alat penyimpangan konsep dalam bahasa pemrograman Java.
- ML-Flex: Paket perangkat lunak yang memungkinkan pengguna berintegrasi dengan paket pembelajaran mesin pihak ketiga yang ditulis dalam bahasa pemrograman apa pun, menjalankan analisis klasifikasi secara paralel di beberapa node komputasi, dan membuat laporan HTML dari hasil klasifikasi.
- Pustaka MLPACK: kumpulan algoritme pembelajaran mesin siap pakai yang ditulis dalam bahasa C ++.
- MEPX - alat lintas platform untuk masalah regresi dan klasifikasi berdasarkan Genetic
- Varian pemrograman.
- NLTK (Natural Language Toolkit): Seperangkat pustaka dan program untuk pemrosesan bahasa alami simbolik dan statistik (NLP) untuk bahasa Python.
- OpenNN: Buka pustaka jaringan neural.
- Oranye: Rangkaian perangkat lunak penambangan data dan pembelajaran mesin berbasis komponen yang ditulis dalam bahasa Python.
- R: Bahasa pemrograman dan lingkungan perangkat lunak untuk komputasi statistik, penggalian data, dan grafik. Ini adalah bagian dari Proyek GNU.

- scikit-learn adalah pustaka pembelajaran mesin open source untuk bahasa pemrograman Python
- Torch: Pustaka deep learning open source untuk bahasa pemrograman Lua dan framework komputasi ilmiah dengan dukungan luas untuk algoritme machine learning.
- UIMA: UIMA (Unstructured Information Management Architecture) adalah kerangka kerja komponen untuk menganalisis konten tidak terstruktur seperti teks, audio dan video - aslinya dikembangkan oleh IBM.
- Weka: Rangkaian aplikasi perangkat lunak pembelajaran mesin yang ditulis dalam bahasa pemrograman Java.

Perangkat Lunak dan Aplikasi Penambangan Data dengan Hak Cipta (Paten)

Aplikasi berikut ini tersedia di bawah lisensi eksklusif.

- Angoss KnowledgeSTUDIO: alat penambangan data yang disediakan oleh Angoss.
- Clarabridge: solusi analisis teks kelas perusahaan.
- Platform HP Vertica Analytics: perangkat lunak penambangan data yang disediakan oleh HP.
- IBM SPSS Modeler: perangkat lunak penambangan data yang disediakan oleh IBM.
- KXEN Modeler: alat penambangan data yang disediakan oleh KXEN.
- LIONsolver: aplikasi perangkat lunak terintegrasi untuk penggalian data, intelijen bisnis, dan pemodelan yang mengimplementasikan pendekatan *Learning and Intelligent OptimizatioN* (LION).
- Megaputer Intelligence: perangkat lunak penambangan data dan teks disebut PolyAnalyst.
- Microsoft Analysis Services: perangkat lunak penambangan data yang disediakan oleh Microsoft.
- NetOwl: rangkaian produk analisis teks dan entitas multibahasa yang memungkinkan penggalian data.
- OpenText ™ Big Data Analytics: Visual Data Mining & Predictive Analysis dari Open Text Corporation
- Oracle Data Mining: perangkat lunak penambangan data oleh Oracle.
- PSeven: platform untuk otomatisasi simulasi dan analisis teknik, pengoptimalan multidisiplin, dan penggalian data yang disediakan oleh DATADVANCE.

- Qlucore Omics Explorer: perangkat lunak penambangan data yang disediakan oleh Qlucore.
- RapidMiner: Lingkungan untuk pembelajaran mesin dan eksperimen penggalian data.
- SAS Enterprise Miner: perangkat lunak penambangan data yang disediakan oleh SAS Institute.
- STATISTICA Data Miner: software data mining yang disediakan oleh StatSoft.
- Tanagra: Perangkat lunak penambangan data berorientasi visualisasi, juga untuk pengajaran.

3.2 Deteksi Anomali

Dalam data mining, deteksi anomali (juga *outlier detection* adalah identifikasi item, peristiwa, atau pengamatan yang tidak sesuai dengan pola yang diharapkan atau item lain dalam kumpulan data. Anomali adalah segala sesuatu yang tidak teratur - sesuatu yang menonjol atau tidak termasuk dalam sebuah cluster. Anomali di sini adalah penyimpangan dari pola data yang Anda amati selama hari kerja. Anomali dapat bervariasi dari satu bisnis ke yang lain, satu niche ke yang lain, satu industri ke yang lain dan banyak lagi. Biasanya item anomali akan diterjemahkan ke beberapa jenis masalah seperti penipuan bank, cacat struktural, masalah medis atau kesalahan dalam teks. Anomali juga disebut sebagai pencilan, kebaruan, kebisingan, penyimpangan, dan pengecualian.

Secara khusus, dalam konteks penyalahgunaan dan deteksi intrusi jaringan, objek yang menarik sering kali bukanlah objek langka, tetapi ledakan aktivitas yang tidak terduga. Pola ini tidak sesuai dengan definisi statistik umum dari pencilan sebagai objek langka, dan banyak metode deteksi pencilan (khususnya metode tanpa pengawasan) akan gagal pada data tersebut, kecuali jika telah diagregasikan dengan tepat. Alih-alih, algoritma analisis cluster mungkin dapat mendeteksi cluster mikro yang dibentuk oleh pola-pola ini.

Ada tiga kategori besar teknik deteksi anomali. Teknik pendeteksian anomali tanpa pengawasan mendeteksi anomali dalam kumpulan data pengujian yang tidak berlabel dengan asumsi bahwa sebagian besar instance dalam kumpulan data adalah normal dengan mencari instance yang tampaknya paling tidak cocok dengan kumpulan data lainnya. Teknik deteksi anomali yang diawasi memerlukan kumpulan data yang telah diberi label sebagai "normal" dan "abnormal" dan melibatkan pelatihan pengklasifikasi (perbedaan utama untuk banyak masalah klasifikasi statistik lainnya adalah sifat tidak seimbang yang melekat pada deteksi pencilan).

Teknik deteksi anomali semi-supervised membuat model yang mewakili perilaku normal dari kumpulan data pelatihan normal tertentu, dan kemudian menguji kemungkinan instance uji untuk dibuat oleh model yang dipelajari.

Penggunaan Deteksi Anomali

Penerapan deteksi anomali, dapat ditemukan pada sejumlah bidang-bidang kegunaan seperti: deteksi penyusupan (*intrusion detection*), deteksi penggelapan (*fraud detection*), gangguan ekosistem (*ecosystem disturbances*), kesehatan masyarakat (*public health*), dan kedokteran. Ini sering digunakan dalam *preprocessing* untuk menghapus data yang tidak wajar dari dataset. Dalam supervisi pembelajaran menghapus data yang tidak wajar dari kumpulan data sering kali menghasilkan peningkatan akurasi yang signifikan secara statistik.

a. Deteksi penyusupan

   Akses ke sumber data dalam instansi, baik komputer maupun jaringan tidak dapat dibantah bahwa datangnya dari tempat umum (internet), tetapi untuk akses yang bisanya bertujuan misalnya untuk mematikan fungsi server atau merusak data yang tersimpan akan memiliki perilaku yang berbeda, misalnya dari isi *header* protokol atau isi pesan tertentu yang disisipkan didalamnya. Dengan mengintegrasikan metode deteksi anomali pada protokol jaringan seperti *firewall* atau *router*, maka diharapkan dapat meningkatkan sistem keamanan jaringan.

b. Deteksi penggelapan

   Pembelian barang secara online dengan kartu kredit memicu munculnya para pencuri kartu kredit yang menggunakan nomor kartu kredit pelanggannya dengan membeli barang-barang yang biasanya mempunyai jenis, jumlah, pola pembelian yang berbeda dengan pelanggan yang seharusnya. Tanpa ada usaha dari bank penyedia layanan kartu kredit untuk secara berkala melakukan deteksi penyimpangan pola transaksi kartu kredit setiap pelanggannya tentu penggelapan isi kartu kredit yang dilakukan pencuri akan merugikan pelanggannya.

c. Gangguan ekosistem

   Dalam kehidupan di bumi, ada sejumlah pola-pola kehidupan seperti musim, pola kehidupan hewan, yang biasanya ada pola tertentu yang mengalami penyimpangan. Penyimpangan ini dapat mengakibatkan adanya gejala alam yang tidak biasa terjadi. Misalnya, pada masalah penyimpangan pola musim yang dapat mengakibatkan pola kehidupan hewan tertentu menjadi berubah. Deteksi penyimpangan kondisi alam sangat penting untuk dilakukan agar dapat mengetahui sejak dini bahaya-bahaya

yang mungkin bisa terjadi, seperti : banjir, kebakaran, gempa, pencemaran, dan sebagainya.

d. Kesehatan masyarakat (*public health*)

Di Indonesia, ada puskesmas di setiap kecamatan atau kabupaten, data-data rekam medis yang tersimpan di setiap puskesmas mencatat kasus-kasus penyakit yang ditangani disana yang umumnya pasiennya dari daerah terdekat dari lokasi puskesmas. Pola-pola penyakit yang diderita pasien dari keseluruhan puskemas dalam regional tertentu dapat diamati untuk mengetahui pola penyakit yang berbeda yang ditangani pada puskesmas kesalahan cara vaksinasi untuk anak-anak.

Teknik yang Populer

Beberapa teknik deteksi anomali telah diajukan dalam literatur. Beberapa teknik yang populer adalah:

- Teknik berbasis kepadatan (k-tetangga terdekat, faktor pencilan lokal, dan banyak lagi variasi dari konsep ini).
- Deteksi outlier berbasis spasi dan korelasi untuk data berdimensi tinggi.
- One class support vector machines.
- Jaringan saraf replikator.
- Deteksi outlier berbasis analisis cluster.
- Penyimpangan dari aturan asosiasi dan kumpulan item
- Deteksi outlier berbasis logika fuzzy.
- Teknik ensemble, menggunakan fitur bagging, normalisasi skor dan sumber keragaman yang berbeda.

Aplikasi Keamanan untuk Deteksi Anomali

Performa metode yang berbeda sangat bergantung pada kumpulan data dan parameter, dan metode memiliki sedikit keunggulan sistematis dibandingkan metode lainnya jika dibandingkan dengan banyak kumpulan data dan parameter.

Deteksi anomali melalui sistem deteksi intrusi (*Intrusion Detection System*-IDS) diperkenalkan oleh Dorothy Denning pada tahun 1986. Deteksi anomali untuk IDS biasanya dicapai dengan ambang batas dan statistik, tetapi juga dapat dilakukan dengan komputasi lunak, dan pembelajaran induktif. Jenis statistik yang diusulkan tahun 1999 termasuk profil pengguna, workstation, jaringan, host jarak jauh, kelompok pengguna, dan program

berdasarkan frekuensi, sarana, varians, kovarian, dan deviasi standar. Pendamping dari deteksi anomali dalam deteksi intrusi adalah deteksi penyalahgunaan.

ELKI adalah toolkit penambangan data Java *open-source* yang berisi beberapa algoritme deteksi anomali, serta akselerasi indeks untuk algoritme tersebut.

3.3 Metode Association Rule

Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan assosiatif yang memenuhi syarat minimum untuk support (*minimum support*) dan syarat minimum untuk confidence (*minimum confidence*).

Analisis asosiasii adalah metode pembelajaran mesin berbasis aturan untuk menemukan hubungan yang menarik antara variabel dalam database besar. Ini dimaksudkan untuk mengidentifikasi aturan kuat yang ditemukan dalam database menggunakan beberapa ukuran yang menarik. Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari berbagai teknik data mining lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frequensi tinggi (frequent pattern mining) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien.

Aturan assosiatif biasanya dinyatakan dalam bentuk :

{roti, mentega} {susu} (support = 40%, confidence = 50%)

Yang artinya : "50% dari transaksi di database yang memuat item roti dan mentega juga memuat item susu. Sedangkan 40% dari seluruh transaksi yang ada di database memuat ketiga item itu."

Dapat juga diartikan : "Seorang konsumen yang membeli roti dan mentega punya kemungkinan 50% untuk juga membeli susu. Aturan ini cukup signifikan karena mewakili 40% dari catatan transaksi selama ini."

Analisis asosiasi atau *association rule mining* adalah teknik data mining untuk menemukan aturan assosiatif antara suatu kombinasi item. Contoh aturan assosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan ini pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu. Analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisa isi keranjang belanja

di pasar swalayan. Analisis asosiasi juga sering disebut dengan istilah market basket analysis.

Penting tidaknya suatu aturan assosiatif dapat diketahui dengan dua parameter, support (nilai penunjang) yaitu persentase kombinasi item tersebut dalam database dan confidence (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan assosiatif.
Berdasarkan konsep aturan yang kuat, Rakesh Agrawal dkk. memperkenalkan aturan asosiasi untuk menemukan keteraturan antara produk dalam data transaksi skala besar yang dicatat oleh sistem point-of-sale (POS) di supermarket.
Misalnya, aturan {bawang, kentang} {burger} yang ditemukan dalam data penjualan supermarket akan menunjukkan bahwa jika pelanggan membeli bawang bombay dan kentang bersama-sama, kemungkinan besar mereka juga akan membeli hamburger.

Informasi tersebut dapat digunakan sebagai dasar untuk mengambil keputusan tentang aktivitas pemasaran seperti, misalnya, harga promosi atau penempatan produk. Selain contoh di atas dari aturan asosiasi analisis keranjang pasar yang digunakan saat ini di banyak area aplikasi termasuk penambangan penggunaan Web, deteksi intrusi, produksi berkelanjutan, dan bioinformatika. Berbeda dengan penambangan urutan, pembelajaran aturan asosiasi biasanya tidak mempertimbangkan urutan item baik dalam transaksi atau di seluruh transaksi.

Pemahaman Lebih Lanjut

Contoh database dengan 5 transaksi dan 5 item					
ID Transaski	Susu	Roti	Mentega	Minuman Botol	Popok Bayi
1	1	1	0	0	0
2	0	0	1	0	0
3	0	0	0	1	1
4	1	1	1	0	0
5	0	1	0	0	0

Mengikuti definisi asli oleh Agrawal dkk. masalah penambangan aturan asosiasi didefinisikan sebagai:
Nyatakan I $\{ i_1, i_2, ... i_n \}$ adalah satu set atribut biner *n* yang disebut *item*
Nyatakan D $\{ t_1, t_2, ... i_m \}$ adalah satu set transaksi yang disebut *database*.
Setiap transaksi di D memiliki ID transaksi unik dan berisi subset item dalam I.

Aturan didefinisikan sebagai implikasi dari formulir: $X \Rightarrow Y$
Di mana, $X, Y \subseteq I$ dan $X \cap Y = \varnothing$..

Setiap aturan disusun oleh dua set item yang berbeda, juga dikenal sebagai set item, X dan Y, di mana X disebut anteseden atau sisi kiri (kiri) dan konsekuensi Y atau sisi kanan (kanan). Untuk mengilustrasikan konsep, kami menggunakan contoh kecil dari domain supermarket. Set item adalah I {susu, roti, mentega, bir, popok} dan dalam tabel ditampilkan database kecil yang berisi item, di mana, di setiap entri, nilai 1 berarti adanya item dalam transaksi yang sesuai, dan nilai 0 mewakili tidak adanya item dalam transaksi tersebut. Contoh aturan untuk supermarket adalah {mentega, roti} {susu} yang berarti jika mentega dan roti dibeli, pelanggan juga membeli susu.

Catatan: contoh ini sangat kecil. Dalam aplikasi praktis, aturan memerlukan dukungan dari beberapa ratus transaksi sebelum dapat dianggap signifikan secara statistik, dan kumpulan data sering kali berisi ribuan atau jutaan transaksi.

Konsep dan Istilah dalam *Association Rule*

Untuk memilih aturan yang menarik dari himpunan semua aturan yang mungkin, batasan pada berbagai ukuran signifikansi dan kepentingan digunakan. Kendala yang paling terkenal adalah ambang batas minimum untuk dukungan dan keyakinan.

Misalkan X adalah himpunan item, $X \Rightarrow Y$ adalah sebuah *association rule* dan T adalah himpunan transaksi dari basis data tertentu.

Support

Support adalah indikasi seberapa sering item-set muncul di database.

Nilai Support X terhadap T didefinisikan sebagai proporsi transaksi dalam basis data yang berisi set item X. Dalam rumus: *supp (X) / N*

Dalam database contoh, item-set {beer, diapers} memiliki Support karena terjadi di 20% dari semua transaksi (1 dari 5 transaksi). Argumen supp () adalah sekumpulan prasyarat, dan dengan demikian menjadi lebih terbatas seiring berkembangnya (daripada lebih inklusif).

Confidence

Confidence adalah indikasi seberapa sering aturan itu terbukti benar.
Keyakinan adalah indikasi seberapa sering aturan tersebut terbukti benar.

Nilai *Confidence* suatu aturan, $X \Rightarrow Y$, sehubungan dengan satu set transaksi T, adalah proporsi transaksi yang mengandung X yang juga mengandung Y.

Confidence didefinisikan sebagai: $\text{conf}(X \Rightarrow Y) = \text{supp}(X \cup Y) / \text{supp}(X).$

Misalnya, *rule* {butter, bread} {milk} memiliki tingkat *confidence* 0,2 > database, yang berarti bahwa untuk 100% transaksi yang mengandung mentega dan roti, *rule* benar (100% frekuensi pelanggan membeli mentega dan roti, susu juga dibeli).

Perhatikan bahwa $\text{supp}(X \cup Y)$ berarti support dari penyatuan item di X dan Y. Ini agak membingungkan karena biasanya kita berpikir dalam istilah probabilitas peristiwa dan bukan kumpulan item. Kita dapat menulis ulang $\text{supp}(X \cup Y)$ sebagai probabilitas gabungan $P(E_X \cap E_Y)$, di mana E_X dan E_Y adalah peristiwa di mana transaksi masing-masing berisi itemset X atau Y. Dengan demikian kepercayaan dapat diartikan sebagai estimasi probabilitas bersyarat $P(E_Y \mid E_X)$, probabilitas menemukan RHS dari aturan dalam transaksi dengan syarat bahwa transaksi tersebut juga mengandung LHS.

Lift

Peningkatan aturan didefinisikan sebagai:

$$\text{lift}(X \Rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X) \times \text{supp}(Y)}$$

atau rasio *support* yang diamati dengan yang diharapkan jika X dan Y independen.
Misalnya, aturan {susu, roti} $\Rightarrow$ {mentega} memiliki *lift* $\frac{0.2}{0.4 \times 0.4} = 1.25.$

Jika aturan memiliki peningkatan 1, itu akan menyiratkan bahwa probabilitas kemunculan anteseden dan konsekuen tidak bergantung satu sama lain. Jika dua peristiwa tidak bergantung satu sama lain, tidak ada aturan yang dapat dibuat yang melibatkan kedua peristiwa tersebut.

Jika peningkatan> 1, itu memungkinkan kita mengetahui sejauh mana kedua kemunculan tersebut bergantung satu sama lain, dan membuat aturan tersebut berpotensi berguna

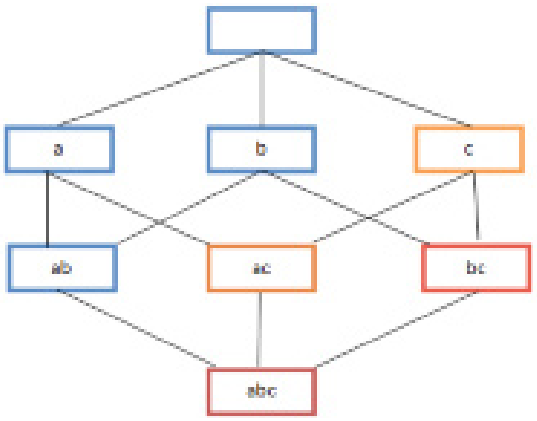
untuk memprediksi konsekuensi dalam kumpulan data mendatang. Nilai *lift* adalah karena mempertimbangkan keyakinan aturan dan kumpulan data secara keseluruhan.

Conviction

Conviction suatu *rule* didefinisikan dengan: $\text{conv}(X \Rightarrow Y) = \frac{1 - \text{supp}(Y)}{1 - \text{conf}(X \Rightarrow Y)}$

Misalnya, *Rule* {susu, roti} $\Rightarrow$ {mentega} memiliki *Conviction* 1 0,4 1,2, dan dapat diartikan sebagai rasio frekuensi yang diharapkan bahwa X muncul tanpa Y (artinya, frekuensi aturan tersebut membuat prediksi salah) jika X dan Y independen dibagi dengan frekuensi pengamatan prediksi salah. Dalam contoh ini, nilai *Conviction* 1,2 menunjukkan bahwa *Rule* {susu, roti} $\Rightarrow$ {mentega} akan salah 20% lebih sering (1,2 kali lebih sering) jika hubungan antara X dan Y murni kebetulan.

Proses



Kisi set *item yang sering, di mana warna kotak menunjukkan berapa banyak transaksi yang* berisi kombinasi *item. Perhatikan bahwa tingkat kisi yang lebih rendah dapat memuat paling* banyak jumlah *minimum item orang tua mereka; misalnya {ac} hanya boleh memiliki paling min* (a, c) *item. Ini disebut properti penutupan-bawah*

Association Rule biasanya diperlukan untuk memenuhi *support* minimum yang ditentukan pengguna dan *Conviction* minimum yang ditentukan pengguna dan keyakinan minimum yang ditentukan pengguna pada saat yang sama. Pembuatan aturan asosiasi biasanya dibagi menjadi dua langkah terpisah:

1. Ambang batas dukungan minimum diterapkan untuk menemukan semua set item yang sering di database.
2. Batasan keyakinan minimum diterapkan pada kumpulan item yang sering ini untuk membentuk aturan.

Meskipun langkah kedua sangat mudah, langkah pertama membutuhkan lebih banyak perhatian.

Menemukan semua set item yang sering dalam database sulit karena melibatkan pencarian semua item- set yang mungkin (kombinasi item). Himpunan himpunan item yang mungkin adalah himpunan daya di atas I dan memiliki ukuran 2n 1 (tidak termasuk himpunan kosong yang bukan merupakan himpunan item yang valid). Meskipun ukuran set daya tumbuh secara eksponensial dalam jumlah item n di I, pencarian yang efisien dimungkinkan menggunakan properti dukungan penutupan ke bawah (juga disebut anti-monotonisitas) yang menjamin bahwa untuk itemset yang sering, semua subsetnya adalah juga sering dan dengan demikian untuk item-set yang jarang, semua supersetnya juga harus jarang. Memanfaatkan properti ini, algoritme yang efisien (mis., Apriori dan Eclat) dapat menemukan semua set item yang sering.

Sejarah

Konsep *association rule* dipopulerkan terutama karena artikel Agrawal dkk tahun 1993, yang telah memperoleh lebih dari 18.000 kutipan menurut Google Cendekia, per Agustus 2015, dan dengan demikian menjadi salah satu makalah yang paling banyak dikutip di bidang *Data Mining*. Namun, ada kemungkinan bahwa apa yang sekarang disebut "aturan asosiasi" mirip dengan apa yang muncul dalam makalah 1966 tentang GUHA, metode penambangan data umum yang dikembangkan oleh Petr Hájek dkk.

Penggunaan awal (sekitar 1989) dari dukungan dan keyakinan minimum untuk menemukan semua aturan asosiasi adalah kerangka Pemodelan Berbasis Fitur, yang menemukan semua aturan dengan supp (X) dan conf (X Y) lebih besar daripada batasan yang ditentukan pengguna.

Association Rule yang Baik Berdasarkan Statistik

Salah satu batasan dari pendekatan standar untuk menemukan asosiasi adalah bahwa dengan mencari sejumlah besar asosiasi yang mungkin untuk mencari koleksi item yang tampaknya terkait, terdapat risiko besar untuk menemukan banyak asosiasi palsu. Ini adalah kumpulan item yang terjadi bersamaan dengan frekuensi yang tidak terduga dalam data, tetapi hanya terjadi secara kebetulan. Sebagai contoh, misalkan kita sedang mempertimbangkan koleksi 10.000 item dan mencari *rule* yang berisi dua item di sisi kiri dan 1 item di sisi kanan. Ada sekitar 1.000.000.000.000 *rule* seperti itu. Jika kita

menerapkan uji statistik untuk independensi dengan tingkat signifikansi 0,05 itu berarti hanya ada peluang 5% untuk menerima suatu *rule* jika tidak ada *association*. Jika kita berasumsi bahwa tidak ada asosiasi, kita tetap berharap menemukan 50.000.000.000 *rule*. Penemuan asosiasi yang sehat secara statistik mengontrol risiko ini, dalam banyak kasus mengurangi risiko menemukan asosiasi palsu ke tingkat signifikansi yang ditentukan pengguna.

Banyak algoritma untuk menghasilkan *association rule* disajikan dari waktu ke waktu. Beberapa algoritma terkenal adalah Apriori, Eclat dan FP-Growth, tetapi mereka hanya melakukan setengah dari pekerjaan, karena mereka adalah algoritma untuk menambang item items. Langkah lain perlu dilakukan setelah menghasilkan aturan dari itemset sering ditemukan dalam database.

Algoritma Apriori

Algoritma apriori merupakan salah satu algoritma klasik data mining. Algoritma apriori digunakan agar komputer dapat mempelajari aturan asosiasi, mencari pola hubungan antar satu atau lebih item dalam suatu dataset. Algoritma apriori banyak digunakan pada data transaksi atau biasa disebut market basket, misalnya sebuah swalayan memiliki market basket, dengan adanya algoritma apriori, pemilik swalayan dapat mengetahui pola pembelian seorang konsumen, jika seorang konsumen membeli item A , B, punya kemungkinan 50% dia akan membeli item C, pola ini sangat signifikan dengan adanya data transaksi selama ini.

Algoritma Eclat

Eclat (singkatan dari *Equivalence Class Transformatio*n) adalah algoritma pencarian kedalaman-pertama yang menggunakan set persimpangan. Ini adalah algoritma elegan alami yang cocok untuk eksekusi berurutan maupun paralel dengan properti peningkat lokalitas. Ini pertama kali diperkenalkan oleh Zaki, Parthasarathy, Li dan Ogihara dalam serangkaian makalah yang ditulis pada tahun 1997. Algoritma Eclat menggunakan basis data dengan tata letak vertikal. Kelebihan dari Eclat adalah proses dan performa penghitungan support dari semua itemsets dilakukan dengan lebih efisien dibandingkan dengan algoritma apriori.

Algoritma FP-growth

Algoritma FP-Growth merupakan pengembangan dari algoritma Apriori. Algoritma Frequent Pattern Growth adalah salah satu alternatif algoritma yang dapat digunakan

untuk menentukan himpunan data yang paling sering muncul (frequent itemset) dalam sebuah kumpulan data. Pada algoritma FP-Growth menggunakan konsep pembangunan tree, yang biasa disebut FP-Tree, dalam pencarian frequent itemsets bukan menggunakan generate candidate seperti yang dilakukan pada algoritma Apriori. Dengan menggunakan konsep tersebut, algoritma FP-Growth menjadi lebih cepat daripada algoritma Apriori. Metode FP-Growth dibagi menjadi tiga tahapan utama, yaitu:

1. tahap pembangkitan conditional pattern base,
2. tahap pembangkitan conditional FP

Langkah pertama, algoritme menghitung kemunculan item (pasangan nilai atribut) dalam kumpulan data, dan menyimpannya ke 'tabel header'. Pada lintasan kedua, ia membangun struktur pohon FP dengan memasukkan instance. Item di setiap contoh harus diurutkan berdasarkan urutan frekuensinya dalam dataset, sehingga pohon dapat diproses dengan cepat. Item di setiap contoh yang tidak memenuhi ambang batas cakupan minimum akan dibuang. Jika banyak instance berbagi item paling sering, FP-tree menyediakan kompresi tinggi di dekat root pohon.

Pemrosesan rekursif dari versi terkompresi dari set data utama ini menumbuhkan kumpulan item besar secara langsung, alih-alih menghasilkan item kandidat dan mengujinya pada seluruh database. Pertumbuhan dimulai dari bagian bawah tabel header (memiliki cabang terpanjang), dengan menemukan semua contoh yang cocok dengan kondisi tertentu. Pohon baru dibuat, dengan jumlah yang diproyeksikan dari pohon asli sesuai dengan kumpulan instance yang bersyarat pada atribut, dengan setiap node mendapatkan jumlah anak-anaknya. Pertumbuhan rekursif berakhir ketika tidak ada item individu yang bersyarat pada atribut memenuhi ambang dukungan minimal, dan pemrosesan berlanjut pada item header yang tersisa dari pohon-FP asli.

Setelah proses rekursif selesai, semua kumpulan item besar dengan cakupan minimum telah ditemukan, dan pembuatan aturan asosiasi dimulai.

Lainnya

AprioriDP

AprioriDP menggunakan Pemrograman Dinamis dalam penambangan itemset yang sering. Prinsip kerjanya adalah menghilangkan kandidat generasi seperti FP-tree, tetapi ia menyimpan hitungan dukungan dalam struktur data khusus alih-alih tree.
Algoritma Aturan Pertambangan Berbasis Konteks

CBPNARM adalah algoritma yang baru dikembangkan yang dikembangkan pada 2013 untuk menambang aturan asosiasi berdasarkan konteks. Ini menggunakan variabel konteks atas dasar di mana dukungan dari itemset diubah atas dasar aturan akhirnya diisi ke aturan yang ditetapkan.

Algoritma berbasis Node-set

FIN, PrePost dan PPV adalah tiga algoritma berdasarkan set simpul. Mereka menggunakan node dalam Fptree pengkodean untuk mewakili itemset, dan menggunakan strategi pencarian kedalaman-pertama untuk menemukan itemset yang sering menggunakan "persimpangan" set node.

Prosedur GUHA ASSOC

GUHA adalah metode umum untuk analisis data eksplorasi yang memiliki dasar teoritis dalam kalkulus pengamatan.

Prosedur ASSOC adalah metode GUHA yang menambang untuk aturan asosiasi umum menggunakan operasi bitstring cepat. Aturan asosiasi yang ditambang dengan metode ini lebih umum daripada yang dihasilkan oleh apriori, misalnya "item" dapat dihubungkan baik dengan konjungsi dan disjungsi dan hubungan antara anteseden dan konsekuensi dari aturan tidak terbatas untuk menetapkan dukungan minimum dan kepercayaan seperti dalam apriori: kombinasi sewenang-wenang dari langkah-langkah bunga yang didukung dapat digunakan.

Pencarian OPUS

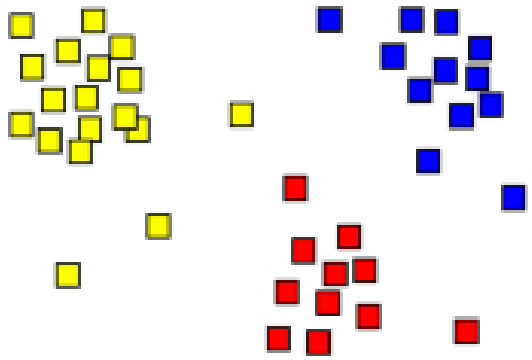
OPUS adalah algoritma yang efisien untuk penemuan aturan yang, berbeda dengan kebanyakan alternatif, tidak memerlukan kendala monoton atau anti-monoton seperti dukungan minimum. Awalnya digunakan untuk menemukan aturan untuk konsekuensi tetap, kemudian diperluas untuk menemukan aturan dengan item apa pun sebagai konsekuensi. Pencarian OPUS adalah teknologi inti dalam sistem penemuan asosiasi Magnum Opus yang populer.

Kasus : Penerapan Association Rule Mining

Kisah terkenal tentang *association rule mining* adalah kisah "*beer and diaper*". Sebuah survei terhadap perilaku pembeli supermarket menemukan bahwa pelanggan (mungkin laki-laki muda) yang membeli popok cenderung juga membeli bir. Anekdot ini menjadi populer sebagai contoh bagaimana aturan asosiasi yang tidak terduga dapat ditemukan dari data sehari-hari. Ada berbagai pendapat tentang seberapa banyak cerita itu benar. Daniel Powers mengatakan:

Pada tahun 1992, Thomas Blischok, manajer grup konsultasi ritel di Teradata, dan stafnya menyiapkan analisis terhadap 1,2 juta keranjang pasar dari sekitar 25 toko Obat Osco. Kueri database dikembangkan untuk mengidentifikasi afinitas. Analisis tersebut “menemukan bahwa antara 5:00 dan 7:00 malam. bahwa konsumen membeli bir dan popok ”. Manajer Osco TIDAK mengeksploitasi hubungan bir dan popok dengan mendekatkan produk di rak.

3.4 Analisis Cluster



Hasil analisis *kluster ditampilkan sebagai pewarnaan kotak menjadi tiga kelompok.*

Analisis cluster atau pengelompokan adalah tugas mengelompokkan sekumpulan objek sedemikian rupa sehingga objek dalam grup yang sama (disebut cluster) lebih mirip (dalam beberapa hal atau lainnya) satu sama lain daripada yang ada di grup lain (cluster) . Ini adalah tugas utama penambangan data eksplorasi, dan teknik umum untuk analisis data statistik, yang digunakan di banyak bidang, termasuk pembelajaran mesin, pengenalan pola, analisis gambar, pengambilan informasi, bioinformatika, kompresi data, dan grafik komputer.

Analisis cluster sendiri bukanlah satu algoritma khusus, tetapi tugas umum yang harus diselesaikan. Hal ini dapat dicapai dengan berbagai algoritme yang berbeda secara signifikan dalam pengertian mereka tentang apa yang membentuk sebuah cluster dan bagaimana menemukannya secara efisien. Pengertian populer tentang cluster termasuk kelompok dengan jarak kecil di antara anggota cluster, area padat ruang data, interval atau distribusi statistik tertentu. Oleh karena itu, pengelompokan dapat dirumuskan sebagai masalah pengoptimalan multi-tujuan. Algoritme pengelompokan yang sesuai dan pengaturan parameter (termasuk nilai seperti fungsi jarak yang akan digunakan, ambang

batas kepadatan atau jumlah kluster yang diharapkan) bergantung pada kumpulan data individu dan tujuan penggunaan hasil. Analisis cluster seperti itu bukanlah tugas otomatis, tetapi proses berulang dari penemuan pengetahuan atau optimasi multi-tujuan interaktif yang melibatkan uji coba dan kegagalan. Sering kali perlu untuk memodifikasi praproses data dan parameter model sampai hasilnya mencapai properti yang diinginkan.

Selain istilah clustering, ada beberapa istilah yang memiliki arti serupa, antara lain klasifikasi otomatis, taksonomi numerik, botriologi, dan analisis tipologi. Perbedaan halus sering kali terjadi dalam penggunaan hasil: sementara dalam data mining, kelompok yang dihasilkan adalah masalah yang diminati, dalam klasifikasi otomatis kekuatan diskriminatif yang dihasilkan menjadi perhatian.

Analisis cluster berasal dari antropologi oleh Driver dan Kroeber pada tahun 1932 dan diperkenalkan ke psikologi oleh Zubin pada tahun 1938 dan Robert Tryon pada tahun 1939 dan terkenal digunakan oleh Cattell mulai tahun 1943 untuk klasifikasi teori sifat dalam psikologi kepribadian..

Definisi

Gagasan tentang "cluster" tidak dapat didefinisikan secara tepat, yang merupakan salah satu alasan mengapa ada begitu banyak algoritma pengelompokan. Ada penyebut yang sama: sekelompok objek data. Namun, peneliti yang berbeda menggunakan model cluster yang berbeda, dan untuk masing-masing model cluster ini lagi-lagi algoritma yang berbeda dapat diberikan. Gagasan tentang cluster, seperti yang ditemukan oleh algoritme yang berbeda, sangat bervariasi dalam propertinya. Memahami "model cluster" ini adalah kunci untuk memahami perbedaan antara berbagai algoritme. Model cluster tipikal meliputi:

- Model konektivitas: misalnya, pengelompokan hierarki membangun model berdasarkan konektivitas jarak.
- Model sentroid: misalnya, algoritme k-means mewakili setiap cluster dengan vektor mean tunggal.
- Model distribusi: cluster dimodelkan menggunakan distribusi statistik, seperti distribusi normal multivariat yang digunakan oleh algoritme Maksimalisasi Ekspektasi.
- Model kepadatan: misalnya, DBSCAN dan OPTICS mendefinisikan cluster sebagai daerah padat yang terhubung di ruang data.
- Model subruang: dalam Biclustering (juga dikenal sebagai Co-clustering atau two-mode-clustering), cluster dimodelkan dengan anggota cluster dan atribut yang relevan.

- Model grup: beberapa algoritme tidak memberikan model yang disempurnakan untuk hasil mereka dan hanya menyediakan informasi pengelompokan.
- Model berbasis grafik: sebuah klik, yaitu subset dari node dalam grafik sedemikian rupa sehingga setiap dua node dalam subset dihubungkan oleh edge dapat dianggap sebagai bentuk prototipe cluster. Relaksasi dari persyaratan konektivitas lengkap (sebagian kecil dari tepinya bisa hilang) dikenal sebagai kuasi-klik, seperti dalam algoritme pengelompokan SKT.

Sebuah "clustering" pada dasarnya adalah sekumpulan cluster seperti itu, biasanya berisi semua objek dalam kumpulan data. Selain itu, ini dapat menentukan hubungan cluster satu sama lain, misalnya, hierarki cluster yang disematkan satu sama lain. Pengelompokan secara kasar dapat dibedakan sebagai:

- hard clustering: setiap objek termasuk dalam cluster atau tidak
- pengelompokan lembut (juga: pengelompokan fuzzy): setiap objek dimiliki oleh setiap kluster dengan derajat tertentu (misalnya, kemungkinan menjadi milik kluster)

Ada juga kemungkinan perbedaan yang lebih halus, misalnya:

- clustering partisi ketat: di sini setiap objek milik tepat satu cluster
- pengelompokan partisi ketat dengan pencilan: objek juga dapat dimiliki oleh tidak ada kluster, dan dianggap pencilan.
- pengelompokan yang tumpang tindih (juga: pengelompokan alternatif, pengelompokan multi-tampilan): meskipun biasanya pengelompokan keras, objek mungkin termasuk dalam lebih dari satu kluster.
- pengelompokan hierarki: objek yang dimiliki kluster anak juga termasuk dalam kluster induk
- pengelompokan subruang: saat pengelompokan yang tumpang tindih, dalam subruang yang ditentukan secara unik, kluster tidak diharapkan untuk tumpang tindih.

Clustering Algorithms

Clustering algorithms dapat dikategorikan berdasarkan model klusternya, seperti yang tercantum di atas. Ringkasan berikut hanya akan mencantumkan contoh paling menonjol dari *Clustering algorithms* (algoritma pengelompokan), karena mungkin ada lebih dari 100 algoritma pengelompokan yang diterbitkan. Tidak semua menyediakan model untuk cluster mereka sehingga tidak dapat dengan mudah dikategorikan.

Tidak ada *Clustering algorithms* yang "benar" secara objektif, tetapi seperti yang dicatat, "pengelompokan bergantung pada yang melihatnya". *Clustering algorithms* yang paling tepat untuk masalah tertentu sering kali perlu dipilih secara eksperimental, kecuali jika ada alasan matematis untuk memilih satu model kluster daripada yang lain. Perlu dicatat bahwa algoritme yang dirancang untuk satu jenis model tidak memiliki peluang pada kumpulan data yang berisi jenis model yang sangat berbeda. Misalnya, k-means tidak dapat menemukan cluster non-cembung.

Clustering Berbasis Konektivitas (*Hierarchical Clustering*)

Pengelompokan berbasis konektivitas, juga dikenal sebagai *hierarchical clustering* (pengelompokan hierarki), didasarkan pada gagasan inti tentang objek yang lebih terkait dengan objek di sekitar daripada ke objek yang lebih jauh. Algoritme ini menghubungkan "objek" untuk membentuk "cluster" berdasarkan jaraknya. Sebuah cluster dapat dijelaskan sebagian besar oleh jarak maksimum yang diperlukan untuk menghubungkan bagian-bagian cluster. Pada jarak yang berbeda, cluster yang berbeda akan terbentuk, yang dapat direpresentasikan menggunakan dendrogram, yang menjelaskan dari mana nama umum "hierarchical clustering" berasal: algoritme ini tidak menyediakan satu partisi dari kumpulan data, tetapi menyediakan hierarki yang luas dari cluster yang bergabung satu sama lain pada jarak tertentu. Dalam dendrogram, sumbu y menandai jarak di mana cluster bergabung, sedangkan objek ditempatkan di sepanjang sumbu x sehingga cluster tidak bercampur.

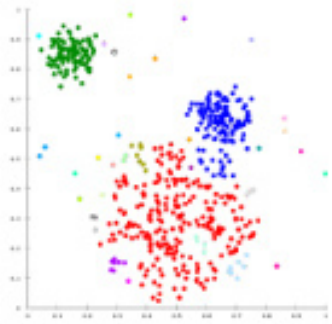
Pengelompokan berbasis konektivitas adalah sekumpulan metode yang berbeda menurut cara penghitungan jarak. Terlepas dari pilihan fungsi jarak yang biasa, pengguna juga perlu memutuskan kriteria hubungan (karena cluster terdiri dari beberapa objek, ada beberapa kandidat untuk menghitung jarak) yang akan digunakan. Pilihan populer dikenal sebagai pengelompokan tautan tunggal (jarak minimum objek), pengelompokan tautan lengkap (jarak maksimum objek) atau UPGMA ("Metode Kelompok Pasangan Tanpa Bobot dengan Rata-Rata Aritmatika", juga dikenal sebagai pengelompokan tautan rata-rata). Lebih jauh, pengelompokan hierarki dapat bersifat aglomeratif (dimulai dengan elemen tunggal dan menggabungkannya menjadi beberapa kluster) atau memecah belah (dimulai dengan kumpulan data lengkap dan membaginya menjadi beberapa partisi).

Metode ini tidak akan menghasilkan partisi unik dari kumpulan data, tetapi hierarki yang darinya pengguna masih perlu memilih cluster yang sesuai. Mereka tidak terlalu kuat terhadap pencilan, yang akan muncul sebagai cluster tambahan atau bahkan

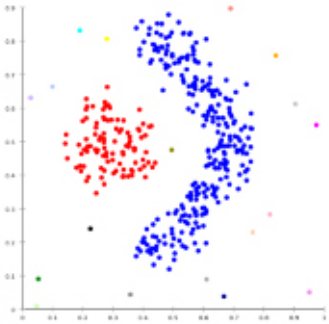
menyebabkan cluster lain bergabung (dikenal sebagai "fenomena rantai", khususnya dengan clustering linkage tunggal). Dalam kasus umum, kompleksitasnya adalah $\mathcal{O}(n^3)$ untuk pengelompokan aglomeratif dan $\mathcal{O}(2^{n-1})$ untuk pengelompokan memecah belah, yang membuatnya terlalu lambat untuk kumpulan data yang besar. Untuk beberapa kasus khusus, metode efisien optimal (kompleksitas $\mathcal{O}(n^2)$) diketahui: SLINK untuk tautan tunggal dan CLINK untuk pengelompokan tautan lengkap. Dalam komunitas data mining, metode ini diakui sebagai landasan teoritis analisis cluster, tetapi sering dianggap usang. Namun mereka memberikan inspirasi untuk metode selanjutnya seperti pengelompokan berbasis kepadatan.

Metode *Single Linkage Clustering*

Metode Single Linkage merupakan metode pengelompokan hierarchical clustering. Metode *Single Linkage* mengelompokkan dokumen didasarkan pada jarak terdekat antar dokumen.



Linkage clustering pada data Gaussian. Pada 35 cluster, cluster terbesar mulai terpecah-pecah menjadi bagian-bagian yang lebih kecil, sementara sebelumnya masih terhubung ke yang terbesar kedua karena efek single-link.



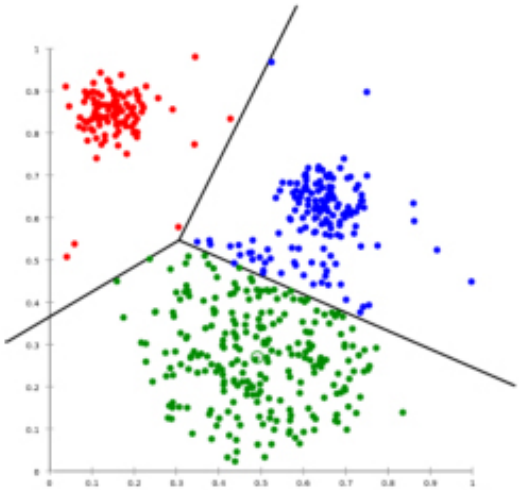
Single Linkage pada kelompok berbasis kepadatan. 20 cluster diekstraksi, yang sebagian besar mengandung elemen tunggal, karena *Linkage Clustering*e tidak memiliki gagasan “noise”.

Clustering Berbasis Centroid

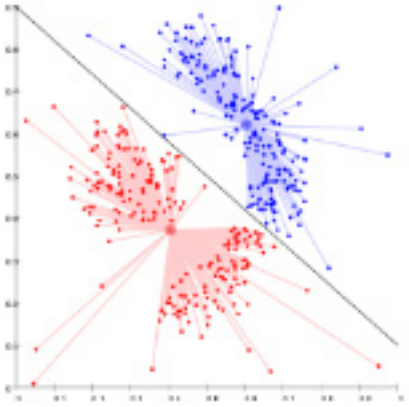
Dalam pengelompokan berbasis sentroid, kluster diwakili oleh vektor pusat, yang mungkin tidak harus menjadi anggota kumpulan data. Ketika jumlah cluster ditetapkan ke k, clustering k-mean memberikan definisi formal sebagai masalah optimasi: temukan pusat cluster k dan tandai objek ke pusat cluster terdekat, sehingga jarak kuadrat dari cluster diminimalkan .

Masalah pengoptimalan itu sendiri dikenal sebagai NP-hard, dan dengan demikian pendekatan yang umum adalah hanya mencari solusi perkiraan. Metode pendekatan yang sangat terkenal adalah algoritme Lloyd, yang sering kali disebut sebagai "algoritme k-means". Namun itu hanya menemukan optimal lokal, dan biasanya dijalankan beberapa kali dengan inisialisasi acak yang berbeda. Variasi k-means sering kali mencakup pengoptimalan seperti memilih yang terbaik dari beberapa proses, tetapi juga membatasi sentroid ke anggota kumpulan data (k-medoids), memilih median (pengelompokan k-median), memilih pusat awal secara tidak acak (K-means ++) atau mengizinkan penetapan cluster fuzzy (Fuzzy c-means).

Sebagian besar algoritme tipe k-means memerlukan jumlah kluster - k - untuk ditentukan terlebih dahulu, yang dianggap sebagai salah satu kelemahan terbesar dari algoritme ini. Selain itu, algoritme lebih memilih cluster dengan ukuran yang kurang lebih sama, karena mereka akan selalu menetapkan objek ke pusat massa terdekat. Hal ini sering menyebabkan kesalahan pemotongan batas di antara cluster (yang tidak mengherankan, karena algoritme mengoptimalkan pusat cluster, bukan batas cluster).



K-means memisahkan *data menjadi sel-sel Voronoi, yang mengasumsikan cluster berukuran* sama (tidak *memadai di sini*



K-means tidak *dapat merepresentasikan cluster berbasis kepadatan*

K-means memiliki sejumlah sifat teoritis yang menarik. Pertama, partisi ruang data menjadi struktur yang dikenal sebagai diagram Voronoi. Kedua, ini secara konseptual dekat dengan klasifikasi tetangga terdekat, dan karena itu populer dalam pembelajaran mesin. Ketiga, dapat dilihat sebagai variasi dari klasifikasi berbasis model, dan algoritma Lloyd sebagai variasi dari algoritma Pemaksimalan ekspektasi untuk model ini yang dibahas di bawah ini.

Clustering Berbasis Distribusi

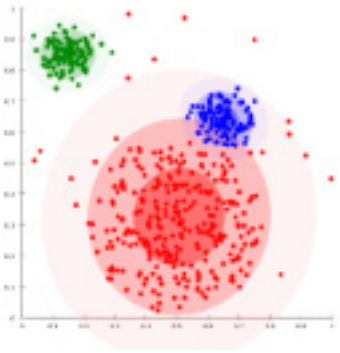
Model pengelompokan yang paling dekat hubungannya dengan statistik didasarkan pada model distribusi. Kluster kemudian dapat dengan mudah didefinisikan sebagai objek yang kemungkinan besar memiliki distribusi yang sama. Properti yang nyaman dari pendekatan ini adalah bahwa ini sangat mirip dengan cara kumpulan data buatan dihasilkan: dengan mengambil sampel objek acak dari suatu distribusi.

Meskipun fondasi teoretis dari metode ini sangat baik, mereka mengalami satu masalah utama yang dikenal sebagai overfitting, kecuali jika kendala diletakkan pada kompleksitas model. Model yang lebih kompleks biasanya dapat menjelaskan data dengan lebih baik, yang membuat pemilihan kompleksitas model yang sesuai secara inheren sulit.

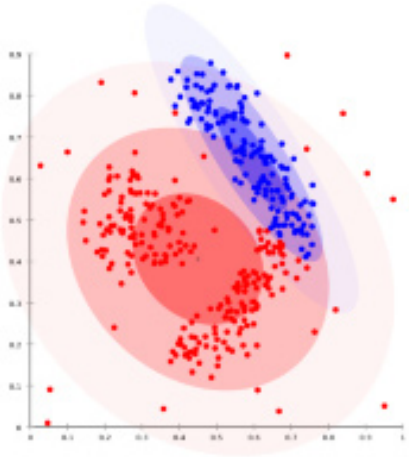
Salah satu metode yang menonjol dikenal sebagai model campuran Gaussian (menggunakan algoritma ekspektasi-maksimisasi). Di sini, kumpulan data biasanya dimodelkan dengan jumlah distribusi Gaussian yang tetap (untuk menghindari overfitting) yang diinisialisasi secara acak dan yang parameternya dioptimalkan secara berulang agar lebih sesuai dengan kumpulan data. Ini akan menyatu ke optimal lokal, sehingga beberapa proses dapat menghasilkan hasil yang berbeda. Untuk mendapatkan pengelompokan keras,

objek sering kali kemudian ditugaskan ke distribusi Gaussian yang kemungkinan besar mereka miliki; untuk pengelompokan lunak, ini tidak perlu.

Pengelompokan berbasis distribusi menghasilkan model kompleks untuk kluster yang dapat menangkap korelasi dan ketergantungan antar atribut. Namun, algoritme ini memberikan beban tambahan pada pengguna: untuk banyak kumpulan data nyata, mungkin tidak ada model matematika yang didefinisikan secara ringkas (misalnya dengan asumsi distribusi Gaussian adalah asumsi yang cukup kuat pada data).
Contoh pengelompokan Expectation-Maximization (EM)



Pada data *terdistribusi Gaussian, EM berfungsi dengan baik, karena EM menggunakan* Gaussians untuk *memodelkan cluster*



Kluster berbasis *kepadatan tidak dapat dimodelkan menggunakan distribusi Gaussian*

Clustering Berbasis Kepadatan

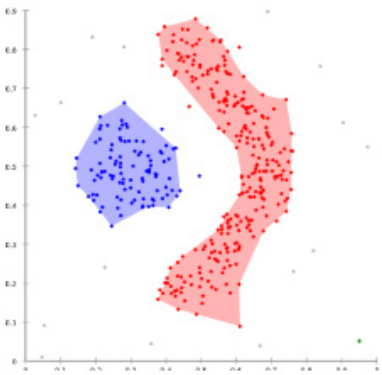
Dalam pengelompokan berbasis kepadatan, kluster didefinisikan sebagai area dengan kepadatan lebih tinggi daripada kumpulan data lainnya. Objek di area yang jarang ini - yang diperlukan untuk memisahkan cluster - biasanya dianggap sebagai titik kebisingan dan perbatasan.

Metode pengelompokan berbasis kepadatan yang paling populer adalah DBSCAN. Berbeda dengan banyak metode yang lebih baru, ini menampilkan model cluster yang didefinisikan dengan baik yang disebut "kepadatan-jangkauan". Mirip dengan pengelompokan berbasis tautan, ini didasarkan pada titik penghubung dalam ambang jarak tertentu. Namun, ini hanya menghubungkan titik-titik yang memenuhi kriteria kepadatan, dalam varian asli yang didefinisikan sebagai jumlah minimum objek lain dalam radius ini. Sebuah cluster terdiri dari semua objek yang terhubung dengan kepadatan (yang dapat membentuk cluster dalam bentuk arbitrer, berbeda dengan banyak metode lainnya) ditambah semua objek yang berada dalam rentang objek ini. Properti menarik lainnya dari DBSCAN adalah kompleksitasnya cukup rendah - ia memerlukan sejumlah kueri rentang linier pada database - dan ia akan menemukan hasil yang pada dasarnya sama (ini deterministik untuk titik inti dan gangguan, tetapi tidak untuk titik perbatasan) dalam setiap proses, oleh karena itu tidak perlu menjalankannya beberapa kali. OPTICS adalah generalisasi DB-SCAN yang menghilangkan kebutuhan untuk memilih nilai yang sesuai untuk parameter range, dan menghasilkan hasil hierarki yang terkait dengan linkage clustering. DeLi-Clu, Density-Link-Clustering menggabungkan ide-ide dari pengelompokan tautan tunggal dan OPTICS, menghilangkan parameter sepenuhnya dan menawarkan peningkatan kinerja atas OPTICS dengan menggunakan indeks R-tree.

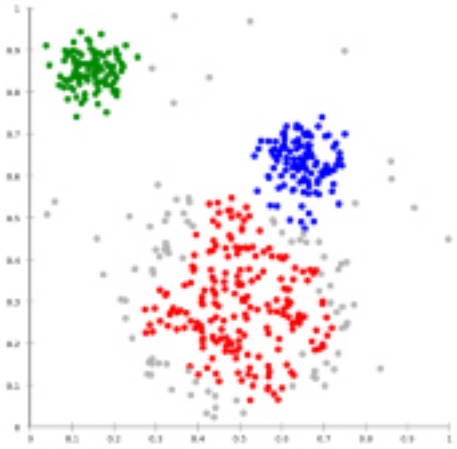
Kelemahan utama DBSCAN dan OPTICS adalah bahwa mereka mengharapkan semacam penurunan kepadatan untuk mendeteksi batas cluster. Selain itu, mereka tidak dapat mendeteksi struktur cluster intrinsik yang lazim di sebagian besar data kehidupan nyata. Variasi DBSCAN, EnDBSCAN, secara efisien mendeteksi jenis struktur seperti itu. Pada kumpulan data dengan, misalnya, distribusi Gaussian yang tumpang tindih - kasus penggunaan umum dalam data buatan - batas cluster yang dihasilkan oleh algoritme ini akan sering terlihat sewenang-wenang, karena kepadatan cluster terus menurun. Pada kumpulan data yang terdiri dari campuran Gaussians, algoritme ini hampir selalu mengungguli metode seperti pengelompokan EM yang mampu memodelkan data semacam ini secara tepat.

Pergeseran rata-rata adalah pendekatan pengelompokan di mana setiap objek dipindahkan ke area terpadat di sekitarnya, berdasarkan estimasi kepadatan kernel. Akhirnya, objek menyatu ke kepadatan maksimum lokal. Mirip dengan pengelompokan k-means, "penarik kepadatan" ini dapat berfungsi sebagai perwakilan untuk kumpulan data, tetapi pergeseran rata-rata dapat mendeteksi kluster berbentuk arbitrer yang mirip dengan DBSCAN. Karena prosedur iteratif dan estimasi kepadatan yang mahal, mean-shift biasanya lebih lambat daripada DBSCAN atau k-Means.

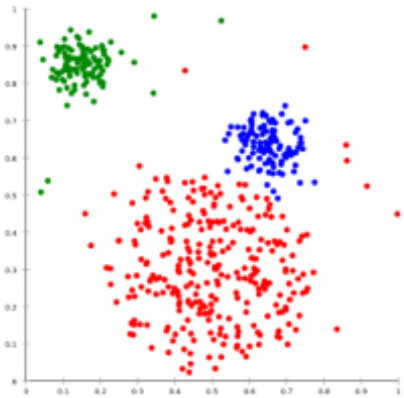
Contoh pengelompokan berbasis kepadatan



Density-based clustering *dengan DBSCAN*



DBSCAN mengasumsikan *cluster dengan kepadatan yang sama, dan mungkin memiliki masalah* dalam memisahkan *cluster terdekat*



OPTICS adalah *varian DBSCAN yang menangani kepadatan berbeda jauh lebih baik*

Perkembangan Terkini

Dalam beberapa tahun terakhir, banyak upaya telah dilakukan untuk meningkatkan kinerja algoritme yang ada. Diantaranya adalah CLARANS (Ng dan Han, 1994), dan BIRCH (Zhang et al., 1996). Dengan kebutuhan baru-baru ini untuk memproses kumpulan data yang lebih besar dan lebih besar (juga dikenal sebagai data besar), kesediaan untuk memperdagangkan makna semantik dari cluster yang dihasilkan untuk kinerja telah meningkat. Hal ini mengarah pada pengembangan metode pra-pengelompokan seperti pengelompokan kanopi, yang dapat memproses kumpulan data yang sangat besar secara efisien, tetapi “kluster” yang dihasilkan hanyalah prartisi kasar dari kumpulan data untuk kemudian menganalisis partisi dengan metode yang lebih lambat seperti sebagai k-means clustering. Berbagai pendekatan lain untuk pengelompokan telah dicoba seperti pengelompokan berbasis benih.

Untuk data berdimensi tinggi, banyak dari metode yang ada gagal karena kutukan dimensionalitas, yang membuat fungsi jarak tertentu bermasalah dalam ruang berdimensi tinggi. Hal ini menyebabkan algoritme pengelompokan baru untuk data berdimensi tinggi yang berfokus pada pengelompokan subruang (di mana hanya beberapa atribut yang digunakan, dan model kluster menyertakan atribut yang relevan untuk kluster) dan pengelompokan korelasi yang juga mencari subruang yang dirotasi sewenang-wenang ("berkorelasi") cluster yang dapat dimodelkan dengan memberikan korelasi atributnya. Contoh untuk algoritma pengelompokan tersebut adalah CLIQUE dan SUBCLU.

Ide dari metode pengelompokan berbasis kepadatan (khususnya keluarga algoritma DBSCAN / OPTICS) telah diadopsi ke pengelompokan subruang (HiSC, pengelompokan subruang hierarkis, dan DiSH) dan pengelompokan korelasi (HiCO, pengelompokan korelasi hierarki, 4C menggunakan "konektivitas korelasi Dan ERiC mengeksplorasi cluster korelasi berbasis kepadatan hirarkis).

Beberapa sistem pengelompokan yang berbeda berdasarkan informasi timbal balik telah diusulkan. Salah satunya adalah variasi metrik informasi Marina Meilă; yang lain menyediakan pengelompokan hierarki. Dengan menggunakan algoritma genetika, berbagai fungsi fit yang berbeda dapat dioptimalkan, termasuk informasi timbal balik. Juga algoritma penyampaian pesan, perkembangan terbaru dalam Ilmu Komputer dan Fisika Statistik, telah menyebabkan penciptaan jenis algoritma pengelompokan (*clustering algorithms*) baru.

Contoh lain adalah Skema Algoritme Sekuensial Dasar (BSAS). Basic Sequential Algorithmic Scheme (BSAS) merupakan salah satu metode analisis cluster yang paling dasar dan mudah digunakan. Theodoridis dkk (2003) mengungkapkan bahwa metode ini merupakan metode yang cepat dan mudah. Sequential clustering algorithms sendiri terbagi menjadi tiga jenis yakni Basic Sequential Algorithmic Scheme (BSAS), Modified Basic Sequential Algorithmic Scheme (MBSAS) dan A Two – Threshold Sequential Scheme. Dari ketiga jenis algoritma ini, yang paling dasar adalah BSAS. Dalam BSAS juga dapat mengatasi kelemahan dari metode hirarki dan non hirarki. Metode BSAS tidak memerlukan parameter jumlah cluster yang diinginkan. Parameter yang dibutuhkan dalam metode ini hanyalah batas nilai pengukuran jarak yang diizinkan.

Evaluasi dan Penilaian

Evaluasi hasil clustering terkadang disebut sebagai validasi cluster.
Ada beberapa saran untuk mengukur kesamaan antara dua clustering. Ukuran seperti itu dapat digunakan untuk membandingkan seberapa baik algoritma pengelompokan data yang berbeda bekerja pada sekumpulan data. Pengukuran ini biasanya terkait dengan jenis kriteria yang dipertimbangkan dalam menilai kualitas metode pengelompokan.

Evaluasi Internal

Jika hasil pengelompokan dievaluasi berdasarkan data yang dikelompokkan itu sendiri, ini disebut evaluasi internal. Metode ini biasanya memberikan skor terbaik pada algoritme yang menghasilkan cluster dengan kesamaan tinggi dalam suatu cluster dan kesamaan rendah antar cluster. Salah satu kelemahan menggunakan kriteria internal dalam evaluasi cluster adalah bahwa skor tinggi pada ukuran internal tidak selalu menghasilkan aplikasi pencarian informasi yang efektif. Selain itu, evaluasi ini bias terhadap algoritma yang menggunakan model cluster yang sama. Misalnya, pengelompokan k-Means secara alami mengoptimalkan jarak objek, dan kriteria internal berbasis jarak kemungkinan akan menilai pengelompokan yang dihasilkan secara berlebihan.

Oleh karena itu, langkah-langkah evaluasi internal paling cocok untuk mendapatkan beberapa wawasan tentang situasi di mana satu algoritma berkinerja lebih baik daripada yang lain, tetapi ini tidak berarti bahwa satu algoritma menghasilkan hasil yang lebih valid daripada yang lain. Validitas yang diukur dengan indeks semacam itu bergantung pada klaim bahwa jenis struktur ini ada dalam kumpulan data. Algoritme yang dirancang untuk beberapa jenis model tidak memiliki peluang jika kumpulan data berisi kumpulan model yang sangat berbeda, atau jika evaluasi mengukur kriteria yang sangat berbeda.

Misalnya, pengelompokan k-means hanya dapat menemukan kluster konveks, dan banyak indeks evaluasi mengasumsikan kluster konveks. Pada kumpulan data dengan cluster non-konveks, baik penggunaan k-means, maupun kriteria evaluasi yang mengasumsikan konveks, tidak baik.

Metode berikut dapat digunakan untuk menilai kualitas algoritma pengelompokan berdasarkan kriteria internal:

- Indeks Davies – Bouldin

  Indeks Davies – Bouldin dapat dihitung dengan rumus berikut:

$$DB = \frac{1}{n}\sum_{i=1}^{n} \max_{j \neq i}\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right)$$

  dimana n adalah banyaknya cluster, c_x adalah centroid cluster x, $\boldsymbol{\sigma}_x$ adalah jarak rata-rata semua elemen pada cluster x ke centroid c_x, dan x (c_i , c_j) adalah jarak antara centroids c_i dan c_j. Karena algoritma yang menghasilkan cluster dengan jarak intra-cluster rendah (kesamaan intra-cluster tinggi) dan jarak antar cluster tinggi (kesamaan antar cluster rendah) akan memiliki indeks Davies-Bouldin yang rendah, algoritma clustering yang menghasilkan kumpulan cluster dengan indeks Davies-Bouldin terkecil dianggap sebagai algoritma terbaik berdasarkan kriteria ini.

- Indeks Dunn

  Indeks Dunn bertujuan untuk mengidentifikasi cluster yang padat dan terpisah dengan baik. Ini didefinisikan sebagai rasio antara jarak antar klaster minimal dengan jarak maksimal intra klaster. Untuk setiap partisi cluster, indeks Dunn dapat dihitung dengan rumus berikut:

$$D = \frac{\min_{1 \le i < j \le n} d(i, j)}{\max_{1 \le k \le n} d'(k)}$$

  dimana d (i, j) mewakili jarak antara cluster i dan j, dan d '(k) mengukur jarak intra-cluster dari cluster k. Jarak antar cluster d (i, j) antara dua cluster dapat berupa sejumlah ukuran jarak, seperti jarak antara centroid cluster. Demikian pula, jarak intra-cluster d '(k) dapat diukur dengan berbagai cara, seperti jarak maksimal antara pasangan elemen dalam cluster k. Karena kriteria internal mencari cluster dengan kesamaan intra-cluster yang tinggi dan kesamaan antar-cluster yang rendah, algoritme yang menghasilkan cluster dengan indeks Dunn tinggi lebih diinginkan.

- Koefisien siluet

  Koefisien siluet mengontraskan jarak rata-rata ke elemen dalam kluster yang sama dengan jarak rata-rata ke elemen di kluster lain. Objek dengan nilai siluet tinggi

dianggap terkluster dengan baik, objek dengan nilai rendah mungkin merupakan pencilan. Indeks ini bekerja dengan baik dengan k-means clustering, dan juga digunakan untuk menentukan jumlah cluster yang optimal.

Evaluasi Eksternal

Dalam evaluasi eksternal, hasil pengelompokan dievaluasi berdasarkan data yang tidak digunakan untuk pengelompokan, seperti label kelas yang diketahui dan tolok ukur eksternal. Tolok ukur tersebut terdiri dari satu set item yang telah diklasifikasikan sebelumnya, dan set ini sering dibuat oleh manusia (ahli). Dengan demikian, kumpulan patokan dapat dianggap sebagai standar emas untuk evaluasi. Jenis metode evaluasi ini mengukur seberapa dekat pengelompokan dengan kelas benchmark yang telah ditentukan. Namun, baru-baru ini telah dibahas apakah ini memadai untuk data nyata, atau hanya pada kumpulan data sintetis dengan kebenaran dasar faktual, karena kelas dapat berisi struktur internal, atribut yang ada mungkin tidak memungkinkan pemisahan cluster atau kelas mungkin berisi anomali. Selain itu, dari sudut pandang penemuan pengetahuan, reproduksi pengetahuan yang diketahui belum tentu merupakan hasil yang diinginkan. Dalam skenario khusus pengelompokan terbatas, di mana informasi meta (seperti label kelas) sudah digunakan dalam proses pengelompokan, penahanan informasi untuk tujuan evaluasi tidak sepele.

Sejumlah ukuran diadaptasi dari varian yang digunakan untuk mengevaluasi tugas klasifikasi. Sebagai ganti penghitungan berapa kali kelas ditetapkan dengan benar ke satu titik data (dikenal sebagai positif benar), metrik penghitungan pasangan tersebut menilai apakah setiap pasangan titik data yang benar-benar dalam kluster yang sama diprediksi berada dalam kluster yang sama pula.

Sebagai ganti penghitungan berapa kali kelas ditetapkan dengan benar ke satu titik data (dikenal sebagai positif benar), metrik penghitungan pasangan tersebut menilai apakah setiap pasangan titik data yang benar-benar dalam kluster yang sama diprediksi berada di tempat yang sama. gugus.

Beberapa ukuran kualitas algoritma kluster menggunakan kriteria eksternal meliputi:

- Rand Measure (William M. Rand 1971)
  Indeks Rand menghitung seberapa mirip cluster (dikembalikan oleh algoritma clustering) dengan klasifikasi benchmark. Seseorang juga dapat melihat indeks Rand

sebagai ukuran persentase keputusan yang benar yang dibuat oleh algoritma. Itu dapat dihitung menggunakan rumus berikut:

$$RI = \frac{TP + TN}{TP + FP + FN + TN}$$

dimana *TP* adalah jumlah positif benar, *TN* adalah jumlah negatif benar, *FP* adalah jumlah positif palsu, dan *FN* adalah jumlah negatif palsu. Satu masalah dengan indeks Rand adalah bahwa positif palsu dan negatif palsu memiliki bobot yang sama. Ini mungkin merupakan karakteristik yang tidak diinginkan untuk beberapa aplikasi pengelompokan. Pengukuran-*F* menangani masalah ini, seperti halnya indeks Rand yang disesuaikan dengan koreksi peluang.

- Pengukuran-F

  Pengukuran-F dapat digunakan untuk menyeimbangkan kontribusi negatif palsu dengan menimbang recall melalui parameter 0.β ≥ Biarkan presisi dan recall didefinisikan sebagai berikut:

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

  di mana P adalah tingkat presisi dan R adalah tingkat penarikan. Kita dapat menghitung ukuran-F dengan menggunakan rumus berikut:

$$F_\beta = \frac{(\beta^2 + 1) \cdot P \cdot R}{\beta^2 \cdot P + R}$$

  Perhatikan bahwa ketika β = 0, $P_0 = P$. Dengan kata lain, mengingat tidak memiliki dampak pada pengukuran-F ketika β = 0 dan meningkatkan β mengalokasikan jumlah berat yang meningkat untuk mengingat dalam pengukuran-F akhir.

- Indeks Jaccard

  Indeks Jaccard digunakan untuk mengukur kesamaan antara dua kumpulan data. Indeks Jaccard mengambil nilai antara 0 dan 1. Indeks 1 berarti bahwa dua dataset identik, dan indeks 0 menunjukkan bahwa dataset tidak memiliki elemen yang sama. Indeks Jaccard ditentukan dengan rumus berikut:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN}$$

  Ini hanyalah jumlah elemen unik yang umum untuk kedua set dibagi dengan jumlah total elemen unik di kedua set.

- Indeks Fowlkes – Mallows (E. B. Fowlkes & C. L. Mallows 1983)
  Indeks Fowlkes-Mallows menghitung kesamaan antara cluster yang dikembalikan oleh algoritma clustering dan klasifikasi benchmark. Semakin tinggi nilai indeks Fowlkes-Mallows semakin mirip klaster dan klasifikasi benchmarknya. Itu dapat dihitung menggunakan rumus berikut:

$$FM = \sqrt{\frac{TP}{TP+FP} \cdot \frac{TP}{TP+FN}}$$

  dimana TP adalah jumlah positif benar, FP adalah jumlah positif palsu, dan FN adalah jumlah negatif palsu. Indeks FM adalah rata-rata geometrik dari presisi dan recall P dan R, sedangkan F-measure adalah nilai harmoniknya. Selain itu, presisi dan panggilan ulang juga dikenal sebagai indeks Wallace B^{I} dan B^{II}.
- Mutual Information adalah ukuran teoritis informasi tentang seberapa banyak informasi dibagi antara pengelompokan dan klasifikasi kebenaran dasar yang dapat mendeteksi kesamaan non-linier antara dua pengelompokan. Informasi timbal balik yang disesuaikan adalah varian yang dikoreksi untuk peluang ini yang memiliki bias yang berkurang untuk berbagai nomor cluster.
- Matriks Konfusi
  Matriks konfusi dapat digunakan untuk memvisualisasikan dengan cepat hasil algoritme klasifikasi (atau pengelompokan). Ini menunjukkan betapa berbedanya sebuah cluster dari cluster standar emas.

Bidang Penerapan

Biologi, biologi komputasi dan bioinformatika

Ekologi tumbuhan dan hewan

analisis cluster digunakan untuk mendeskripsikan dan membuat perbandingan spasial dan temporal dari komunitas (kumpulan) organisme di lingkungan yang heterogen; itu juga digunakan dalam sistematika tumbuhan untuk menghasilkan filogeni buatan atau kelompok organisme (individu) pada spesies, genus atau tingkat yang lebih tinggi yang memiliki sejumlah atribut

Transcriptomics

pengelompokan digunakan untuk membangun kelompok gen dengan pola ekspresi terkait (juga dikenal sebagai gen yang diekspresikan bersama) seperti dalam algoritme pengelompokan SKT. Seringkali kelompok seperti itu mengandung protein yang berhubungan secara fungsional, seperti enzim untuk jalur tertentu, atau gen yang diatur bersama. Eksperimen throughput tinggi menggunakan tag urutan yang diekspresikan

(EST) atau mikroarray DNA dapat menjadi alat yang ampuh untuk anotasi genom, aspek umum genomik.

Analisis urutan (Sequence Analysis)

clustering digunakan untuk mengelompokkan urutan homolog ke dalam keluarga gen. Ini adalah konsep yang sangat penting dalam bioinformatika, dan biologi evolusi pada umumnya.

Platform genotipe throughput tinggi

algoritma pengelompokan digunakan untuk menetapkan genotipe secara otomatis.

Pengelompokan genetik manusia

Kesamaan data genetik digunakan dalam pengelompokan untuk menyimpulkan struktur populasi.

Pengobatan

Pencitraan medis

Pada pemindaian PET, analisis kelompok dapat digunakan untuk membedakan berbagai jenis jaringan dalam gambar tiga dimensi untuk berbagai tujuan.

Bisnis dan pemasaran

Riset pasar

Analisis cluster secara luas digunakan dalam riset pasar ketika bekerja dengan data multivariat dari survei dan panel uji. Peneliti pasar menggunakan analisis kluster untuk membagi populasi umum konsumen ke dalam segmen pasar dan untuk lebih memahami hubungan antara berbagai kelompok konsumen / pelanggan potensial, dan untuk digunakan dalam segmentasi pasar, penentuan posisi produk, pengembangan produk baru dan pemilihan pasar uji.

Pengelompokan barang belanjaan

Clustering dapat digunakan untuk mengelompokkan semua item belanja yang tersedia di web ke dalam satu set produk unik. Misalnya, semua barang di eBay dapat dikelompokkan menjadi produk unik. (eBay tidak memiliki konsep SKU)

World Wide Web

Analisis jejaring sosial

Dalam studi jejaring sosial, pengelompokan dapat digunakan untuk mengenali komunitas dalam kelompok besar orang.

Pengelompokan hasil pencarian

Dalam proses pengelompokan file dan situs web yang cerdas, pengelompokan dapat digunakan untuk membuat hasil pencarian yang lebih relevan dibandingkan dengan mesin pencari biasa seperti Google. Saat ini ada sejumlah alat pengelompokan berbasis web seperti Clusty.

Optimasi slippy map

Peta foto Flickr dan situs peta lainnya menggunakan pengelompokan untuk mengurangi jumlah penanda pada peta. Ini membuatnya lebih cepat dan mengurangi jumlah kekacauan visual.

Ilmu Komputer

Evolusi perangkat lunak

Pengelompokan berguna dalam evolusi perangkat lunak karena membantu mengurangi properti warisan dalam kode dengan mereformasi fungsionalitas yang telah tersebar. Ini adalah bentuk restrukturisasi dan karenanya merupakan cara pemeliharaan preventif langsung.

Segmentasi gambar

Clustering dapat digunakan untuk membagi gambar digital menjadi beberapa wilayah berbeda untuk deteksi perbatasan atau pengenalan objek.

Algoritme evolusioner

Pengelompokan dapat digunakan untuk mengidentifikasi relung yang berbeda dalam populasi algoritme evolusi sehingga peluang reproduksi dapat didistribusikan lebih merata di antara spesies atau subspesies yang berkembang.

Sistem pemberi rekomendasi

Sistem pemberi rekomendasi dirancang untuk merekomendasikan item baru berdasarkan selera pengguna. Mereka terkadang menggunakan algoritme pengelompokan untuk memprediksi preferensi pengguna berdasarkan preferensi pengguna lain di kluster pengguna.

Metode Markov chain Monte Carlo

Clustering sering digunakan untuk mencari dan mengkarakterisasi ekstrema dalam distribusi target.

Deteksi anomali

Anomali / pencilan biasanya - baik secara eksplisit maupun implisit - didefinisikan sehubungan dengan struktur pengelompokan dalam data.

Ilmu Sosial dan Kemasyarakatan

Analisis kejahatan

Analisis cluster dapat digunakan untuk mengidentifikasi area-area di mana terdapat insiden yang lebih besar dari jenis kejahatan tertentu. Dengan mengidentifikasi area-area berbeda atau "hot spot" di mana kejahatan serupa telah terjadi selama periode waktu tertentu, dimungkinkan untuk mengelola sumber daya penegakan hukum lebih lanjut secara efektif.

Penambangan data pendidikan

Analisis cluster misalnya digunakan untuk mengidentifikasi kelompok sekolah atau siswa dengan sifat yang serupa.

Tipologi

Dari data jajak pendapat, proyek-proyek seperti yang dilakukan oleh *Pew Research Center* menggunakan analisis kluster untuk membedakan tipologi pendapat, kebiasaan, dan demografi yang mungkin berguna dalam politik dan pemasaran.

Penggunaan dan Penerapan Bidang Lainnya

Robotika

Algoritma pengelompokan digunakan untuk kesadaran situasional robotik untuk melacak objek dan mendeteksi pencilan dalam data sensor.

Kimia matematika

Untuk mencari kemiripan struktur, dll., Misalnya, 3000 senyawa kimia dikelompokkan dalam ruang 90 indeks topologi.

Klimatologi

Untuk menemukan rezim cuaca atau pola atmosfer tekanan permukaan laut yang disukai.

Geologi perminyakan

Analisis cluster digunakan untuk merekonstruksi data inti lubang bawah yang hilang atau kurva log yang hilang untuk mengevaluasi properti reservoir.

Geografi fisik

Pengelompokan sifat kimia di lokasi sampel yang berbeda.

3.5 Klasifikasi Statistik

Dalam pembelajaran mesin dan statistik, klasifikasi adalah masalah untuk mengidentifikasi kumpulan kategori (sub-populasi) yang mana dari pengamatan baru itu, berdasarkan kumpulan data pelatihan yang berisi pengamatan (atau contoh) yang keanggotaan kategorinya diketahui.

Contohnya adalah menetapkan email tertentu ke dalam kelas "spam" atau "non-spam" atau menetapkan diagnosis untuk pasien tertentu seperti yang dijelaskan oleh karakteristik pasien yang diamati (jenis kelamin, tekanan darah, ada atau tidak adanya gejala tertentu, dll.). Klasifikasi adalah contoh pengenalan pola.

Dalam terminologi pembelajaran mesin, klasifikasi dianggap sebagai contoh pembelajaran yang diawasi, yaitu pembelajaran di mana kumpulan pelatihan pengamatan yang diidentifikasi dengan benar tersedia. Prosedur tanpa pengawasan yang sesuai dikenal sebagai pengelompokan, dan melibatkan pengelompokan data ke dalam kategori berdasarkan beberapa ukuran kesamaan atau jarak yang melekat.

Seringkali, pengamatan individu dianalisis menjadi satu set properti yang dapat diukur, yang dikenal dengan berbagai variasi sebagai variabel atau fitur penjelas. Properti ini dapat bermacam-macam kategori (misalnya "A", "B", "AB" atau "O", untuk golongan darah), ordinal (misalnya "besar", "sedang" atau "kecil"), bernilai bilangan bulat (misalnya jumlah kemunculan kata tertentu dalam email) atau nilai riil (misalnya pengukuran tekanan darah). Pengklasifikasi lain bekerja dengan membandingkan pengamatan dengan pengamatan sebelumnya dengan menggunakan fungsi kesamaan atau jarak.

Algoritme yang mengimplementasikan klasifikasi, terutama dalam implementasi konkret, dikenal sebagai pengklasifikasi. Istilah "pengklasifikasi" terkadang juga mengacu pada fungsi matematika, yang diterapkan oleh algoritme klasifikasi, yang memetakan data masukan ke suatu kategori.

Terminologi lintas bidang cukup beragam. Dalam statistik, di mana klasifikasi sering dilakukan dengan regresi logistik atau prosedur serupa, sifat pengamatan disebut variabel penjelas (atau variabel independen, regressor, dll.), Dan kategori yang akan diprediksi dikenal sebagai hasil, yang dianggap sebagai menjadi nilai yang mungkin dari variabel dependen. Dalam pembelajaran mesin, observasi sering disebut sebagai contoh, variabel penjelas disebut fitur (dikelompokkan ke dalam vektor fitur), dan kategori yang mungkin

untuk diprediksi adalah kelas. Bidang lain mungkin menggunakan terminologi berbeda: mis. dalam ekologi komunitas, istilah "klasifikasi" biasanya mengacu pada analisis cluster, yaitu jenis pembelajaran tanpa pengawasan, bukan pembelajaran yang diawasi yang dijelaskan dalam artikel ini.

Keterkaitan antara Klasifikasi dan Clustering

Klasifikasi dan pengelompokan adalah contoh dari masalah pengenalan pola yang lebih umum, yang merupakan penugasan semacam nilai keluaran ke nilai masukan yang diberikan. Contoh lainnya adalah regresi, yang memberikan keluaran bernilai nyata untuk setiap masukan; pelabelan urutan, yang menetapkan kelas ke setiap anggota urutan nilai (misalnya, penandaan bagian ucapan, yang menetapkan bagian ucapan untuk setiap kata dalam kalimat masukan); parsing, yang menetapkan pohon parse ke kalimat masukan, yang menjelaskan struktur sintaksis kalimat; dll.

Subkelas klasifikasi yang umum adalah klasifikasi probabilistik. Algoritme seperti ini menggunakan inferensi statistik untuk menemukan kelas terbaik untuk instance tertentu. Tidak seperti algoritme lain, yang hanya mengeluarkan kelas "terbaik", algoritme probabilistik mengeluarkan probabilitas instans menjadi anggota dari setiap kelas yang memungkinkan. Kelas terbaik biasanya kemudian dipilih sebagai kelas dengan probabilitas tertinggi. Namun, algoritme semacam itu memiliki banyak keunggulan dibandingkan pengklasifikasi non-probabilistik:

- Dapat mengeluarkan nilai kepercayaan yang terkait dengan pilihannya (secara umum, pengklasifikasi yang dapat melakukan ini dikenal sebagai pengklasifikasi berbobot kepercayaan).
- Sejalan dengan itu, ia dapat melakukan abstain ketika keyakinannya untuk memilih keluaran tertentu terlalu rendah.
- Karena probabilitas yang dihasilkan, pengklasifikasi probabilistik dapat lebih efektif digabungkan ke dalam tugas pembelajaran mesin yang lebih besar, dengan cara yang sebagian atau seluruhnya menghindari masalah penyebaran kesalahan.

Prosedur Frequentist

Pekerjaan awal pada klasifikasi statistik kenalkan oleh Fisher, dalam konteks masalah dua kelompok, yang mengarah ke fungsi diskriminan linier Fisher sebagai aturan untuk menetapkan kelompok ke pengamatan baru. Pekerjaan awal ini mengasumsikan bahwa

nilai data dalam masing-masing dari dua kelompok memiliki distribusi normal multi-variat. Perluasan konteks yang sama ini ke lebih dari dua grup juga telah dipertimbangkan dengan pembatasan yang diberlakukan bahwa aturan klasifikasi harus linier. Pekerjaan selanjutnya untuk distribusi normal multivariat memungkinkan pengklasifikasi menjadi nonlinier: beberapa aturan klasifikasi dapat diturunkan berdasarkan penyesuaian yang sedikit berbeda dari jarak Mahalanobis, dengan pengamatan baru yang ditugaskan pada kelompok yang pusatnya memiliki jarak penyesuaian terendah dari pengamatan.

Prosedur Bayesian

Tidak seperti prosedur frequentist, prosedur klasifikasi Bayesian memberikan cara alami untuk memperhitungkan informasi yang tersedia tentang ukuran relatif sub-populasi yang terkait dengan kelompok berbeda dalam keseluruhan populasi. Prosedur Bayesian cenderung mahal secara komputasi dan, pada hari-hari sebelum penghitungan Monte Carlo rantai Markov dikembangkan, perkiraan untuk aturan pengelompokan Bayesian dibuat.

Beberapa prosedur Bayesian melibatkan penghitungan probabilitas keanggotaan grup: ini dapat dilihat sebagai memberikan hasil analisis data yang lebih informatif daripada atribusi sederhana dari label grup tunggal ke setiap observasi baru.

Klasifikasi Biner dan Multiclass

Klasifikasi dapat dianggap sebagai dua masalah terpisah - klasifikasi biner dan klasifikasi multikelas. Dalam klasifikasi biner, tugas yang lebih dipahami, hanya dua kelas yang terlibat, sedangkan klasifikasi multikelas melibatkan penugasan objek ke salah satu dari beberapa kelas. Karena banyak metode klasifikasi telah dikembangkan secara khusus untuk klasifikasi biner, klasifikasi multikelas sering kali memerlukan penggunaan gabungan dari beberapa pengklasifikasi biner.

Vektor Fitur

Sebagian besar algoritme mendeskripsikan instance individu yang kategorinya akan diprediksi menggunakan vektor fitur dari properti individual yang dapat diukur dari instance tersebut. Setiap properti disebut fitur, juga dikenal dalam statistik sebagai variabel penjelas (atau variabel independen, meskipun fitur mungkin atau mungkin tidak independen secara statistik). Berbagai fitur dapat berupa biner (mis. "Pria" atau "wanita"); kategorikal (misalnya, "A", "B", "AB" atau "O", untuk golongan darah); ordinal (misalnya "besar", "sedang", atau

"kecil"); nilai integer (misalnya jumlah kemunculan kata tertentu dalam email); atau nilai riil (misalnya pengukuran tekanan darah). Jika instance adalah gambar, nilai fitur mungkin sesuai dengan piksel gambar; jika instance adalah sepotong teks, nilai fitur mungkin merupakan frekuensi kemunculan kata-kata yang berbeda. Beberapa algoritme hanya bekerja dalam hal data diskrit dan memerlukan nilai riil atau bilangan bulat.

Pengklasifikasi Linier

Sejumlah besar algoritme untuk klasifikasi dapat disusun dalam bentuk fungsi linier yang memberikan skor ke setiap kemungkinan kategori k dengan menggabungkan vektor fitur dari sebuah instance dengan vektor bobot, menggunakan perkalian titik. Kategori yang diprediksi adalah kategori dengan skor tertinggi. Jenis fungsi skor ini dikenal sebagai fungsi prediktor linier dan memiliki bentuk umum berikut:

$$score(\mathbf{X}_i, k) = \beta_k \cdot \mathbf{X}_i$$

dimana X adalah vektor fitur misalnya i, β adalah vektor bobot yang sesuai dengan kategori k, dan skor (X, k) adalah skor yang terkait dengan menetapkan instans i ke kategori k. Dalam teori pilihan diskrit, di mana contoh mewakili orang dan kategori mewakili pilihan, skor dianggap utilitas yang terkait dengan orang yang saya pilih kategori k.

Algoritma dengan pengaturan dasar ini dikenal sebagai pengklasifikasi linier. Yang membedakan mereka adalah prosedur untuk menentukan (melatih) bobot / koefisien yang optimal dan cara menafsirkan skor tersebut.
Contoh dari algoritma tersebut adalah

- Regresi logistik dan regresi logistik multinomial
- Regresi Probit
- Algoritma perceptron
- Mendukung mesin vektor
- Analisis diskriminan linier.

Algoritma

Contoh algoritma klasifikasi meliputi:

- Pengklasifikasi linier
- Diskriminan linier Fisher
- Regresi logistik
- Pengklasifikasi Naive Bayes

- Perceptron
- Support vector machines
- Least squares support vector machines
- Quadratic classifiers
- Kernel estimation
- k-nearest neighbor
- Boosting (meta-algorithm)
- Decision trees
- Random forests
- Neural networks
- FMM Neural Networks
- Learning vector quantization

Evaluasi

Kinerja classifier sangat tergantung pada karakteristik data yang akan diklasifikasikan. Tidak ada pengklasifikasi tunggal yang bekerja paling baik pada semua masalah yang diberikan (sebuah fenomena yang dapat dijelaskan oleh teorema tanpa makan siang). Berbagai tes empiris telah dilakukan untuk membandingkan kinerja classifier dan untuk menemukan karakteristik data yang menentukan kinerja classifier. Namun, menentukan pengelompokan yang cocok untuk masalah tertentu masih lebih merupakan seni daripada sains.

Pengukuran presisi dan penarikan kembali adalah metrik populer yang digunakan untuk mengevaluasi kualitas sistem klasifikasi. Baru-baru ini, kurva karakteristik operasi penerima (ROC) telah digunakan untuk mengevaluasi pertukaran antara tingkat benar dan salah-positif dari algoritma klasifikasi.
Sebagai metrik kinerja, koefisien ketidakpastian memiliki keunggulan dibandingkan akurasi sederhana karena tidak dipengaruhi oleh ukuran relatif dari kelas yang berbeda. Lebih lanjut, itu tidak akan menghukum algoritma karena hanya mengatur ulang kelas.

Domain Aplikasi
Klasifikasi memiliki banyak aplikasi. Dalam beberapa hal ini digunakan sebagai prosedur penggalian data, sementara yang lain dilakukan pemodelan statistik yang lebih rinci.

- Visi komputer
  - Pencitraan medis dan analisis citra medis
  - Pengenalan karakter optik
  - Pelacakan video

- Penemuan dan pengembangan obat
  - Toksikogenomik
  - Hubungan struktur-aktivitas kuantitatif
- Geostatistik
- Pengenalan suara
- Pengenalan tulisan tangan
- Identifikasi biometrik
- Klasifikasi biologis
- Pemrosesan bahasa alami secara statistik
- Klasifikasi dokumen
- Mesin pencari internet
- Penilaian kredit
- Pengenalan pola
- Klasifikasi larik mikro

3.6 Analisis Regresi

Dalam pemodelan statistik, analisis regresi adalah proses statistik untuk memperkirakan hubungan antar variabel. Ini mencakup banyak teknik untuk memodelkan dan menganalisis beberapa variabel, ketika fokusnya adalah pada hubungan antara variabel dependen dan satu atau lebih variabel independen (atau 'prediktor'). Lebih khusus lagi, analisis regresi membantu seseorang memahami bagaimana nilai khas dari variabel dependen (atau 'variabel kriteria') berubah ketika salah satu variabel independen bervariasi, sedangkan variabel independen lainnya ditetapkan tetap. Paling umum, analisis regresi memperkirakan ekspektasi bersyarat dari variabel dependen yang diberikan variabel dependen - yaitu, nilai rata-rata variabel dependen ketika variabel independen ditetapkan. Lebih jarang, fokusnya adalah pada kuantil, atau parameter lokasi lain dari distribusi bersyarat dari variabel dependen yang diberikan variabel independen. Dalam semua kasus, target estimasi adalah fungsi dari variabel independen yang disebut fungsi regresi. Dalam analisis regresi, juga menarik untuk mengkarakterisasi variasi variabel dependen di sekitar fungsi regresi yang dapat dijelaskan dengan distribusi probabilitas. Pendekatan yang terkait tetapi berbeda diperlukan analisis kondisi (NCA), yang memperkirakan nilai maksimum (bukan rata-rata) dari variabel dependen untuk nilai tertentu dari variabel independen (garis langit-langit daripada garis tengah) untuk mengidentifikasi nilai dari variabel dependen. variabel independen diperlukan tetapi tidak cukup untuk nilai tertentu dari variabel dependen.

Analisis regresi banyak digunakan untuk prediksi dan perkiraan, yang penggunaannya secara substansial tumpang tindih dengan bidang pembelajaran mesin. Analisis regresi juga digunakan untuk memahami variabel independen mana yang terkait dengan variabel dependen, dan untuk mengeksplorasi bentuk-bentuk hubungan tersebut. Dalam keadaan terbatas, analisis regresi dapat digunakan untuk menyimpulkan hubungan kausal antara variabel independen dan dependen. Namun hal ini dapat menyebabkan ilusi atau hubungan yang salah, jadi berhati-hatilah; misalnya, korelasi tidak berarti sebab-akibat.

Banyak teknik untuk melakukan analisis regresi telah dikembangkan. Metode yang sudah dikenal seperti regresi linier dan regresi kuadrat terkecil biasa adalah parametrik, di mana fungsi regresi didefinisikan dalam istilah sejumlah parameter yang tidak diketahui yang diperkirakan dari data. Regresi nonparametrik mengacu pada teknik yang memungkinkan fungsi regresi berada dalam sekumpulan fungsi tertentu, yang mungkin berdimensi tak hingga.

Kinerja metode analisis regresi dalam praktiknya bergantung pada bentuk proses penghasil data, dan bagaimana hubungannya dengan pendekatan regresi yang digunakan. Karena bentuk sebenarnya dari proses penghasil data umumnya tidak diketahui, analisis regresi seringkali bergantung pada pembuatan asumsi tentang proses ini. Asumsi ini terkadang dapat diuji jika tersedia cukup data. Model regresi untuk prediksi sering kali berguna bahkan ketika asumsi dilanggar secara moderat, meskipun performanya mungkin tidak optimal. Namun, dalam banyak penerapan, terutama dengan efek kecil atau pertanyaan kausalitas berdasarkan data observasi, metode regresi dapat memberikan hasil yang menyesatkan.

Dalam arti yang lebih sempit, regresi dapat merujuk secara khusus pada estimasi variabel respon kontinu, sebagai lawan dari variabel respon diskrit yang digunakan dalam klasifikasi. Kasus variabel keluaran berkelanjutan mungkin lebih spesifik disebut sebagai regresi metrik untuk membedakannya dari masalah terkait.

Sejarah

Bentuk paling awal dari regresi adalah metode kuadrat terkecil, yang diterbitkan oleh Legendre pada tahun 1805, dan oleh Gauss pada tahun 1809. Legendre dan Gauss sama-sama menerapkan metode tersebut pada masalah penentuan, dari pengamatan astronomi, orbit benda-benda di sekitar Matahari (kebanyakan komet, tetapi juga kemudian

planet minor yang baru ditemukan). Gauss menerbitkan perkembangan lebih lanjut dari teori kuadrat terkecil pada tahun 1821, termasuk versi teorema Gauss-Markov.

Istilah "regresi" diciptakan oleh Francis Galton pada abad kesembilan belas untuk menggambarkan fenomena biologis. Fenomena yang terjadi adalah tinggi badan keturunan dari nenek moyang yang tinggi cenderung menurun menuju rata-rata normal (fenomena yang juga dikenal sebagai regresi menuju mean). Bagi Galton, regresi hanya memiliki makna biologis ini, tetapi karyanya kemudian diperluas oleh Udny Yule dan Karl Pearson ke konteks statistik yang lebih umum. Dalam penelitian Yule dan Pearson, distribusi gabungan dari respon dan variabel penjelas diasumsikan sebagai Gaussian. Ini

asumsi dilemahkan oleh R.A. Fisher dalam karyanya tahun 1922 dan 1925. Fisher berasumsi bahwa distribusi bersyarat dari variabel respon adalah Gaussian, tetapi distribusi gabungannya tidak perlu. Dalam hal ini, asumsi Fisher lebih dekat dengan rumusan Gauss tahun 1821.

Pada 1950-an dan 1960-an, para ekonom menggunakan kalkulator meja elektromekanis untuk menghitung regresi. Sebelum tahun 1970, terkadang butuh waktu hingga 24 jam untuk menerima hasil dari satu regresi.

Metode regresi terus menjadi bidang penelitian aktif. Dalam beberapa dekade terakhir, metode baru telah dikembangkan untuk regresi yang kuat, regresi yang melibatkan respons berkorelasi seperti deret waktu dan kurva pertumbuhan, regresi di mana prediktor (variabel independen) atau variabel respons adalah kurva, gambar, grafik, atau objek data kompleks lainnya, metode regresi yang menampung berbagai jenis data yang hilang, regresi nonparametrik, metode regresi Bayesian, regresi di mana variabel prediktor diukur dengan kesalahan, regresi dengan variabel prediktor lebih banyak daripada observasi, dan kesimpulan kausal dengan regresi..

Model Regresi

Model regresi melibatkan variabel berikut:

- Parameter yang tidak diketahui, dilambangkan sebagai β, yang mungkin mewakili skalar atau vektor.
- Variabel independen, X.
- Variabel terikat, Y.

Dalam berbagai bidang aplikasi, terminologi yang berbeda digunakan untuk menggantikan variabel dependen dan independen.

Model regresi menghubungkan Y dengan fungsi X dan β.

$$Y \approx f(X, \beta)$$

Perkiraan biasanya diformalkan sebagai $E(Y \mid X) = f(X, \beta)$. Untuk melakukan analisis regresi, bentuk dari fungsi f harus ditentukan. Terkadang bentuk fungsi ini didasarkan pada pengetahuan tentang hubungan antara Y dan Y yang tidak mengandalkan data. Jika pengetahuan seperti itu tidak tersedia, bentuk yang fleksibel atau nyaman untuk f dipilih.

Asumsikan sekarang bahwa vektor parameter yang tidak diketahui β memiliki panjang k. Untuk melakukan analisis regresi, pengguna harus memberikan informasi tentang variabel dependen Y:

- Jika N titik data dalam bentuk (Y, X) diamati, di mana $N < k$, sebagian besar pendekatan klasik untuk analisis regresi tidak dapat dilakukan: karena sistem persamaan yang mendefinisikan model regresi kurang ditentukan, tidak ada cukup data untuk dipulihkan β.
- Jika tepat N = k titik data diamati, dan fungsi f linier, persamaan Y = f (X, β) dapat diselesaikan dengan tepat daripada perkiraan. Ini mereduksi menjadi menyelesaikan satu set persamaan N dengan N yang tidak diketahui (elemen β), yang memiliki solusi unik selama X bebas linier. Jika f nonlinier, solusi mungkin tidak ada, atau banyak solusi mungkin ada.
- Situasi yang paling umum adalah saat $N > k$ titik data diamati. Dalam hal ini, terdapat cukup informasi dalam data untuk memperkirakan nilai unik untuk β yang paling sesuai dengan data dalam beberapa hal, dan model regresi bila diterapkan pada data dapat dilihat sebagai sistem yang ditentukan secara berlebihan dalam β.

Dalam kasus terakhir, analisis regresi menyediakan alat untuk:

1. Menemukan solusi untuk parameter yang tidak diketahui β yang akan, misalnya, meminimalkan jarak antara nilai terukur dan prediksi dari variabel dependen Y (juga dikenal sebagai metode kuadrat terkecil).
2. Berdasarkan asumsi statistik tertentu, analisis regresi menggunakan informasi surplus untuk memberikan informasi statistik tentang parameter β yang tidak diketahui dan nilai prediksi dari variabel dependen Y.

3.6.3 Jumlah Pengukuran Independen yang Diperlukan

Misalnya model regresi yang memiliki tiga parameter yang tidak diketahui, β_0, β_1, dan β_2. Misalkan seorang pelaku eksperimen melakukan 10 pengukuran semua pada nilai yang sama persis dari variabel independen vektor X (yang berisi variabel independen X_1, X_2, dan X_3). Dalam kasus ini, analisis regresi gagal memberikan satu set unik nilai taksiran untuk tiga parameter yang tidak diketahui; pelaku eksperimen tidak memberikan informasi yang cukup. Yang terbaik yang dapat dilakukan adalah memperkirakan nilai rata-rata dan deviasi standar variabel dependen Y. Demikian pula, mengukur pada dua nilai X yang berbeda akan memberikan data yang cukup untuk regresi dengan dua ketidakpastian, tetapi tidak untuk tiga atau lebih ketidakpastian.

Jika pelaku eksperimen telah melakukan pengukuran pada tiga nilai berbeda dari variabel independen vektor X, maka analisis regresi akan memberikan satu set perkiraan unik untuk tiga parameter yang tidak diketahui di β. Dalam kasus regresi linier umum, pernyataan di atas setara dengan persyaratan bahwa matriks X^TX dapat dibalik.

Asumsi dalan Uji Statistik

Jika jumlah pengukuran, N lebih besar dari jumlah parameter yang tidak diketahui, k, dan kesalahan pengukuran ε terdistribusi normal maka kelebihan informasi yang terdapat pada pengukuran (N - k) digunakan untuk membuat prediksi statistik tentang parameter yang tidak diketahui. . Kelebihan informasi ini disebut sebagai derajat kebebasan regresi.

Asumsi yang Mendasari

Asumsi adalah sebuah perkiraan yang biasa dibuat oleh manusia untuk menyederhanakan suatu masalah. Biasanya ia digunakan ketika menganalisa suatu masalah dikarenakan adanya variabel-variabel tertentu yang tidak terukur / diketahui. Asumsi juga biasa digunakan untuk menyingkat waktu penyelesaian masalah dan biasanya amat erat terkait dengan pengalaman pribadi penggunanya

Asumsi klasik untuk analisis regresi meliputi:

- Sampel mewakili populasi untuk prediksi inferensi.
- Error adalah variabel acak dengan mean nol tergantung pada variabel penjelas.
- Variabel independen diukur tanpa kesalahan. (Catatan: Jika tidak demikian, pemodelan dapat dilakukan menggunakan teknik model error-in-variable).

- Variabel bebas (prediktor) adalah bebas linier, yaitu tidak mungkin untuk menyatakan prediktor apapun sebagai kombinasi linier dari yang lain.
- Kesalahan tidak berkorelasi, yaitu matriks varians-kovarian dari kesalahan adalah diagonal dan setiap elemen bukan nol adalah varian kesalahan.
- Varians kesalahan konstan di seluruh pengamatan (homoskedastisitas). Jika tidak, kotak terkecil berbobot atau metode lain dapat digunakan.

Ini adalah kondisi yang cukup bagi estimator kuadrat-terkecil untuk memiliki properti yang diinginkan; Secara khusus, asumsi ini menyiratkan bahwa estimasi parameter akan tidak bias, konsisten, dan efisien di kelas penduga yang tidak bias linier. Penting untuk dicatat bahwa data aktual jarang memenuhi asumsi. Artinya, metode tersebut digunakan meskipun asumsi-asumsi tersebut tidak benar. Variasi dari asumsi terkadang dapat digunakan sebagai ukuran sejauh mana model tersebut bermanfaat. Banyak dari asumsi ini dapat dilonggarkan dalam perawatan yang lebih maju. Laporan analisis statistik biasanya mencakup analisis pengujian pada data sampel dan metodologi untuk kesesuaian dan kegunaan model.

Asumsi termasuk dukungan geometris dari variabel. Variabel independen dan dependen sering mengacu pada nilai yang diukur di lokasi titik. Mungkin ada tren spasial dan autokorelasi spasial dalam variabel yang melanggar asumsi statistik regresi. Regresi berbobot geografis adalah salah satu teknik untuk menangani data tersebut. Selain itu, variabel dapat mencakup nilai yang dikumpulkan berdasarkan area. Dengan data agregat, masalah unit areal yang dapat dimodifikasi dapat menyebabkan variasi ekstrim dalam parameter regresi. Saat menganalisis data yang dikumpulkan berdasarkan batas-batas politik, hasil kode pos atau wilayah sensus mungkin sangat berbeda dengan pilihan unit yang berbeda.

Regresi Linier

Dalam regresi linier, spesifikasi modelnya adalah variabel dependen, y_i adalah kombinasi linier dari parameter (tetapi tidak harus linier dalam variabel independen). Misalnya, dalam regresi linier sederhana untuk pemodelan n titik data terdapat satu variabel independen: x_i, dan dua parameter, β_0 and β_1 :

garis lurus: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n.$

Dalam regresi linier berganda, terdapat beberapa variabel independen atau fungsi variabel independen.

Menambahkan suku dalam x_i^2 ke regresi sebelumnya menghasilkan:

parabola: $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i,\ i = 1, \ldots, n.$

Ini masih regresi linier; meskipun ekspresi di sisi kanan berbentuk kuadrat dalam variabel independen x_i, itu linier dalam parameter β_0 dan β_1, β_2

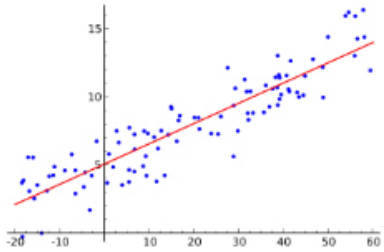
Dalam kedua kasus, ε_i adalah istilah kesalahan dan subskrip i mengindeks pengamatan tertentu.

Mengembalikan perhatian kami pada kasus garis lurus: Diberikan sampel acak dari populasi, kami memperkirakan parameter populasi dan memperoleh model regresi linier sampel: $\widehat{y_i} = \hat{\beta}_0 + \hat{\beta}_1 x_i$

Residual, $e_i = y_i - \hat{y}_i$, adalah selisih antara nilai variabel dependen yang diprediksi oleh model, $\widehat{y}_i$,, dan nilai sebenarnya dari variabel dependen, y_i. Salah satu metode estimasi adalah kuadrat terkecil biasa. Metode ini mendapatkan estimasi parameter yang meminimalkan jumlah kuadrat residual, SSE, terkadang juga disebut RSS:

$$SSE = \sum_{i=1}^{n} e_i^2$$

Minimisasi fungsi ini menghasilkan sekumpulan persamaan normal, sekumpulan persamaan linier simultan dalam parameter, yang diselesaikan untuk menghasilkan penduga parameter, $\widehat{\beta}_0$, $\widehat{\beta}_1$



Ilustrasi regresi *linier pada kumpulan data.*

Dalam kasus regresi sederhana, rumus untuk perkiraan kuadrat terkecil adalah

$$\widehat{\beta}_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2} \text{ and } \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$$

di mana $\overline{x}$ adalah mean (rata-rata) dari nilai x dan $\overline{y}$ adalah mean dari nilai y.

Dengan asumsi suku kesalahan populasi memiliki varians konstan, estimasi varians tersebut diberikan oleh:

$$\hat{\sigma}_\varepsilon^2 = \frac{SSE}{n-2}.$$

Ini disebut mean square error (MSE) dari regresi. Penyebut adalah ukuran sampel yang dikurangi dengan jumlah parameter model yang diperkirakan dari data yang sama, (n-p) untuk regressor p atau (n-p-1) jika intersep digunakan. Dalam hal ini, p = 1 jadi penyebutnya adalah n-2.

Kesalahan standar dari perkiraan parameter dinyatakan dengan :

$$\hat{\sigma}_{\beta_0} = \hat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{\overline{x}^2}{\sum (x_i - \overline{x})^2}}$$

$$\hat{\sigma}_{\beta_1} = \hat{\sigma}_\varepsilon \sqrt{\frac{1}{\sum (x_i - \overline{x})^2}}.$$

Dengan asumsi lebih lanjut bahwa istilah kesalahan populasi terdistribusi normal, peneliti dapat menggunakan kesalahan standar yang diperkirakan ini untuk membuat interval kepercayaan dan melakukan uji hipotesis tentang parameter populasi.

Model Linier Umum

Dalam model regresi berganda yang lebih umum, ada p variabel independen:

Rumus $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i,$

di mana x_{ij} adalah i^{th} pengamatan pada j^{th} variabel independen. Jika variabel independen pertama mengambil nilai 1 untuk semua x, x_{i1} = 1, maka β_1 disebut intersep regresi.

Estimasi parameter kuadrat terkecil diperoleh dari persamaan normal p. Sisa bisa ditulis sebagai : $\varepsilon_i = y_i - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_p x_{ip}.$

Persamaan normalnya adalah :

$$\sum_{i=1}^{n}\sum_{k=1}^{p} X_{ij}X_{ik}\hat{\beta}_k = \sum_{i=1}^{n} X_{ij}y_i,\ j = 1,\ldots,p.$$

Dalam notasi matriks, persamaan normal ditulis:

$$(X^\top X)\hat{\beta} = X^\top Y,$$

dimana elemen ij dari x adalah x_{ij}, elemen i dari vektor kolom Y adalah y_1, dan elemen j dari $\hat{\beta}$ adalah $\hat{\beta}_j$. Jadi X adalah n × p, Y adalah n × 1, dan $\hat{\beta}$ adalah n × 1. Solusinya adalah : $\hat{\beta} = (X^\top X)^{-1}X^\top Y.$

Diagnostik

Setelah model regresi dibuat, mungkin penting untuk memastikan kesesuaian model dan signifikansi statistik dari parameter yang diperkirakan. Pemeriksaan kesesuaian yang umum digunakan meliputi R-squared, analisis pola residual, dan pengujian hipotesis. Signifikansi statistik dapat diperiksa dengan uji-F dari kesesuaian keseluruhan, diikuti dengan uji-t untuk parameter individu.

Interpretasi dari tes diagnostik ini sangat bergantung pada asumsi model. Meskipun pemeriksaan residual dapat digunakan untuk membuat model tidak valid, hasil uji-t atau uji-F terkadang lebih sulit diinterpretasikan jika asumsi model dilanggar. Misalnya, jika istilah kesalahan tidak memiliki distribusi normal, dalam sampel kecil, parameter yang diperkirakan tidak akan mengikuti distribusi normal dan memperumit kesimpulan. Dengan sampel yang relatif besar, bagaimanapun, teorema batas pusat dapat digunakan sehingga pengujian hipotesis dapat dilanjutkan dengan menggunakan pendekatan asimtotik.

Model Limited Dependent Variables

Ungkapan *Limited Dependent Variables* digunakan dalam statistik ekonometrik untuk variabel kategori dan variabel terbatas. Variabel respon mungkin non-kontinu (“terbatas” terletak pada beberapa bagian dari garis nyata). Untuk variabel biner (nol atau satu), jika analisis dilanjutkan dengan regresi linier kuadrat-terkecil, model ini disebut model probabilitas linier. Model nonlinear untuk variabel dependen biner termasuk model probit dan logit. Model probit multivariat adalah metode standar untuk memperkirakan hubungan bersama antara beberapa variabel dependen biner dan beberapa variabel independen. Untuk variabel kategori dengan lebih dari dua nilai ada multinomial logit. Untuk variabel ordinal dengan lebih dari dua nilai, ada logit yang dipesan dan model probit yang dipesan. Model regresi yang disensor dapat digunakan ketika variabel dependen hanya kadang-

kadang diamati, dan model tipe koreksi Heckman dapat digunakan ketika sampel tidak dipilih secara acak dari populasi yang diminati. Alternatif untuk prosedur tersebut adalah regresi linier berdasarkan korelasi polikorik (atau korelasi poliserial) antara variabel kategori. Prosedur seperti itu berbeda dalam asumsi yang dibuat tentang distribusi variabel dalam populasi. Jika variabel positif dengan nilai-nilai rendah dan mewakili pengulangan terjadinya suatu peristiwa, maka hitung model seperti regresi Poisson atau model binomial negatif yang dapat digunakan sebagai gantinya.

Interpolasi dan Ekstrapolasi

Model regresi memprediksi nilai variabel Y mengingat nilai variabel X yang diketahui. Prediksi dalam kisaran nilai dalam dataset yang digunakan untuk model-fitting dikenal secara informal sebagai interpolasi. Prediksi di luar rentang data ini dikenal sebagai ekstrapolasi. Melakukan ekstrapolasi sangat bergantung pada asumsi regresi. Semakin jauh ekstrapolasi keluar dari data, semakin banyak ruang bagi model untuk gagal karena perbedaan antara asumsi dan data sampel atau nilai sebenarnya.

Secara umum disarankan bahwa ketika melakukan ekstrapolasi, seseorang harus menyertai nilai estimasi dari variabel dependen dengan interval prediksi yang mewakili ketidakpastian. Interval seperti itu cenderung berkembang pesat karena nilai variabel independen bergerak di luar kisaran yang dicakup oleh data yang diamati. Untuk alasan seperti itu dan yang lainnya, beberapa cenderung mengatakan bahwa mungkin tidak bijaksana untuk melakukan ekstrapolasi. Namun, ini tidak mencakup kesalahan pemodelan yang mungkin dibuat: khususnya, asumsi bentuk tertentu untuk hubungan antara Y dan X. Analisis regresi yang dilakukan dengan benar akan mencakup penilaian tentang seberapa baik bentuk yang diasumsikan dicocokkan dengan data yang diamati, tetapi itu hanya dapat dilakukan dalam kisaran nilai variabel independen yang sebenarnya tersedia. Ini berarti bahwa setiap ekstrapolasi sangat bergantung pada asumsi yang dibuat tentang bentuk struktural dari hubungan regresi. Saran praktik terbaik di sini adalah bahwa hubungan linier-dalam-variabel dan linier-dalam-parameter tidak boleh dipilih hanya untuk kenyamanan komputasi, tetapi semua pengetahuan yang tersedia harus diterapkan dalam membangun model regresi. Jika pengetahuan ini mencakup fakta bahwa variabel dependen tidak dapat keluar dari rentang nilai tertentu, ini dapat digunakan dalam memilih model - bahkan jika dataset yang diamati tidak memiliki nilai yang sangat dekat dengan batas tersebut. Implikasi dari langkah memilih bentuk fungsional yang sesuai untuk regresi dapat menjadi besar ketika ekstrapolasi dipertimbangkan. Minimal, dapat

memastikan bahwa setiap ekstrapolasi yang timbul dari model pas adalah "realistis" (atau sesuai dengan apa yang diketahui).

Regresi Nonlinear

Ketika fungsi model tidak linier dalam parameter, jumlah kuadrat harus diminimalkan dengan prosedur berulang. Ini memperkenalkan banyak komplikasi yang dirangkum dalam Perbedaan antara kuadrat terkecil linier dan non-linier

Perhitungan Daya dan Ukuran Sampel

Tidak ada metode yang disepakati secara umum untuk menghubungkan jumlah observasi versus jumlah variabel independen dalam model. Salah satu aturan praktis yang disarankan oleh Good dan Hardin adalah $N = m^n$, di mana N adalah ukuran sampel, n adalah jumlah variabel independen dan m adalah jumlah observasi yang diperlukan untuk mencapai presisi yang diinginkan jika model hanya memiliki satu variabel independen. Misalnya, seorang peneliti sedang membangun model regresi linier menggunakan dataset yang berisi 1000 pasien (N). Jika peneliti memutuskan bahwa lima pengamatan diperlukan untuk secara tepat menentukan garis lurus (m), maka jumlah maksimum variabel independen yang dapat didukung model adalah 4, karena

$$\frac{\log 1000}{\log 5} = 4.29.$$

Metode Lainnya

Meskipun parameter model regresi biasanya diperkirakan menggunakan metode kuadrat terkecil, metode lain yang telah digunakan meliputi:

- Metode Bayesian, mis. Regresi linier Bayesian
- Persentase regresi, untuk situasi di mana mengurangi kesalahan persentase dianggap lebih tepat.
- *Least absolute deviations*, yang lebih kuat dengan adanya outlier, yang mengarah ke regresi kuadrat
- Regresi nonparametrik, membutuhkan banyak pengamatan dan intensif secara komputasi
- *Distance metric learning,* yang dipelajari dengan mencari metrik jarak yang bermakna di ruang input yang diberikan.

Perangkat Lunak Model Regresi

Semua paket perangkat lunak statistik utama melakukan analisis regresi dan inferensi kuadrat terkecil. Regresi linier sederhana dan regresi berganda menggunakan kuadrat terkecil dapat dilakukan di beberapa aplikasi spreadsheet dan beberapa kalkulator. Meskipun banyak paket perangkat lunak statistik dapat melakukan berbagai jenis regresi nonparametrik dan regresi kuat, metode ini kurang terstandarisasi; paket perangkat lunak yang berbeda menerapkan metode yang berbeda, dan metode dengan nama tertentu dapat diterapkan secara berbeda dalam paket yang berbeda. Perangkat lunak regresi khusus telah dikembangkan untuk digunakan dalam bidang-bidang seperti analisis survei dan pencitraan saraf

3.7 Peringkas Teks Otomatis

Peringkasan otomatis adalah proses mengurangi dokumen teks dengan program komputer untuk membuat ringkasan yang mempertahankan poin paling penting dari dokumen asli. Teknologi yang dapat membuat ringkasan yang koheren memperhitungkan variabel akun seperti panjang, gaya penulisan, dan sintaksis. Peringkasan data otomatis adalah bagian dari pembelajaran mesin dan penggalian data. Gagasan utama peringkasan adalah menemukan subset representatif dari data, yang berisi informasi seluruh rangkaian. Teknologi peringkasan digunakan dalam sejumlah besar sektor di industri saat ini. Contoh penggunaan teknologi peringkasan adalah mesin pencari seperti Google. Contoh lain termasuk peringkasan dokumen, peringkasan koleksi gambar dan peringkasan video. Peringkasan dokumen, mencoba untuk secara otomatis membuat ringkasan representatif atau abstrak dari seluruh dokumen, dengan menemukan kalimat yang paling informatif. Demikian pula, dalam peringkasan gambar sistem menemukan gambar yang paling representatif dan penting (atau menonjol). Demikian pula, dalam video konsumen, seseorang ingin menghapus adegan yang membosankan atau berulang, dan mengekstrak versi video yang jauh lebih pendek dan ringkas. Ini juga penting, katakanlah untuk video pengawasan, di mana orang mungkin ingin mengekstrak hanya peristiwa penting dalam video yang direkam, karena sebagian besar video mungkin tidak menarik tanpa terjadi apa-apa. Ketika masalah kelebihan informasi tumbuh, dan ketika jumlah data meningkat, minat dalam peringkasan otomatis juga meningkat.

Secara umum, ada dua pendekatan untuk peringkasan otomatis: ekstraksi dan abstraksi. Metode ekstraktif bekerja dengan memilih subset dari kata, frasa, atau kalimat yang ada dalam teks asli untuk membentuk ringkasan. Sebaliknya, metode abstraktif membangun

representasi semantik internal dan kemudian menggunakan teknik generasi bahasa alami untuk membuat ringkasan yang lebih dekat dengan apa yang dihasilkan manusia. Ringkasan seperti itu mungkin mengandung kata-kata yang tidak secara eksplisit ada dalam aslinya. Penelitian mengenai metode abstraktif adalah bidang penelitian yang semakin penting dan aktif, namun karena kendala kompleksitas, penelitian hingga saat ini telah difokuskan terutama pada metode ekstraktif. Dalam beberapa domain aplikasi, summarization ekstraktif lebih masuk akal. Contohnya termasuk perangkuman koleksi gambar dan perangkuman video.

Summarization Berbasis Ekstraksi

Dalam tugas peringkasan ini, sistem otomatis mengekstrak objek dari seluruh koleksi, tanpa memodifikasi objek itu sendiri. Contohnya termasuk ekstraksi keyphrase, di mana tujuannya adalah untuk memilih kata atau frasa individual untuk “menandai” dokumen, dan meringkas dokumen, di mana tujuannya adalah untuk memilih seluruh kalimat (tanpa memodifikasi mereka) untuk membuat ringkasan paragraf pendek. Demikian pula, dalam ringkasan koleksi gambar, sistem mengekstraksi gambar dari koleksi tanpa memodifikasi gambar itu sendiri.

Peringkasan Berbasis Abstraksi

Teknik ekstraksi hanya menyalin informasi yang dianggap paling penting oleh sistem ke ringkasan (misalnya, klausa kunci, kalimat atau paragraf), sementara abstraksi melibatkan bagian parafrase pada dokumen sumber. Secara umum, abstraksi dapat menyingkat teks lebih kuat daripada ekstraksi, tetapi program yang dapat melakukan ini lebih sulit untuk dikembangkan karena mereka membutuhkan penggunaan teknologi generasi bahasa alami, yang dengan sendirinya merupakan bidang yang berkembang. Sementara beberapa pekerjaan telah dilakukan dalam peringkasan abstraktif (membuat sinopsis abstrak seperti manusia), sebagian besar sistem peringkasan bersifat ekstraktif (memilih subset kalimat untuk ditempatkan dalam ringkasan).

Aided Summarization

Teknik pembelajaran mesin dari bidang yang terkait erat seperti pengambilan informasi atau penambangan teks telah berhasil diadaptasi untuk membantu peringkasan otomatis.

Selain Peringkasan Otomatis Penuh (FAS), ada sistem yang membantu pengguna dengan tugas peringkasan (MAHS = Peringkasan Manusia Berbantuan Mesin), misalnya

dengan menyorot bagian kandidat untuk dimasukkan dalam ringkasan, dan ada sistem yang bergantung pada pos -proses oleh manusia (HAMS = Human Aided Machine Summarization).

Aplikasi dan Sistem untuk Summarization

Ada dua jenis tugas peringkasan ekstraktif secara luas, bergantung pada apa yang menjadi fokus program peringkasan. Yang pertama adalah ringkasan umum, yang berfokus pada perolehan ringkasan umum atau abstrak dari koleksi (baik dokumen, atau kumpulan gambar, atau video, berita, dll.). Yang kedua adalah peringkasan yang relevan kueri, terkadang disebut peringkasan berbasis kueri, yang meringkas objek khusus untuk kueri. Sistem peringkasan dapat membuat ringkasan teks kueri yang relevan dan ringkasan yang dibuat oleh mesin umum, bergantung pada apa yang dibutuhkan pengguna.

Contoh masalah peringkasan adalah peringkasan dokumen, yang mencoba menghasilkan abstrak dari dokumen tertentu secara otomatis. Terkadang seseorang mungkin tertarik untuk membuat ringkasan dari satu dokumen sumber, sementara yang lain dapat menggunakan beberapa dokumen sumber (misalnya, sekumpulan artikel tentang topik yang sama). Masalah ini disebut peringkasan multi-dokumen. Aplikasi terkait meringkas artikel berita. Bayangkan sebuah sistem, yang secara otomatis mengumpulkan artikel berita tentang topik tertentu (dari web), dan secara ringkas menyajikan berita terbaru sebagai ringkasan.

Peringkasan kumpulan gambar adalah contoh aplikasi lain dari peringkasan otomatis. Ini terdiri dari pemilihan kumpulan gambar yang representatif dari kumpulan gambar yang lebih besar. Ringkasan dalam konteks ini berguna untuk menampilkan gambar hasil yang paling representatif dalam sistem eksplorasi koleksi gambar. Peringkasan video adalah domain terkait, di mana sistem secara otomatis membuat cuplikan dari video panjang. Ini juga memiliki aplikasi dalam video konsumen atau pribadi, di mana orang mungkin ingin melewatkan tindakan membosankan atau berulang. Demikian pula, dalam video pengawasan, seseorang ingin mengekstrak aktivitas penting dan mencurigakan, sambil mengabaikan semua bingkai membosankan dan berlebihan yang ditangkap.

Pada tingkat yang sangat tinggi, algoritme peringkasan mencoba menemukan subset objek (seperti kumpulan kalimat, atau kumpulan gambar), yang mencakup informasi dari keseluruhan kumpulan. Ini juga disebut set inti. Algoritme ini memodelkan gagasan

seperti keragaman, cakupan, informasi, dan keterwakilan ringkasan. Teknik peringkasan berbasis kueri, selain itu memodelkan relevansi ringkasan dengan kueri. Beberapa teknik dan algoritma yang secara alami memodelkan masalah peringkasan adalah TextRank dan PageRank, Fungsi set submodular, Proses titik determinan, relevansi marginal maksimal (MMR) dll.

Ekstraksi Keyphrase

Tugasnya adalah sebagai berikut. Anda diberi sepotong teks, seperti artikel jurnal, dan Anda harus membuat daftar kata kunci atau [frase] kunci yang menangkap topik utama yang dibahas dalam teks. Dalam kasus artikel penelitian, banyak penulis memberikan kata kunci yang ditetapkan secara manual, tetapi sebagian besar teks tidak memiliki frasa kunci yang sudah ada sebelumnya. Misalnya, artikel berita jarang memiliki frasa kunci yang dilampirkan, tetapi akan berguna jika dapat melakukannya secara otomatis untuk sejumlah aplikasi yang dibahas di bawah ini. Perhatikan contoh teks dari artikel berita:

> *"Korps Insinyur Angkatan Darat, bergegas untuk memenuhi janji Presiden Bush untuk melindungi New Orleans pada awal musim badai 2006, memasang pompa pengendali banjir yang rusak tahun lalu meskipun ada peringatan dari ahlinya bahwa peralatan akan gagal selama badai, menurut ke dokumen yang diperoleh The Associated Press ".*

Ekstraktor frasa kunci mungkin memilih "Korps Insinyur Angkatan Darat", "Presiden Bush", "New Orleans", dan "pompa pengendali banjir yang rusak" sebagai frasa kunci. Ini ditarik langsung dari teks. Sebaliknya, sistem frasa kunci abstraktif entah bagaimana akan menginternalisasi konten dan menghasilkan frasa unik yang tidak muncul dalam teks, tetapi lebih mirip dengan apa yang mungkin dihasilkan manusia, seperti "kelalaian politik" atau "perlindungan yang tidak memadai dari banjir". Abstraksi membutuhkan pemahaman yang mendalam tentang teks, yang menyulitkan sistem komputer. Frase unik memiliki banyak aplikasi. Mereka dapat mengaktifkan penjelajahan dokumen dengan memberikan ringkasan singkat, meningkatkan pengambilan informasi (jika dokumen memiliki frasa kunci yang ditetapkan, pengguna dapat mencari dengan frasa kunci untuk menghasilkan klik yang lebih andal daripada pencarian teks lengkap), dan digunakan untuk menghasilkan entri indeks yang besar. teks korpus.

Tergantung pada literatur yang berbeda dan definisi istilah kunci, kata atau frase, tema yang sangat terkait tentunya ekstraksi Kata Kunci

Pendekatan Supervised Learning

Dimulai dengan pekerjaan Turney, banyak peneliti telah mendekati ekstraksi frasa kunci sebagai masalah pembelajaran mesin yang diawasi. Diberikan sebuah dokumen, kami membuat contoh untuk setiap unigram, bigram, dan trigram yang ditemukan dalam teks (meskipun unit teks lain juga dimungkinkan, seperti dibahas di bawah). Kami kemudian menghitung berbagai fitur yang menjelaskan setiap contoh (misalnya, apakah frasa dimulai dengan huruf besar?). Kami berasumsi ada frasa kunci yang diketahui tersedia untuk sekumpulan dokumen pelatihan. Menggunakan Frase unik yang diketahui, kita dapat memberikan label positif atau negatif ke contoh. Kemudian kita mempelajari pengklasifikasi yang dapat membedakan antara contoh positif dan negatif sebagai fungsi fitur. Beberapa pengklasifikasi membuat klasifikasi biner untuk contoh pengujian, sementara pengklasifikasi lainnya menetapkan kemungkinan menjadi frasa kunci. Misalnya, pada teks di atas, kita mungkin mempelajari aturan yang mengatakan frasa dengan huruf kapital kemungkinan besar menjadi Frase unik. Setelah melatih pelajar, kita dapat memilih Frase unik untuk dokumen pengujian dengan cara berikut. Kami menerapkan strategi pembuatan contoh yang sama ke dokumen pengujian, lalu menjalankan setiap contoh melalui pelajar. Kita dapat menentukan Frase unik dengan melihat keputusan klasifikasi biner atau probabilitas yang dikembalikan dari model yang kita pelajari. Jika probabilitas diberikan, ambang batas digunakan untuk memilih frasa kunci. Ekstraktor frasa kunci umumnya dievaluasi menggunakan presisi dan penarikan kembali. Ketepatan mengukur berapa banyak frasa kunci yang diusulkan sebenarnya benar. Ingat mengukur berapa banyak frase kunci yang benar yang diusulkan sistem Anda. Kedua ukuran tersebut dapat digabungkan dalam skor-F, yang merupakan rata-rata harmonis dari keduanya (F = 2PR / (P + R)). Kecocokan antara frasa kunci yang diusulkan dan frasa unik yang diketahui dapat diperiksa setelah stemming atau menerapkan beberapa normalisasi teks lainnya.

Merancang sistem ekstraksi frasa kunci yang diawasi melibatkan penentuan beberapa pilihan (beberapa di antaranya juga berlaku untuk tanpa pengawasan). Pilihan pertama persis bagaimana menghasilkan contoh. Turney dan yang lainnya telah menggunakan semua kemungkinan unigram, bigram, dan trigram tanpa mengintervensi tanda baca dan setelah menghapus stopwords. Hulth menunjukkan bahwa Anda bisa mendapatkan beberapa peningkatan dengan memilih contoh untuk menjadi urutan token yang cocok dengan pola tertentu dari tag bagian ucapan. Idealnya, mekanisme untuk menghasilkan contoh menghasilkan semua frasa kunci berlabel yang dikenal sebagai kandidat, meskipun ini sering kali tidak terjadi. Misalnya, jika kita hanya menggunakan unigram, bigram, dan trigram, maka kita tidak akan pernah bisa mengekstrak frasa kunci yang diketahui

yang mengandung empat kata. Dengan demikian, ingatan mungkin menderita. Namun, menghasilkan terlalu banyak contoh juga dapat menyebabkan presisi yang rendah.

Kita juga perlu membuat fitur yang mendeskripsikan contoh dan cukup informatif untuk memungkinkan algoritma pembelajaran membedakan Frase Utama dari Frase Non-Frase. Biasanya fitur melibatkan berbagai frekuensi istilah (berapa kali frase muncul dalam teks saat ini atau dalam korpus yang lebih besar), panjang contoh, posisi relatif kejadian pertama, berbagai fitur sintaksis boolean (misalnya, berisi semua huruf besar), dll. Kertas Turney menggunakan sekitar 12 fitur tersebut. Hulth menggunakan serangkaian fitur yang dikurangi, yang ditemukan paling berhasil dalam karya KEA (Keyphrase Extraction Algorithm) yang berasal dari makalah mani Turney.

Pada akhirnya, sistem perlu mengembalikan daftar frasa kunci untuk dokumen pengujian, jadi kita perlu memiliki cara untuk membatasi jumlahnya. Metode ensemble (yaitu, menggunakan suara dari beberapa pengklasifikasi) telah digunakan untuk menghasilkan skor numerik yang dapat diberi ambang batas untuk memberikan jumlah frasa kunci yang diberikan pengguna. Ini adalah teknik yang digunakan oleh Turney dengan pohon keputusan C4.5. Hulth menggunakan pengklasifikasi biner tunggal sehingga algoritme pembelajaran secara implisit menentukan angka yang sesuai.

Setelah contoh dan fitur dibuat, kita memerlukan cara untuk belajar memprediksi frasa kunci. Hampir semua algoritma pembelajaran yang diawasi dapat digunakan, seperti pohon keputusan, Naive Bayes, dan instruksi aturan. Dalam kasus algoritme GenEx Turney, algoritme genetika digunakan untuk mempelajari parameter untuk algoritme ekstraksi frasa kunci khusus domain. Ekstraktor mengikuti serangkaian heuristik untuk mengidentifikasi frasa kunci. Algoritme genetik mengoptimalkan parameter untuk heuristik ini sehubungan dengan performa pada dokumen pelatihan dengan frasa kunci yang diketahui.

Pendekatan Unsupervised Learning: Textrank

Algoritma ekstraksi frasa kunci lainnya adalah TextRank. Meskipun metode yang diawasi memiliki beberapa properti bagus, seperti mampu menghasilkan aturan yang dapat ditafsirkan untuk fitur apa yang menjadi ciri frase kunci, metode tersebut juga memerlukan data pelatihan dalam jumlah besar. Diperlukan banyak dokumen dengan frasa kunci yang diketahui. Selain itu, pelatihan pada domain tertentu cenderung menyesuaikan proses ekstraksi ke domain tersebut, sehingga pengklasifikasi yang dihasilkan belum tentu

portabel, seperti yang ditunjukkan oleh beberapa hasil Turney. Ekstraksi frasa kunci tanpa pengawasan menghilangkan kebutuhan akan data pelatihan. Ini mendekati masalah dari sudut yang berbeda. Alih-alih mencoba mempelajari fitur eksplisit yang menjadi ciri frasa unik, algoritme TextRank mengeksploitasi struktur teks itu sendiri untuk menentukan Frasa unik yang tampak "sentral" pada teks dengan cara yang sama seperti PageRank memilih halaman Web penting. Ingat ini didasarkan pada gagasan "prestise" atau "rekomendasi" dari jejaring sosial. Dengan cara ini, TextRank tidak bergantung pada data pelatihan sebelumnya sama sekali, melainkan dapat dijalankan pada bagian teks manapun, dan dapat menghasilkan keluaran hanya berdasarkan properti intrinsik teks. Dengan demikian algoritme ini mudah dibawa-bawa ke domain dan bahasa baru.

TextRank adalah algoritma peringkat berbasis grafik untuk tujuan umum untuk NLP. Pada dasarnya, ini menjalankan PageRank pada grafik yang dirancang khusus untuk tugas NLP tertentu. Untuk ekstraksi frasa kunci, itu membangun grafik menggunakan beberapa set unit teks sebagai simpul. Tepi didasarkan pada beberapa ukuran kemiripan semantik atau leksikal antara simpul unit teks. Tidak seperti PageRank, tepi biasanya tidak diarahkan dan dapat diberi bobot untuk mencerminkan tingkat kesamaan. Setelah grafik dibuat, grafik tersebut digunakan untuk membentuk matriks stokastik, dikombinasikan dengan faktor redaman (seperti dalam “model surfer acak”), dan peringkat di atas simpul diperoleh dengan mencari vektor eigen yang sesuai dengan nilai eigen 1 (yaitu, distribusi stasioner dari jalan acak pada grafik).

Simpul harus sesuai dengan apa yang ingin kita rangking. Secara potensial, kita dapat melakukan sesuatu yang mirip dengan metode yang diawasi dan membuat simpul untuk setiap unigram, bigram, trigram, dll. Namun, untuk menjaga grafik tetap kecil, penulis memutuskan untuk memberi peringkat unigram individu pada langkah pertama, dan kemudian memasukkan yang kedua langkah yang menggabungkan unigram berdekatan berperingkat tinggi untuk membentuk frase multi-kata. Ini memiliki efek samping yang bagus karena memungkinkan kita menghasilkan Frasa unik dengan panjang yang berubah-ubah. Misalnya, jika kita memberi peringkat unigram dan menemukan bahwa "lanjutan", "alami", "bahasa", dan "pemrosesan" semuanya mendapat peringkat tinggi, maka kita akan melihat teks asli dan melihat bahwa kata-kata ini muncul secara berurutan dan membuat final Frasa unik menggunakan keempatnya secara bersamaan. Perhatikan bahwa unigram yang ditempatkan pada grafik dapat difilter berdasarkan part of speech. Penulis menemukan bahwa kata sifat dan kata benda adalah yang terbaik untuk disertakan. Dengan demikian, beberapa pengetahuan linguistik ikut bermain dalam langkah ini.

Tepi dibuat berdasarkan kemunculan kata dalam aplikasi TextRank ini. Dua simpul dihubungkan oleh sebuah tepi jika unigram muncul dalam jendela berukuran N dalam teks asli. N biasanya sekitar 2-10. Jadi, "natural" dan "language" mungkin terkait dalam teks tentang NLP. "Alami" dan "pemrosesan" juga akan dihubungkan karena keduanya akan muncul dalam string N kata yang sama. Tepi ini dibangun di atas gagasan "kohesi teks" dan gagasan bahwa kata-kata yang muncul di dekat satu sama lain kemungkinan besar terkait dengan cara yang berarti dan "merekomendasikan" satu sama lain kepada pembaca.

Karena metode ini hanya memberi peringkat pada simpul individu, kita memerlukan cara untuk melakukan ambang batas atau menghasilkan frasa kunci dalam jumlah terbatas. Teknik yang dipilih adalah menyetel hitungan T menjadi pecahan yang ditentukan pengguna dari jumlah total simpul pada grafik. Kemudian T simpul / unigram teratas dipilih berdasarkan probabilitas stasionernya. Langkah pasca-pemrosesan kemudian diterapkan untuk menggabungkan contoh yang berdekatan dari T unigram ini. Akibatnya, frasa kunci akhir yang berpotensi lebih atau kurang dari T final akan dihasilkan, tetapi jumlahnya harus kira-kira sebanding dengan panjang teks asli.

Pada awalnya tidak jelas mengapa menerapkan PageRank ke grafik kejadian bersama akan menghasilkan frase kunci yang berguna. Salah satu cara untuk memikirkannya adalah sebagai berikut. Sebuah kata yang muncul beberapa kali di seluruh teks mungkin memiliki banyak tetangga yang muncul bersamaan. Misalnya, dalam teks tentang pembelajaran mesin, "pembelajaran" unigram mungkin muncul bersamaan dengan "mesin", "diawasi", "tidak diawasi", dan "semi-supervisi" dalam empat kalimat berbeda. Jadi, simpul "belajar" akan menjadi "pusat" pusat yang menghubungkan ke kata-kata pengubah lainnya ini. Menjalankan PageRank / TextRank pada grafik kemungkinan akan mendapat peringkat "belajar" tinggi. Demikian pula, jika teks berisi frasa "klasifikasi terbimbing", maka akan ada tepi antara "terbimbing" dan "klasifikasi". Jika "klasifikasi" muncul di beberapa tempat lain dan dengan demikian memiliki banyak tetangga, kepentingannya akan berkontribusi pada pentingnya "diawasi". Jika itu berakhir dengan peringkat tinggi, itu akan dipilih sebagai salah satu T unigram teratas, bersama dengan "pembelajaran" dan mungkin "klasifikasi". Pada langkah pasca-pemrosesan terakhir, kita akan berakhir dengan frasa kunci "pembelajaran yang diawasi" dan "klasifikasi yang diawasi".

Singkatnya, grafik kejadian bersama akan berisi wilayah yang terhubung dengan rapat untuk istilah yang sering muncul dan dalam konteks yang berbeda. Jalan acak pada

grafik ini akan memiliki distribusi stasioner yang memberikan probabilitas besar pada suku-suku di pusat cluster. Ini mirip dengan halaman Web yang terhubung dengan rapat yang mendapatkan peringkat tinggi oleh PageRank. Pendekatan ini juga telah digunakan dalam peringkasan dokumen, yang dibahas di bawah ini.

Peringkasan Dokumen

Seperti ekstraksi frasa kunci, peringkasan dokumen bertujuan untuk mengidentifikasi esensi teks. Satu-satunya perbedaan nyata adalah bahwa sekarang kita berurusan dengan unit teks yang lebih besar — seluruh kalimat, bukan kata dan frasa.

Sebelum membahas detail beberapa metode peringkasan, kami akan menyebutkan bagaimana sistem peringkasan biasanya dievaluasi. Cara yang paling umum adalah menggunakan apa yang disebut ukuran ROUGE (Recall-Oriented Understudy for Gisting Evaluation). Ini adalah ukuran berbasis ingatan yang menentukan seberapa baik ringkasan yang dihasilkan sistem mencakup konten yang ada dalam satu atau beberapa ringkasan model buatan manusia yang dikenal sebagai referensi. Itu berbasis ingatan untuk mendorong sistem untuk memasukkan semua topik penting dalam teks. Penarikan kembali dapat dihitung sehubungan dengan unigram, bigram, trigram, atau pencocokan 4 gram. Misalnya, ROUGE-1 dihitung sebagai pembagian jumlah unigram dalam referensi yang muncul di sistem dan jumlah unigram dalam ringkasan referensi.

Jika ada banyak referensi, skor ROUGE-1 akan dirata-ratakan. Karena ROUGE hanya didasarkan pada konten yang tumpang tindih, ROUGE dapat menentukan apakah konsep umum yang sama dibahas antara ringkasan otomatis dan ringkasan referensi, tetapi ROUGE tidak dapat menentukan apakah hasilnya koheren atau kalimat mengalir bersama dengan cara yang masuk akal. Pengukuran ROUGE n-gram orde tinggi mencoba menilai kelancaran sampai tingkat tertentu. Perhatikan bahwa ROUGE mirip dengan ukuran BLEU untuk terjemahan mesin, tetapi BLEU berbasis presisi, karena sistem terjemahan mendukung akurasi.

Baris yang menjanjikan dalam peringkasan dokumen adalah peringkasan dokumen / teks adaptif. Ide dari peringkasan adaptif melibatkan pengenalan awal dari genre dokumen / teks dan aplikasi selanjutnya dari algoritma peringkasan yang dioptimalkan untuk genre ini. Ringkasan pertama yang melakukan peringkasan adaptif telah dibuat.

Pendekatan Supervised learning

Supervised Learning sangat mirip dengan ekstraksi frasa kunci yang diawasi. Pada dasarnya, jika Anda memiliki kumpulan dokumen dan ringkasan yang dibuat oleh manusia, Anda dapat mempelajari fitur kalimat yang menjadikannya kandidat yang baik untuk disertakan dalam ringkasan. Fitur mungkin termasuk posisi dalam dokumen (yaitu, beberapa kalimat pertama mungkin penting), jumlah kata dalam kalimat, dll. Kesulitan utama dalam peringkasan ekstraktif terbimbing adalah bahwa ringkasan yang diketahui harus dibuat secara manual dengan mengekstraksi kalimat sehingga kalimat dalam dokumen pelatihan asli dapat diberi label sebagai "dalam ringkasan" atau "tidak dalam ringkasan". Ini bukan cara orang membuat ringkasan, jadi hanya menggunakan abstrak jurnal atau ringkasan yang ada biasanya tidak cukup. Kalimat dalam ringkasan ini tidak selalu cocok dengan kalimat dalam teks aslinya, sehingga akan sulit untuk memberikan label pada contoh untuk pelatihan. Namun, perhatikan bahwa ringkasan alami ini masih dapat digunakan untuk tujuan evaluasi, karena ROUGE-1 hanya peduli pada unigram.

Peringkasan Berbasis Entropi Maksimum

Selama lokakarya evaluasi DUC 2001 dan 2002, TNO mengembangkan sistem ekstraksi kalimat untuk peringkasan multi-dokumen dalam domain berita. Sistem ini didasarkan pada sistem hybrid menggunakan pengklasifikasi bayes naif dan model bahasa statistik untuk salience pemodelan. Meskipun sistem menunjukkan hasil yang baik, para peneliti ingin mengeksplorasi keefektifan pengklasifikasi entropi maksimum (ME) untuk tugas ringkasan rapat, karena ME dikenal kuat terhadap dependensi fitur. Entropi maksimum juga telah berhasil diterapkan untuk peringkasan dalam domain berita siaran.

TextRank dan LexRank

Pendekatan tanpa pengawasan untuk peringkasan juga sangat mirip dengan ekstraksi frase kunci tanpa pengawasan dan mengatasi masalah data pelatihan yang mahal. Beberapa pendekatan peringkasan tanpa pengawasan didasarkan pada pencarian kalimat "sentroid", yang merupakan vektor kata rata-rata dari semua kalimat dalam dokumen. Kemudian kalimat tersebut dapat diurutkan berdasarkan kemiripannya dengan kalimat sentroid ini.

Cara yang lebih berprinsip untuk memperkirakan kepentingan kalimat adalah menggunakan jalan acak dan sentralitas vektor eigen. LexRank pada dasarnya adalah algoritme yang identik dengan TextRank, dan keduanya menggunakan pendekatan ini untuk meringkas dokumen. Kedua metode tersebut dikembangkan oleh kelompok yang berbeda pada

waktu yang sama, dan LexRank hanya berfokus pada peringkasan, tetapi dapat dengan mudah digunakan untuk ekstraksi frase kunci atau tugas pemeringkatan NLP lainnya.

Di LexRank dan TextRank, grafik dibuat dengan membuat titik sudut untuk setiap kalimat di dokumen.

Tepi antar kalimat didasarkan pada beberapa bentuk kesamaan semantik atau konten yang tumpang tindih. Sementara LexRank menggunakan kesamaan kosinus dari vektor TF-IDF, TextRank menggunakan ukuran yang sangat mirip berdasarkan jumlah kata yang dimiliki dua kalimat yang sama (dinormalisasi dengan panjang kalimat). Kertas LexRank dieksplorasi menggunakan tepi tidak berbobot setelah menerapkan ambang batas ke nilai cosinus, tetapi juga bereksperimen dengan menggunakan tepi dengan bobot yang sama dengan skor kesamaan. TextRank menggunakan skor kesamaan berkelanjutan sebagai bobot.

Dalam kedua algoritme, kalimat diberi peringkat dengan menerapkan PageRank ke grafik yang dihasilkan. Ringkasan dibentuk dengan menggabungkan kalimat peringkat teratas, menggunakan ambang batas atau batas panjang untuk membatasi ukuran ringkasan.

Perlu dicatat bahwa TextRank diterapkan pada peringkasan persis seperti yang dijelaskan di sini, sedangkan LexRank digunakan sebagai bagian dari sistem peringkasan yang lebih besar (MEAD) yang menggabungkan skor LexRank (probabilitas stasioner) dengan fitur lain seperti posisi dan panjang kalimat menggunakan kombinasi linier dengan bobot yang ditentukan pengguna atau disetel secara otomatis. Dalam kasus ini, beberapa dokumen pelatihan mungkin diperlukan, meskipun hasil TextRank menunjukkan bahwa fitur tambahan tidak mutlak diperlukan.

Perbedaan penting lainnya adalah TextRank digunakan untuk peringkasan dokumen tunggal, sedangkan LexRank telah diterapkan pada peringkasan multi-dokumen. Tugasnya tetap sama di kedua kasus — hanya jumlah kalimat yang dipilih yang bertambah. Namun, saat meringkas banyak dokumen, ada risiko lebih besar untuk memilih kalimat duplikat atau sangat berlebihan untuk ditempatkan dalam ringkasan yang sama. Bayangkan Anda memiliki sekumpulan artikel berita tentang peristiwa tertentu, dan Anda ingin membuat satu ringkasan. Setiap artikel kemungkinan besar memiliki banyak kalimat yang mirip, dan Anda hanya ingin memasukkan ide yang berbeda dalam ringkasan. Untuk mengatasi masalah ini, LexRank menerapkan langkah pasca-pemrosesan heuristik yang membangun ringkasan dengan menambahkan kalimat dalam urutan peringkat, tetapi membuang

kalimat yang terlalu mirip dengan yang sudah ditempatkan di ringkasan. Metode yang digunakan disebut Cross-Sentence Information Subsumption (CSIS).

Metode ini bekerja berdasarkan gagasan bahwa kalimat “merekomendasikan” kalimat lain yang serupa kepada pembaca. Jadi, jika satu kalimat sangat mirip dengan banyak kalimat lainnya, kemungkinan besar itu akan menjadi kalimat yang sangat penting. Pentingnya kalimat ini juga bermula dari pentingnya kalimat “merekomendasikan” nya. Jadi, untuk mendapatkan peringkat tinggi dan ditempatkan dalam ringkasan, sebuah kalimat harus mirip dengan banyak kalimat yang pada gilirannya juga mirip dengan banyak kalimat lainnya. Ini masuk akal secara intuitif dan memungkinkan algoritme diterapkan ke teks baru sembarang. Metodenya tidak bergantung pada domain dan mudah dibawa-bawa. Dapat dibayangkan fitur yang menunjukkan kalimat penting dalam domain berita mungkin sangat berbeda dari domain biomedis. Namun, pendekatan berbasis "rekomendasi" tanpa pengawasan berlaku untuk domain apa pun.

Peringkasan Multi-Dokumen

Peringkasan multi-dokumen adalah prosedur otomatis yang ditujukan untuk mengekstraksi informasi dari beberapa teks yang ditulis tentang topik yang sama. Laporan ringkasan yang dihasilkan memungkinkan pengguna individu, seperti konsumen informasi profesional, untuk dengan cepat membiasakan diri dengan informasi yang terdapat dalam kumpulan dokumen yang besar. Sedemikian rupa, sistem peringkasan multi-dokumen melengkapi agregator berita yang melakukan langkah selanjutnya untuk mengatasi kelebihan informasi. Peringkasan multi-dokumen juga dapat dilakukan untuk menjawab pertanyaan.

Peringkasan multi-dokumen membuat laporan informasi yang ringkas dan komprehensif. Dengan berbagai pendapat yang dikumpulkan dan digarisbawahi, setiap topik dijelaskan dari berbagai perspektif dalam satu dokumen. Meskipun tujuan dari ringkasan singkat adalah untuk menyederhanakan pencarian informasi dan memotong waktu dengan menunjuk ke dokumen sumber yang paling relevan, ringkasan multi-dokumen yang komprehensif itu sendiri harus berisi informasi yang diperlukan, sehingga membatasi kebutuhan untuk mengakses file asli hanya pada kasus-kasus ketika perbaikan dilakukan. yg dibutuhkan. Ringkasan otomatis menyajikan informasi yang diambil dari berbagai sumber secara algoritme, tanpa sentuhan editorial atau campur tangan manusia yang subjektif, sehingga membuatnya benar-benar tidak bias.

Menggabungkan Keragaman

Ringkasan ekstraktif multi-dokumen menghadapi masalah potensi redundansi. Idealnya, kami ingin mengekstrak kalimat yang "sentral" (yaitu, berisi ide utama) dan "beragam" (yaitu, keduanya berbeda satu sama lain). LexRank menangani keragaman sebagai tahap akhir heuristik menggunakan CSIS, dan sistem lain telah menggunakan metode serupa, seperti Maximal Marginal Relevance (MMR), dalam mencoba menghilangkan redundansi dalam hasil pencarian informasi. Ada algoritme peringkat berbasis grafik untuk tujuan umum seperti Page / Lex / TextRank yang menangani "sentralitas" dan "keragaman" dalam kerangka kerja matematika terpadu berdasarkan serapan acak rantai Markov. (Jalan acak yang menyerap adalah seperti jalan acak standar, kecuali beberapa negara bagian sekarang menyerap keadaan yang bertindak sebagai "lubang hitam" yang menyebabkan jalan tersebut berakhir tiba-tiba pada keadaan tersebut.) Algoritme ini disebut GRASS-HOPPER. Selain secara eksplisit mempromosikan keragaman selama proses pemeringkatan, GRASSHOPPER memasukkan peringkat sebelumnya (berdasarkan posisi kalimat dalam kasus peringkasan).

Hasil mutakhir untuk ringkasan multi-dokumen, bagaimanapun, diperoleh dengan menggunakan campuran fungsi submodular. Metode ini telah mencapai hasil mutakhir untuk Document Summarization Corpora, DUC 04 - 07. Hasil serupa juga dicapai dengan penggunaan proses titik determinan (yang merupakan kasus khusus fungsi submodular) untuk DUC-04

Fungsi Submodular sebagai Alat Generik untuk Peringkasan

Ide tentang fungsi himpunan Submodular baru-baru ini muncul sebagai alat pemodelan yang kuat untuk berbagai masalah peringkasan. Fungsi submodular secara alami memodelkan pengertian cakupan, informasi, representasi dan keragaman. Selain itu, beberapa masalah optimasi kombinatorial penting terjadi sebagai contoh khusus dari optimasi submodular. Misalnya, masalah set cover adalah kasus khusus dari optimasi submodular, karena fungsi set cover adalah submodular. Fungsi set cover mencoba menemukan subset objek yang mencakup sekumpulan konsep tertentu. Misalnya, dalam peringkasan dokumen, seseorang ingin ringkasan mencakup semua konsep penting dan relevan dalam dokumen. Ini adalah contoh set sampul. Demikian pula, masalah lokasi fasilitas adalah kasus khusus dari fungsi submodular. Fungsi Lokasi Fasilitas juga secara alami memodelkan cakupan dan keragaman. Contoh lain dari masalah pengoptimalan submodular menggunakan proses titik determinan untuk model keragaman. Demikian pula, prosedur Relevansi Marginal Maksimum juga dapat dilihat sebagai contoh pengoptimalan

submodular. Semua model penting ini mendorong cakupan, keragaman, dan informasi semuanya submodular. Selain itu, fungsi submodular dapat digabungkan bersama secara efisien, dan fungsi yang dihasilkan masih submodular. Oleh karena itu, seseorang dapat menggabungkan satu fungsi submodular yang memodelkan keragaman, fungsi lainnya yang memodelkan cakupan dan menggunakan pengawasan manusia untuk mempelajari model yang tepat dari fungsi submodular untuk masalah tersebut.

Sementara fungsi submodular adalah masalah pas untuk peringkasan, mereka juga mengakui algoritma yang sangat efisien untuk pengoptimalan. Misalnya, algoritme serakah sederhana mengakui jaminan faktor konstan. Selain itu, algoritme greedy sangat sederhana untuk diterapkan dan dapat diskalakan ke kumpulan data besar, yang sangat penting untuk masalah peringkasan.

Fungsi submodular telah mencapai state-of-the-art untuk hampir semua masalah peringkasan. Misalnya, karya Lin dan Bilmes, 2012 menunjukkan bahwa fungsi submodular mencapai hasil terbaik hingga saat ini pada sistem DUC-04, DUC-05, DUC-06 dan DUC-07 untuk peringkasan dokumen. Demikian pula, karya Lin dan Bilmes, 2011, menunjukkan bahwa banyak sistem yang ada untuk peringkasan otomatis adalah contoh fungsi submodular. Ini adalah hasil terobosan yang menetapkan fungsi submodular sebagai model yang tepat untuk masalah peringkasan.

Fungsi Submodular juga telah digunakan untuk tugas peringkasan lainnya. Tschiatschek et al., 2014 menunjukkan bahwa campuran fungsi submodular mencapai hasil mutakhir untuk peringkasan koleksi gambar. Demikian pula, Bairi et al., 2015 menunjukkan kegunaan fungsi submodular untuk meringkas hierarki topik multi-dokumen. Fungsi Submodular juga berhasil digunakan untuk meringkas kumpulan data machine learning.

Teknik Evaluasi

Cara paling umum untuk mengevaluasi keinformatifan ringkasan otomatis adalah membandingkannya dengan ringkasan model buatan manusia. Teknik evaluasi termasuk dalam intrinsik dan ekstrinsik, intertekstual dan intra-tekstual.

Evaluasi Intrinsik dan Ekstrinsik

Evaluasi intrinsik menguji sistem peringkasan dengan sendirinya sementara evaluasi ekstrinsik menguji peringkasan berdasarkan bagaimana hal itu memengaruhi penyelesaian beberapa tugas lain. Evaluasi intrinsik terutama menilai koherensi dan keinformatifan

ringkasan. Evaluasi ekstrinsik, di sisi lain, telah menguji dampak peringkasan pada tugas-tugas seperti penilaian relevansi, pemahaman bacaan, dll.

Intertekstual dan Intra-tekstual

Metode intra-tekstual menilai keluaran dari sistem peringkasan tertentu, dan metode antar-tekstual berfokus pada analisis kontrastif terhadap keluaran dari beberapa sistem peringkasan.

Penilaian manusia sering kali memiliki perbedaan yang luas tentang apa yang dianggap sebagai ringkasan yang "baik", yang berarti bahwa membuat proses evaluasi otomatis sangat sulit. Evaluasi manual dapat digunakan, tetapi ini memakan waktu dan tenaga karena mengharuskan manusia untuk membaca tidak hanya ringkasan tetapi juga dokumen sumber. Masalah lainnya adalah tentang koherensi dan cakupan.

Salah satu metrik yang digunakan dalam Konferensi Pemahaman Dokumen tahunan NIST, di mana kelompok penelitian mengirimkan sistem mereka untuk tugas ringkasan dan terjemahan, adalah metrik ROUGE (Pemahaman Berorientasi Pemahaman untuk Evaluasi Gisting). Ini pada dasarnya menghitung over-lap n-gram antara ringkasan yang dibuat secara otomatis dan ringkasan manusia yang ditulis sebelumnya. Tingkat tumpang tindih yang tinggi harus menunjukkan tingkat konsep bersama yang tinggi di antara kedua ringkasan. Perhatikan bahwa metrik yang tumpang tindih seperti ini tidak dapat memberikan masukan apa pun tentang koherensi ringkasan. Resolusi anafor tetap menjadi masalah lain yang belum sepenuhnya terpecahkan. Demikian pula, untuk penjumlahan gambar, Tschiatschek dkk., Mengembangkan skor Visual-ROUGE yang menilai kinerja algoritma untuk peringkasan gambar.

Tantangan Saat Ini dalam Mengevaluasi Ringkasan Secara Otomatis

Mengevaluasi ringkasan, baik secara manual atau otomatis, adalah tugas yang sulit. Kesulitan utama dalam evaluasi berasal dari ketidakmungkinan membangun standar emas yang adil yang dengannya hasil sistem dapat dibandingkan. Selain itu, juga sangat sulit untuk menentukan ringkasan yang benar, karena selalu ada kemungkinan sistem menghasilkan ringkasan yang baik yang sangat berbeda dari ringkasan manusia yang digunakan sebagai perkiraan untuk keluaran yang benar.

Pemilihan konten bukanlah masalah deterministik. Orang-orang subjektif, dan penulis yang berbeda akan memilih kalimat yang berbeda. Dan individu mungkin tidak konsisten.

Orang tertentu mungkin memilih kalimat yang berbeda pada waktu yang berbeda. Dua kalimat berbeda yang diungkapkan dengan kata berbeda dapat mengungkapkan makna yang sama. Fenomena ini dikenal sebagai parafrase. Kita dapat menemukan pendekatan untuk mengevaluasi ringkasan secara otomatis menggunakan parafrase (ParaEval).

Kebanyakan sistem peringkasan melakukan pendekatan ekstraktif, memilih dan menyalin kalimat penting dari dokumen sumber. Meskipun manusia juga dapat memotong dan menempelkan informasi yang relevan dari sebuah teks, seringkali mereka mengubah kalimat jika diperlukan, atau mereka menggabungkan informasi terkait yang berbeda ke dalam satu kalimat.

Teknik Peringkasan Independen Spesifik vs. Domain

Teknik peringkasan independen domain umumnya menerapkan serangkaian fitur umum yang dapat digunakan untuk mengidentifikasi segmen teks yang kaya informasi. Fokus penelitian baru-baru ini telah beralih ke teknik peringkasan khusus domain yang memanfaatkan pengetahuan yang tersedia khusus untuk domain teks. Sebagai contoh, penelitian peringkasan otomatis pada teks medis umumnya berusaha untuk memanfaatkan berbagai sumber pengetahuan medis dan ontologi yang dikodifikasikan.

Mengevaluasi Ringkasan Secara Kualitatif

Kelemahan utama dari sistem evaluasi yang ada sejauh ini adalah bahwa kami memerlukan setidaknya satu ringkasan referensi, dan untuk beberapa metode lebih dari satu, untuk dapat membandingkan ringkasan otomatis dengan model. Ini adalah tugas yang sulit dan mahal. Banyak upaya yang harus dilakukan untuk mendapatkan korpus teks dan ringkasannya yang sesuai. Selain itu, untuk beberapa metode, kita tidak hanya perlu memiliki ringkasan buatan manusia yang tersedia untuk perbandingan, tetapi juga penjelasan manual harus dilakukan di beberapa metode (misalnya SCU dalam Metode Piramida). Bagaimanapun, yang dibutuhkan metode evaluasi sebagai masukan, adalah seperangkat ringkasan untuk dijadikan standar emas dan satu set ringkasan otomatis. Selain itu, mereka semua melakukan evaluasi kuantitatif sehubungan dengan metrik kesamaan yang berbeda. Untuk mengatasi masalah ini, menurut kami evaluasi kuantitatif mungkin bukan satu-satunya cara untuk mengevaluasi ringkasan, dan evaluasi otomatis kualitatif juga penting.

3.8 Penggunaan Data Mining

Data Mining telah digunakan di banyak aplikasi. Beberapa contoh penggunaan yang penting adalah:

Games

Sejak awal 1960-an, dengan ketersediaan oracle untuk permainan kombinatorial tertentu, juga disebut tablebases (misalnya untuk catur 3x3) dengan konfigurasi awal apa pun, titik dan kotak papan kecil, heksa papan kecil, dan permainan akhir tertentu dalam catur, titik-dan-kotak, dan hex; area baru untuk data mining telah dibuka. Ini adalah ekstraksi strategi yang dapat digunakan manusia dari oracle ini. Pendekatan pengenalan pola saat ini tampaknya tidak sepenuhnya memperoleh abstraksi tingkat tinggi yang diperlukan untuk berhasil diterapkan. Sebaliknya, eksperimen ekstensif dengan tablebases - dikombinasikan dengan studi intensif dari tablebase-jawaban untuk masalah yang dirancang dengan baik, dan dengan pengetahuan seni sebelumnya (yaitu, pengetahuan pra-tablebase) - digunakan untuk menghasilkan pola yang mendalam. Berlekamp (dalam titik-dan-kotak, dll.) Dan John Nunn (dalam permainan akhir catur) adalah contoh penting dari peneliti yang melakukan pekerjaan ini, meskipun mereka tidak - dan tidak - terlibat dalam pembuatan tabel.

Bisnis

Di bidang bisnis, data mining adalah analisis aktivitas bisnis historis, disimpan sebagai data statis dalam database gudang data. Tujuannya adalah untuk mengungkap pola dan tren yang tersembunyi. Perangkat lunak penambangan data menggunakan algoritme pengenalan pola canggih untuk menyaring data dalam jumlah besar guna membantu menemukan informasi bisnis strategis yang sebelumnya tidak diketahui. Contoh dari apa yang bisnis menggunakan data mining adalah memasukkan melakukan analisis pasar untuk mengidentifikasi bundel produk baru, menemukan akar penyebab masalah manufaktur, untuk mencegah pengurangan pelanggan dan mendapatkan pelanggan baru, penjualan silang ke pelanggan yang sudah ada, dan membuat profil pelanggan dengan lebih akurat.

• Di dunia saat ini, data mentah dikumpulkan oleh perusahaan dengan kecepatan yang sangat tinggi. Misalnya, Walmart memproses lebih dari 20 juta transaksi point-of-sale setiap hari. Informasi ini disimpan dalam database terpusat, tetapi tidak akan berguna

tanpa beberapa jenis perangkat lunak data mining untuk menganalisisnya. Jika Walmart menganalisis data titik penjualan mereka dengan teknik penggalian data, mereka akan dapat menentukan tren penjualan, mengembangkan kampanye pemasaran, dan memprediksi loyalitas pelanggan dengan lebih akurat.

- Kategorisasi item yang tersedia di situs e-commerce adalah masalah mendasar. Sistem kategorisasi item yang benar sangat penting untuk pengalaman pengguna karena membantu menentukan item yang relevan untuknya untuk pencarian dan penjelajahan. Kategorisasi item dapat dirumuskan sebagai masalah klasifikasi terbimbing dalam data mining di mana kategorinya adalah kelas sasaran dan fiturnya adalah kata-kata yang menyusun beberapa deskripsi tekstual item. Salah satu pendekatannya adalah menemukan kelompok-kelompok yang pada awalnya serupa dan menempatkan mereka bersama-sama dalam kelompok laten. Sekarang diberikan item baru, pertama-tama klasifikasikan menjadi grup laten yang disebut klasifikasi tingkat kasar. Kemudian, lakukan klasifikasi putaran kedua untuk menemukan kategori item tersebut.
- Setiap kali kartu kredit atau kartu loyalitas toko digunakan, atau kartu garansi sedang diisi, data tentang perilaku pengguna sedang dikumpulkan. Banyak orang menganggap jumlah informasi yang disimpan tentang kami dari perusahaan, seperti Google, Facebook, dan Amazon, mengganggu dan mengkhawatirkan privasi. Meskipun ada potensi data pribadi kita digunakan dengan cara yang berbahaya, atau tidak diinginkan, data itu juga digunakan untuk membuat hidup kita lebih baik. Misalnya, Ford dan Audi berharap suatu hari dapat mengumpulkan informasi tentang pola mengemudi pelanggan sehingga mereka dapat merekomendasikan rute yang lebih aman dan memperingatkan pengemudi tentang kondisi jalan yang berbahaya.
- Data Mining dalam aplikasi manajemen hubungan pelanggan dapat memberikan kontribusi signifikan pada keuntungan. Daripada menghubungi prospek atau pelanggan secara acak melalui call center atau mengirim email, perusahaan dapat memusatkan upayanya pada prospek yang diperkirakan memiliki kemungkinan tinggi untuk menanggapi suatu penawaran. Metode yang lebih canggih dapat digunakan untuk mengoptimalkan sumber daya di seluruh kampanye sehingga seseorang dapat memprediksi saluran mana dan penawaran mana yang kemungkinan besar akan direspons oleh individu (di semua penawaran potensial). Selain itu, aplikasi canggih dapat digunakan untuk mengotomatiskan pengiriman surat. Setelah hasil dari data mining (calon prospek / pelanggan dan saluran / penawaran) ditentukan, "aplikasi canggih" ini dapat secara otomatis mengirim email atau surat biasa. Terakhir, dalam kasus di mana banyak orang akan mengambil tindakan tanpa tawaran, "model peningkatan" dapat digunakan untuk menentukan orang mana yang memiliki peningkatan respons

terbesar jika diberi tawaran. Dengan demikian, pemodelan Uplift memungkinkan pemasar untuk memfokuskan pengiriman dan penawaran pada orang yang dapat dibujuk, dan tidak mengirimkan penawaran kepada orang yang akan membeli produk tanpa penawaran. Pengelompokan data juga dapat digunakan untuk secara otomatis menemukan segmen atau grup dalam kumpulan data pelanggan.

Bisnis yang menggunakan data mining mungkin melihat laba atas investasi, tetapi mereka juga menyadari bahwa jumlah model prediktif dapat dengan cepat menjadi sangat besar. Misalnya, daripada menggunakan satu model untuk memprediksi berapa banyak pelanggan yang akan keluar, sebuah bisnis dapat memilih untuk membuat model terpisah untuk setiap wilayah dan jenis pelanggan. Dalam situasi di mana sejumlah besar model perlu dipertahankan, beberapa bisnis beralih ke metodologi data mining yang lebih otomatis.

- Data mining dapat membantu departemen sumber daya manusia (SDM) dalam mengidentifikasi karakteristik karyawan mereka yang paling sukses. Informasi yang diperoleh - seperti universitas yang dihadiri oleh karyawan yang sangat sukses - dapat membantu HR memfokuskan upaya perekrutan yang sesuai. Selain itu, aplikasi Manajemen Perusahaan Strategis membantu perusahaan menerjemahkan tujuan tingkat perusahaan, seperti target laba dan margin, menjadi keputusan operasional, seperti rencana produksi dan tingkat tenaga kerja.
- Analisis keranjang pasar, terkait dengan penggunaan data-mining dalam penjualan eceran. Jika toko pakaian mencatat pembelian pelanggan, sistem data mining dapat mengidentifikasi pelanggan yang lebih menyukai kemeja sutra daripada katun. Meskipun beberapa penjelasan tentang hubungan mungkin sulit, memanfaatkannya lebih mudah. Contoh tersebut berkaitan dengan aturan asosiasi dalam data berbasis transaksi. Tidak semua data berbasis transaksi dan logis, atau aturan yang tidak tepat mungkin juga ada dalam database.
- Analisis keranjang pasar telah digunakan untuk mengidentifikasi pola pembelian Konsumen Alpha. Menganalisis data yang dikumpulkan pada jenis pengguna ini memungkinkan perusahaan untuk memprediksi tren pembelian di masa depan dan memperkirakan permintaan penawaran.
- Data mining adalah alat yang sangat efektif dalam industri pemasaran katalog. Kataloger memiliki database yang kaya tentang riwayat transaksi pelanggan mereka untuk jutaan pelanggan sejak beberapa tahun yang lalu. Alat Data Mining dapat mengidentifikasi pola di antara pelanggan dan membantu mengidentifikasi pelanggan yang paling mungkin merespons kampanye pengiriman surat yang akan datang.

- Data Mining untuk aplikasi bisnis dapat diintegrasikan ke dalam proses pemodelan dan pengambilan keputusan yang kompleks. LIONsolver menggunakan kecerdasan bisnis Reaktif (RBI) untuk mendukung pendekatan "holistik" yang mengintegrasikan Data Mining, pemodelan, dan visualisasi interaktif ke dalam penemuan ujung-ke-ujung dan proses inovasi berkelanjutan yang didukung oleh pembelajaran manusia dan otomatis.
- Di bidang pengambilan keputusan, pendekatan RBI telah digunakan untuk menggali pengetahuan yang secara progresif diperoleh dari pengambil keputusan, dan kemudian menyesuaikan sendiri metode keputusan yang sesuai. Hubungan antara kualitas sistem data mining dan jumlah investasi yang bersedia dibuat oleh pembuat keputusan diformalkan dengan memberikan perspektif ekonomi tentang nilai "pengetahuan yang diekstraksi" dalam kaitannya dengan hasilnya bagi organisasi. Teori-keputusan ini Kerangka kerja klasifikasi diterapkan pada lini manufaktur wafer semikonduktor dunia nyata, di mana aturan keputusan untuk secara efektif memantau dan mengendalikan lini fabrikasi wafer semikonduktor dikembangkan.
- Contoh data mining yang terkait dengan lini produksi sirkuit terintegrasi (IC) dijelaskan dalam makalah "Menambang Data Uji IC untuk Mengoptimalkan Pengujian VLSI". Dalam makalah ini, aplikasi data mining dan analisis keputusan untuk masalah pengujian fungsional tingkat-mati dijelaskan. Eksperimen yang disebutkan menunjukkan kemampuan untuk menerapkan sistem Data Mining uji-mati historis untuk membuat model probabilistik pola kegagalan cetakan. Pola ini kemudian digunakan untuk memutuskan, dalam waktu nyata, mana yang mati untuk diuji berikutnya dan kapan harus menghentikan pengujian. Sistem ini telah terbukti, berdasarkan eksperimen dengan data uji historis, memiliki potensi untuk meningkatkan keuntungan pada produk IC yang matang. Contoh lain dari penerapan metodologi *Data Mining* di lingkungan manufaktur semikonduktor menunjukkan bahwa metodologi Data Mining mungkin sangat berguna ketika data langka, dan berbagai parameter fisik dan kimia yang mempengaruhi proses menunjukkan interaksi yang sangat kompleks. Implikasi lain adalah bahwa pemantauan on-line dari proses manufaktur semikonduktor menggunakan data mining mungkin sangat efektif.

Bidang Sains dan Teknik

Dalam beberapa tahun terakhir, data mining telah digunakan secara luas di bidang sains dan teknik, seperti bioinformatika, genetika, kedokteran, pendidikan, dan teknik tenaga listrik.

- Dalam studi genetika manusia, penambangan sekuens membantu mengatasi tujuan penting untuk memahami hubungan pemetaan antara variasi antar-individu dalam sekuens DNA manusia dan variabilitas dalam kerentanan penyakit. Secara sederhana, ini bertujuan untuk mengetahui bagaimana perubahan dalam urutan DNA seseorang memengaruhi risiko pengembangan penyakit umum seperti kanker, yang sangat penting untuk meningkatkan metode diagnosis, pencegahan, dan pengobatan penyakit ini. Salah satu metode data mining yang digunakan untuk melakukan tugas ini dikenal sebagai reduksi dimensi multifaktor.
- Di bidang teknik ketenagalistrikan, metode data mining telah banyak digunakan untuk pemantauan kondisi peralatan listrik tegangan tinggi. Tujuan dari pemantauan kondisi adalah untuk mendapatkan informasi berharga, misalnya, status insulasi (atau parameter terkait keselamatan penting lainnya). Teknik pengelompokan data - seperti self-organizing map (SOM), telah diterapkan pada pemantauan getaran dan analisis transformator on-load tap-changer (OLTCS). Dengan pemantauan getaran, dapat diamati bahwa setiap operasi pergantian tap menghasilkan sinyal yang berisi informasi tentang kondisi kontak tap changer dan mekanisme penggerak. Jelas, posisi tap yang berbeda akan menghasilkan sinyal yang berbeda. Namun, ada cukup banyak variabilitas di antara sinyal kondisi normal untuk posisi tap yang sama persis. SOM telah diterapkan untuk mendeteksi kondisi abnormal dan membuat hipotesis tentang sifat kelainan tersebut.
- Metode data mining telah diterapkan pada analisis gas terlarut (DGA) di transformator daya. DGA, sebagai diagnostik untuk transformator daya, telah tersedia selama bertahun-tahun.
- Metode seperti SOM telah diterapkan untuk menganalisis data yang dihasilkan dan untuk menentukan tren yang tidak jelas bagi metode rasio DGA standar (seperti Duval Triangle).
- Dalam penelitian pendidikan, di mana data mining telah digunakan untuk mempelajari faktor-faktor yang mengarahkan siswa untuk memilih terlibat dalam perilaku yang mengurangi pembelajaran mereka, dan untuk memahami faktor-faktor yang mempengaruhi retensi mahasiswa. Contoh serupa dari aplikasi sosial data mining adalah penggunaannya dalam sistem penemuan keahlian, di mana deskriptor keahlian manusia diekstraksi, dinormalisasi, dan diklasifikasikan untuk memfasilitasi penemuan para ahli, terutama di bidang ilmiah dan teknis. Dengan cara ini, data mining dapat memfasilitasi memori institusional.
- Metode data mining dari data biomedis yang difasilitasi oleh ontologi domain, data uji klinis mining, dan analisis lalu lintas menggunakan SOM.

- Dalam pengawasan reaksi obat yang merugikan, Pusat Pemantauan Uppsala, sejak tahun 1998, menggunakan metode penggalian data untuk secara rutin menyaring pola pelaporan yang menunjukkan masalah keamanan obat yang muncul dalam database global WHO tentang 4,6 juta dugaan insiden reaksi obat yang merugikan. Baru-baru ini, metodologi serupa telah dikembangkan untuk menambang koleksi besar catatan kesehatan elektronik untuk pola temporal yang menghubungkan resep obat dengan diagnosis medis.
- Data Mining telah diterapkan pada artefak perangkat lunak dalam bidang rekayasa perangkat lunak: Mining Software Repositories.

DI Bidang Hak asasi Manusia

Penggalian data dari catatan pemerintah - terutama catatan dari sistem peradilan (yaitu, pengadilan, penjara) - memungkinkan ditemukannya pelanggaran hak asasi manusia yang sistemik sehubungan dengan pembuatan dan publikasi catatan hukum yang tidak sah atau curang oleh berbagai badan pemerintah.

Di Bidang Medis

Beberapa algoritma pembelajaran mesin dapat diterapkan di bidang medis sebagai alat diagnostik pendapat kedua dan sebagai alat untuk fase ekstraksi pengetahuan dalam proses penemuan pengetahuan di database. Salah satu pengklasifikasi ini (disebut Prototype exemplar learning classifier (PEL-C) mampu menemukan sindrom serta kasus klinis atipikal.

Pada tahun 2011, kasus Sorrell v. IMS Health, Inc., yang diputuskan oleh Mahkamah Agung Amerika Serikat, memutuskan bahwa apotek dapat berbagi informasi dengan perusahaan luar. Praktik ini diotorisasi di bawah Amandemen Pertama Konstitusi, melindungi "kebebasan berbicara". Namun, berlakunya Health Information Technology for Economic and Clinical Health Act (HITECH Act) membantu memulai adopsi catatan kesehatan elektronik (EHR) dan teknologi pendukung di Amerika Serikat. Undang-Undang HITECH ditandatangani menjadi undang-undang pada 17 Februari 2009 sebagai bagian dari Undang-Undang Pemulihan dan Investasi Amerika (ARRA) dan membantu membuka pintu untuk penambangan data medis. Sebelum penandatanganan undang-undang ini, diperkirakan hanya 20% dari dokter yang berbasis di Amerika Serikat yang menggunakan catatan pasien elektronik. Søren Brunak mencatat bahwa "catatan pasien menjadi kaya informasi" dan dengan demikian "memaksimalkan peluang data mining." Oleh karena

itu, catatan pasien elektronik semakin memperluas kemungkinan terkait penggalian data medis sehingga membuka pintu ke sumber analisis data medis yang luas.

Data Mining Spasial

Penambangan data spasial adalah penerapan metode data mining ke data spasial. Tujuan akhir dari penambangan data spasial adalah menemukan pola dalam data yang berkaitan dengan geografi. Sejauh ini, penambangan data dan Sistem Informasi Geografis (GIS) telah ada sebagai dua teknologi terpisah, masing-masing dengan metode, tradisi, dan pendekatannya sendiri untuk visualisasi dan analisis data. Secara khusus, kebanyakan GIS kontemporer hanya memiliki fungsi analisis spasial yang sangat dasar. Ledakan besar dalam data yang direferensikan secara geografis disebabkan oleh perkembangan di bidang TI, pemetaan digital, penginderaan jauh, dan penyebaran GIS secara global menekankan pentingnya mengembangkan pendekatan induktif berbasis data untuk analisis dan pemodelan geografis.

Penambangan data menawarkan manfaat potensial yang besar untuk pengambilan keputusan terapan berbasis GIS. Baru-baru ini, tugas untuk mengintegrasikan kedua teknologi ini menjadi sangat penting, terutama karena berbagai organisasi sektor publik dan swasta yang memiliki database besar dengan data tematik dan referensi geografis mulai menyadari potensi besar dari informasi yang terkandung di dalamnya. Di antara organisasi tersebut adalah:

- Kantor yang membutuhkan analisis atau penyebaran data statistik geo-referensi
- Layanan kesehatan masyarakat yang mencari penjelasan tentang pengelompokan penyakit
- Lembaga lingkungan yang menilai dampak dari perubahan pola penggunaan lahan terhadap perubahan iklim
- Perusahaan geo-marketing melakukan segmentasi pelanggan berdasarkan lokasi spasial. Tantangan dalam penambangan spasial: Repositori data geospasial cenderung sangat besar. Selain itu, kumpulan data GIS yang ada sering kali dipecah menjadi komponen fitur dan atribut yang secara konvensional diarsipkan dalam sistem manajemen data hybrid. Persyaratan algoritme berbeda secara substansial untuk manajemen data relasional (atribut) dan untuk manajemen data topologi (fitur). Berkaitan dengan hal tersebut adalah ragam dan keragaman format data geografis, yang menghadirkan tantangan unik. Revolusi data geografis digital menciptakan jenis format data baru di luar format "vektor" dan "raster" tradisional. Penyimpanan data

geografis semakin banyak mencakup data yang tidak terstruktur, seperti citra dan multi-media referensi geografis.

Ada beberapa tantangan penelitian kritis dalam penemuan pengetahuan geografis dan penggalian data.

Miller dan Han menawarkan daftar topik penelitian yang muncul berikut di lapangan:

- Mengembangkan dan mendukung gudang data geografis (GDW's): Properti spasial sering kali direduksi menjadi atribut spasial sederhana di gudang data arus utama. Membuat GDW terintegrasi membutuhkan penyelesaian masalah interoperabilitas data spasial dan temporal - termasuk perbedaan dalam semantik, sistem referensi, geometri, akurasi, dan posisi.
- Representasi spasial-temporal yang lebih baik dalam penemuan pengetahuan geografis: Metode penemuan pengetahuan geografis (GKD) saat ini umumnya menggunakan representasi yang sangat sederhana dari objek geografis dan hubungan spasial. Metode penambangan data geografis harus mengenali objek geografis yang lebih kompleks (yaitu, garis dan poligon) dan hubungan (yaitu, jarak non-Euclidean, arah, konektivitas, dan interaksi melalui ruang geografis yang dikaitkan seperti medan). Selain itu, dimensi waktu perlu diintegrasikan sepenuhnya ke dalam representasi dan hubungan geografis ini.
- Penemuan pengetahuan geografi dengan menggunakan tipe data yang beragam: Metode GKD harus dikembangkan yang dapat menangani tipe data yang berbeda di luar model raster dan vektor tradisional, termasuk citra dan multimedia referensi geografis, serta tipe data dinamis (aliran video, animasi).

Penambangan Data Temporal

Data dapat berisi atribut yang dihasilkan dan direkam pada waktu yang berbeda. Dalam hal ini menemukan hubungan yang bermakna dalam data mungkin perlu mempertimbangkan urutan temporal atribut. Hubungan temporal dapat mengindikasikan hubungan kausal, atau hanya sebuah asosiasi.

Penambangan Data Sensor

Jaringan sensor nirkabel dapat digunakan untuk memfasilitasi pengumpulan data untuk penambangan data spasial untuk berbagai aplikasi seperti pemantauan polusi udara. Karakteristik dari jaringan tersebut adalah bahwa node sensor terdekat yang memantau fitur lingkungan biasanya mendaftarkan nilai yang sama. Jenis redundansi data ini karena korelasi spasial antara pengamatan sensor menginspirasi teknik untuk agregasi data dan

penambangan data dalam jaringan. Dengan mengukur korelasi spasial antara data sampel dengan sensor yang berbeda, kelas luas dari algoritma khusus dapat dikembangkan untuk mengembangkan algoritma penambangan data spasial yang lebih efisien.

Penambangan Data Visual

Dalam proses beralih dari analog ke digital, set data besar telah dihasilkan, dikumpulkan, dan disimpan menemukan pola statistik, tren dan informasi yang tersembunyi dalam data, untuk membangun pola prediksi. Studi menunjukkan penambangan data visual lebih cepat dan jauh lebih intuitif daripada penambangan data tradisional.

Penambangan Data Musik

Teknik penambangan data, dan khususnya analisis kejadian bersama, telah digunakan untuk menemukan kesamaan yang relevan di antara korpora musik (daftar radio, basis data CD) untuk keperluan termasuk Mengklasifikasikan musik ke dalam genre dengan cara yang lebih objektif.

Pengamatan

Penambangan data telah digunakan oleh pemerintah AS. Program tersebut mencakup program Total Information Awareness (TIA), Secure Flight (sebelumnya dikenal sebagai Computer-Assisted Passenger Prescreening System (CAPPS II)), Analysis, Dissemination, Visualization, Insight, Semantic Enhancement (ADVISE), dan Multi-state Anti- Pertukaran Informasi Terorisme (MATRIX). Program-program ini telah dihentikan karena kontroversi mengenai apakah mereka melanggar Amandemen ke-4 Konstitusi Amerika Serikat, meskipun banyak program yang dibentuk di bawahnya terus didanai oleh organisasi yang berbeda atau dengan nama yang berbeda.

Dalam konteks memerangi terorisme, dua metode data mining yang sangat masuk akal adalah "penambangan pola" dan "penambangan data berbasis subjek".

Pattern Mining

"Pattern mining" adalah metode data mining yang melibatkan menemukan pola yang ada dalam data. Dalam konteks ini pola sering diartikan sebagai aturan asosiasi. Motivasi asli untuk mencari aturan asosiasi berasal dari keinginan untuk menganalisis data transaksi supermarket, yaitu untuk memeriksa perilaku pelanggan dalam kaitannya dengan produk yang dibeli. Misalnya, aturan asosiasi "bir □ keripik kentang (80%)" menyatakan bahwa empat dari lima pelanggan yang membeli bir juga membeli keripik kentang.

Dalam konteks penambangan pola sebagai alat untuk mengidentifikasi aktivitas teroris, Dewan Riset Nasional memberikan definisi berikut: “Penggalian data berbasis pola mencari pola (termasuk pola data yang tidak wajar) yang mungkin terkait dengan aktivitas teroris - pola-pola ini mungkin dianggap sebagai sinyal kecil di lautan kebisingan yang luas. " Pattern Mining mencakup area baru seperti Music Information Retrieval (MIR) di mana pola yang terlihat baik dalam domain temporal dan non temporal diimpor ke metode pencarian penemuan pengetahuan klasik.

Data Mining berbasis subjek

"Penambangan data berbasis subjek" adalah metode penambangan data yang melibatkan pencarian asosiasi antar individu dalam data. Dalam konteks pemberantasan terorisme, National Research Council memberikan definisi sebagai berikut: “Penambangan data berbasis subjek menggunakan individu pemrakarsa atau datum lain yang dianggap, berdasarkan informasi lain, memiliki minat tinggi, dan tujuannya adalah untuk menentukan apa orang lain atau transaksi keuangan atau pergerakan, dll., terkait dengan data awal itu. "

Knowledge Grid

Penemuan pengetahuan “On the Grid” umumnya mengacu pada melakukan penemuan pengetahuan di lingkungan terbuka menggunakan konsep komputasi grid, memungkinkan pengguna untuk mengintegrasikan data dari berbagai sumber data online, serta menggunakan sumber daya jarak jauh, untuk menjalankan tugas penambangan data mereka. Contoh paling awal adalah Discovery Net, yang dikembangkan di Imperial College London, yang memenangkan "Penghargaan Aplikasi Intensif Data Paling Inovatif" di konferensi dan pameran ACM SC02 (Supercomputing 2002), berdasarkan demonstrasi aplikasi penemuan pengetahuan terdistribusi yang sepenuhnya interaktif untuk aplikasi bioinformatika. Contoh lain termasuk pekerjaan yang dilakukan oleh para peneliti di University of Calabria, yang mengembangkan arsitektur Knowledge Grid untuk penemuan pengetahuan terdistribusi, berdasarkan komputasi grid.

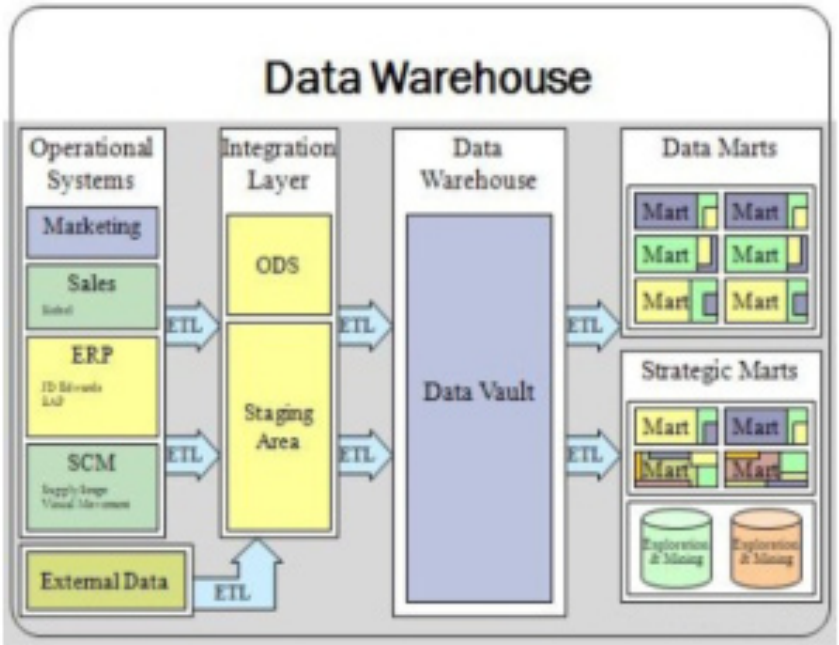
Bab 4 : Data Warehouse

Data Warehouse (Indonesia : Gudang Data) adalah inti dari intelijen bisnis. Ini terutama digunakan untuk melaporkan dan menganalisis data. Data mart, manajemen data master, dimensi, dimensi yang berubah perlahan dan skema bintang.

4.1 Pengertian

Data warehouse merupakan perpaduan dari berbagai teknologi yang membantu mengelola penggunaan data. Sebagai penyimpanan elektronik, data warehouse mampu menyimpan informasi bisnis dalam jumlah yang besar baik untuk query maupun analisis, dan bukan untuk pemrosesan transaksi. Data dari Data warehouse ini bentuknya informasi dan tersedia secara real-time. Data Warehouse adalah kumpulan data dari berbagai sumber yang menyediakan data-data perusahaan untuk kepentingan bisnis. Jika diartikan secara harfiah dalam Bahasa Indonesia, Data Warehouse adalah gudang data.

Sebagaimana fungsi gudang pada umumnya, data warehouse adalah sistem komputer BI (Business Intelligence) yang dibangun untuk menyimpan berbagai data. Data-data ini berupa data penjualan, gaji, dan informasi harian lainnya.



Data Warehouse *Overview*

Data yang disimpan di gudang diunggah dari sistem operasional (seperti pemasaran atau penjualan). Data dapat melewati penyimpanan data operasional untuk operasi tambahan sebelum digunakan dalam DW untuk pelaporan.

Jenis-Jenis Data Warehouse

Data Mart

Data mart adalah bentuk sederhana dari gudang data yang difokuskan pada satu subjek (atau area fungsional), karenanya mereka mengambil data dari sejumlah sumber terbatas seperti penjualan, keuangan, atau pemasaran. Data mart sering dibangun dan dikendalikan oleh satu departemen dalam suatu organisasi. Sumber dapat berupa sistem operasional internal, gudang data pusat, atau data eksternal. Denormalisasi adalah norma untuk teknik pemodelan data dalam sistem ini. Mengingat bahwa data mart umumnya hanya mencakup sebagian dari data yang terkandung dalam data warehouse, mereka seringkali lebih mudah dan lebih cepat untuk diimplementasikan.

Jenis-jenis Data mart

- Dependent data mart
- Independent data mart
- Hybrid data mart

Online Analytical Processing (OLAP)

Online Analytical Processing (OLAP) merupakan suatu metode pendekatan untuk menyajikan jawaban dari permintaan proses analisis yang bersifat dimensional secara cepat, yaitu desain dari aplikasi dan teknologi yang dapat mengoleksi, menyimpan, memanipulasi suatu data multidimensi untuk tujuan analisis. OLAP adalah teknologi yang memproses data di dalam data warehouse dalam struktur multidimensi, menyediakan jawaban yang cepat untuk query analisis yang kompleks. OLAP seringkali disebut analisis data multidimensi. Data multidimensi adalah data yang dapat dimodelkan sebagai atribut dimensi dan atribut ukuran. Contoh atribut dimensi adalah nama barang dan warna barang, sedangkan contoh atribut ukuran adalah jumlah barang.

Tekhnik OLAP

Teknik OLAP dapat dirangkum menjadi 5 garis besar yaitu Fast Analysis of Shared Multidimensional Information atau disingkat menjadi FASMI yang masing-masing berarti sebagai berikut:

FAST, berarti sistem ditargetkan untuk memberikan response terhadap user dengan secepat mungkin, sesuai dengan analisis yang dilakukan.

ANALYSIS, berarti sistem dapat mengatasi berbagai logika bisnis dan analisis statistik yang relevan dengan aplikasi dan user, dan mudah.

SHARED, berarti sistem melaksanakan seluruh kebutuhan pengamanan data, jika dibutuhkan banyak akses penulisan terhadap data, disesuaikan dengan level dari user. Tidak semua aplikasi membutuhkan user untuk menulis data kembali. Sistem harus dapat meng-handle multiple update dalam satu waktu secara aman.

MULTIDIMENSIONAL, berarti sistem harus menghasilkan conceptual view dari data secara multidimensional, meliputi full support untuk hierarki dan mutiple hierarki. Hal ini merupakan cara yang logic untuk menganalisis bisnis dan organisasi.

INFORMATION, adalah semua data dan informasi yang dibutuhkan dan relevan untuk aplikasi. Kapasitas produk OLAP berbeda untuk menghandle input data tergantung beberapa pertimbangan meliputi duplikasi data, RAM yang dibutuhkan, penggunaan disk space, performance, integrasi dengan data warehouse, dan lainnya.

On-Line Transaction Processing (OLTP)

Merupakan suatu pemrosesan yang menyimpan data mengenai kegiatan operasional transaksi sehari-hari. OLTP ditandai dengan sejumlah besar transaksi online pendek (INSERT, UPDATE, DELETE). Sistem OLTP menekankan pemrosesan kueri yang sangat cepat dan menjaga integritas data dalam lingkungan multi-akses. Untuk sistem OLTP, efektivitas diukur dengan jumlah transaksi per detik. Database OLTP berisi data terperinci dan terkini. Skema yang digunakan untuk menyimpan database transaksional adalah model entitas (biasanya 3NF). Normalisasi adalah norma untuk teknik pemodelan data dalam sistem ini.

Analisis Prediktif

Analisis prediktif adalah tentang menemukan dan mengukur pola tersembunyi dalam data menggunakan model matematika kompleks yang dapat digunakan untuk memprediksi hasil di masa mendatang. Analisis prediktif berbeda dari OLAP karena OLAP berfokus pada analisis data historis dan bersifat reaktif, sedangkan analisis prediktif berfokus pada masa depan. Sistem ini juga digunakan untuk *Customer Relationship Management*/CRM (manajemen hubungan pelanggan).

Komponen Data Warehouse

Source Data

Source data merupakan gudang data dari berbagai sumber, yakni dari data internal, data eksternal, archived, dan lain-lain.

Data Stagging

Data yang diekstrak dan load dalam format yang sama yang tidak merubah nilai data.

Data Storage

Media penyimpanan data yang dihasilkan dari data stagging.

Metadata

Komponen yang memberi penjelasan tentang data melebihi kamus data. Metadata terbagi atas Metadata Operasional, Metadata Transformasi & Ekstrak, serta Metadata User.

Information Delivery

Penyampaian informasi pada pengguna yang mana terdapat teknik online, infraned serta email.

Management and Control

Merupakan pengelolaan serta pengendalian yang ada pada data stagging serta metadata

Manfaat Data Warehouse

Sebuah gudang data menyimpan salinan informasi dari sistem transaksi sumber. Kompleksitas arsitektur ini memberikan peluang untuk:

- Mengintegrasikan data dari berbagai sumber ke dalam satu database dan model data. Sekadar penggabungan data ke database tunggal sehingga mesin kueri tunggal dapat digunakan untuk menyajikan data adalah ODS.
- Mengurangi masalah pertentangan kunci level isolasi database dalam sistem pemrosesan transaksi yang disebabkan oleh upaya untuk menjalankan kueri analisis yang besar dan berjalan lama dalam database pro transaksi.
- Menjaga riwayat data, meskipun sistem transaksi sumber tidak.
- Mengintegrasikan data dari berbagai sistem sumber, yang memungkinkan tampilan terpusat di seluruh perusahaan. Manfaat ini selalu berharga, tetapi terutama jika organisasi telah berkembang melalui merger.

- Meningkatkan kualitas data, dengan memberikan kode dan deskripsi yang konsisten, menandai atau bahkan memperbaiki data yang buruk.
- Menyajikan informasi organisasi secara konsisten.
- Menyediakan satu model data umum untuk semua data minat apa pun sumber datanya.
- Mengatur ulang data agar masuk akal bagi pengguna bisnis.
- Mengatur ulang data sehingga memberikan kinerja kueri yang sangat baik, bahkan untuk kueri analitik yang kompleks, tanpa memengaruhi sistem operasional.
- Menambahkan nilai pada aplikasi bisnis operasional, terutama sistem manajemen hubungan pelanggan (CRM).
- Membuat kueri dukungan keputusan lebih mudah untuk ditulis.
- Arsitektur gudang data yang dioptimalkan memungkinkan ilmuwan data untuk mengatur dan membedakan data berulang.

Konsep Dasar Data Warehouse

- Data-data yang berorientasi pada subyek, mempunyai dimensi waktu, terintegrasi, dan merupakan koleksi lengkap, yang dipergunakan dalam rangka mendukung proses pengambilan keputusan yang dilakukan oleh para manager pada tiap-tiap jenjang. Terutama jenjang manager dengan peringkat yang tinggi.
- Pusat repousitori informasi yang dapat memberikan database yang berorientasi pada subyek bagi informasi yang bersifat historis yang memberi dukungan Decision Support System (DSS) serta Executive Information System (EIS).
- Merupakan data yang didapatkan dari proses yang mana sebuah organisasi mengekstraksi makna dari aset informasi yang dimiliki. Data warehouse merupakan penemuan baru dalam bidang teknologi informasi. Dimulai sekitar 15 tahun lalu konsep data warehouse terus berkembang secara pesat sehingga konsep ini merupakan konsep yang paling banyak diperbincangkan oleh para ahli.
- Adalah salinan dari transaksi data yang terstruktur secara spesifik pada query serta analisa.
- Merupakan salinan dari transaksi data yang terstruktur secara spesifik pada query serta laporan.

Sejarah Perkembangan

Konsep pergudangan data dimulai pada akhir 1980-an ketika peneliti IBM Barry Devlin dan Paul Murphy mengembangkan "gudang data bisnis". Intinya, konsep data

warehousing dimaksudkan untuk menyediakan model arsitektural untuk aliran data dari sistem operasional ke lingkungan pendukung keputusan. Konsep tersebut berusaha untuk mengatasi berbagai masalah yang terkait dengan aliran ini, terutama biaya tinggi yang terkait dengannya. Dengan tidak adanya arsitektur data warehousing, diperlukan sejumlah besar redundansi untuk mendukung berbagai lingkungan pendukung keputusan. Di perusahaan yang lebih besar, biasanya lingkungan pendukung keputusan ganda beroperasi secara independen. Meskipun setiap lingkungan melayani pengguna yang berbeda, mereka seringkali membutuhkan banyak data tersimpan yang sama. Proses pengumpulan, pembersihan, dan pengintegrasian data dari berbagai sumber, biasanya dari sistem operasional jangka panjang yang ada (biasanya disebut sebagai sistem warisan), biasanya sebagian direplikasi untuk setiap lingkungan. Selain itu, sistem operasional sering dikaji ulang saat persyaratan pendukung keputusan baru muncul. Seringkali persyaratan baru mengharuskan pengumpulan, pembersihan, dan pengintegrasian data baru dari "data mart" yang disesuaikan untuk akses yang siap oleh pengguna.

Perkembangan utama di tahun-tahun awal data warehousing adalah:

- 1960-an - General Mills dan Dartmouth College, dalam proyek penelitian bersama, mengembangkan istilah dimensi dan fakta.
- 1970-an - ACNielsen dan IRI menyediakan data mart dimensi untuk penjualan ritel.
- 1970-an - Bill Inmon mulai mendefinisikan dan membahas istilah: Gudang Data.
- 1975 - Sperry Univac memperkenalkan MAPPER (MAintain, Mempersiapkan, dan Menghasilkan Laporan Eksekutif) adalah manajemen database dan sistem pelaporan yang mencakup 4GL pertama di dunia. Platform pertama yang dirancang untuk membangun Pusat Informasi (cikal bakal platform Enterprise Data Warehousing kontemporer)
- 1983 - Teradata memperkenalkan sistem manajemen basis data yang dirancang khusus untuk dukungan keputusan.
- 1984 - Metaphor Computer Systems, didirikan oleh David Liddle dan Don Massaro, merilis Data Interpretation System (DIS). DIS adalah paket perangkat keras / perangkat lunak dan GUI untuk pengguna bisnis untuk membuat manajemen database dan sistem analitik.
- 1988 - Barry Devlin dan Paul Murphy menerbitkan artikel Arsitektur untuk bisnis dan sistem informasi di mana mereka memperkenalkan istilah "gudang data bisnis".

- 1990 - Red Brick Systems, didirikan oleh Ralph Kimball, memperkenalkan Red Brick Warehouse, sebuah sistem manajemen database khusus untuk data warehousing.
- 1991 - Prism Solutions, didirikan oleh Bill Inmon, memperkenalkan Prism Warehouse Manager, perangkat lunak untuk mengembangkan gudang data.
- 1992 - Bill Inmon menerbitkan buku Membangun Gudang Data.
- 1995 - Data Warehousing Institute, sebuah organisasi nirlaba yang mempromosikan gudang data, didirikan.
- 1996 - Ralph Kimball menerbitkan buku The Data Warehouse Toolkit.
- 2012 - Bill Inmon mengembangkan dan membuat teknologi publik yang dikenal sebagai "textual disambiguation". Disambiguasi tekstual menerapkan konteks ke teks mentah dan memformat ulang teks mentah dan konteks ke dalam format basis data standar. Setelah teks mentah melewati disambiguasi tekstual, teks mentah dapat diakses dan dianalisis dengan mudah dan efisien oleh teknologi intelijen bisnis standar. Disambiguasi tekstual dicapai melalui eksekusi ETL tekstual. Disambiguasi tekstual berguna dimanapun teks mentah ditemukan, seperti dalam dokumen, Hadoop, email, dan lain sebagainya.

Penyimpanan Informasi

Fakta

Fakta adalah nilai atau ukuran, yang merepresentasikan fakta tentang entitas atau sistem yang dikelola. Fakta yang dilaporkan oleh entitas pelapor dikatakan berada pada level mentah. Misalnya. Dalam sistem telepon seluler, jika BTS (base transceiver station) menerima 1.000 permintaan untuk alokasi saluran lalu lintas, dialokasikan untuk 820 dan menolak sisanya maka akan melaporkan 3 fakta atau pengukuran ke sistem manajemen:

- tch_req_total = 1000
- tch_req_success = 820
- tch_req_fail = 180

Fakta di tingkat mentah selanjutnya dikumpulkan ke tingkat yang lebih tinggi dalam berbagai dimensi untuk mengekstrak lebih banyak layanan atau informasi yang relevan dengan bisnis darinya. Ini disebut agregat atau ringkasan atau kumpulan fakta.

Misalnya jika dalam suatu kota terdapat 3 BTS, maka fakta-fakta di atas dapat diagregasikan dari BTS tersebut ke tingkat kota dalam dimensi jaringan. Sebagai contoh:

- *tch _ req _ sukses _ kota tch _ req _ sukses _ bts1 tch _ req _ success _ bts2 tch _ req _ success _ bts3*
- *avg_tch_req_success_city (tch_req_success_bts1 tch_req_success_bts2 tch_ req_success_bts3) / 3*

Pendekatan Dimensi versus Normalisasi dalam Penyimpanan Data

Ada tiga atau lebih pendekatan utama untuk menyimpan data di gudang data - pendekatan yang paling penting adalah pendekatan dimensi dan pendekatan normalisasi.

Pendekatan dimensional mengacu pada pendekatan Ralph Kimball yang menyatakan bahwa data warehouse harus dimodelkan menggunakan Model Dimensi / skema bintang. Pendekatan yang dinormalisasi, disebut juga dengan model 3NF (Third Normal Form) mengacu pada pendekatan Bill Inmon yang menyatakan bahwa data warehouse harus dimodelkan menggunakan model ER / model normalisasi.

Dalam pendekatan dimensional, data transaksi dipartisi menjadi “fakta”, yang umumnya merupakan data transaksi numerik, dan “dimensi”, yang merupakan informasi referensi yang memberikan konteks pada fakta. Misalnya, transaksi penjualan dapat dipecah menjadi fakta-fakta seperti jumlah produk yang dipesan dan harga total yang dibayarkan untuk produk tersebut, dan menjadi dimensi seperti tanggal pemesanan, nama pelanggan, nomor produk, pengiriman pesanan, dan tagih ke. lokasi, dan staf penjualan yang bertanggung jawab untuk menerima pesanan.

Keuntungan utama dari pendekatan dimensional adalah bahwa data warehouse lebih mudah dipahami dan digunakan oleh pengguna. Selain itu, pengambilan data dari gudang data cenderung beroperasi dengan sangat cepat. Struktur dimensi mudah dipahami oleh pengguna bisnis, karena struktur terbagi menjadi ukuran / fakta dan konteks / dimensi. Fakta terkait dengan proses bisnis organisasi dan sistem operasional sedangkan dimensi yang mengelilinginya mengandung konteks tentang pengukuran (Kimball, Ralph 2008). Keuntungan lain yang ditawarkan oleh model dimensional adalah tidak melibatkan database relasional setiap saat. Dengan demikian, jenis teknik pemodelan ini sangat berguna untuk kueri pengguna akhir di gudang data.
Kerugian utama dari pendekatan dimensional adalah sebagai berikut:

1. Untuk menjaga integritas fakta dan dimensi, memuat data warehouse dengan data dari sistem operasional yang berbeda rumit.

2. Sulit untuk mengubah struktur gudang data jika organisasi yang mengadopsi pendekatan dimensional mengubah cara berbisnisnya.

Dalam pendekatan yang dinormalisasi, data di gudang data disimpan dengan mengikuti aturan normalisasi database. Tabel dikelompokkan berdasarkan bidang subjek yang mencerminkan kategori data umum (misalnya, data tentang pelanggan, produk, keuangan, dll.). Struktur yang dinormalisasi membagi data menjadi entitas, yang membuat beberapa tabel dalam database relasional. Ketika diterapkan di perusahaan besar, hasilnya adalah lusinan tabel yang ditautkan bersama oleh jaringan gabungan. Selanjutnya, setiap entitas yang dibuat diubah menjadi tabel fisik terpisah ketika database diimplementasikan (Kimball, Ralph 2008). Keuntungan utama dari pendekatan ini adalah menambahkan informasi ke dalam database secara langsung. Beberapa kelemahan dari pendekatan ini adalah, karena banyaknya tabel yang terlibat, akan sulit bagi pengguna untuk menggabungkan data dari sumber yang berbeda menjadi informasi yang berarti dan untuk mengakses informasi tanpa pemahaman yang tepat tentang sumber data dan struktur data dari gudang data.

Kedua model yang dinormalisasi dan dimensional dapat direpresentasikan dalam diagram hubungan-entitas karena keduanya berisi tabel relasional yang digabungkan. Perbedaan antara kedua model tersebut adalah derajat normalisasi (juga dikenal sebagai Bentuk Normal). Pendekatan ini tidak saling eksklusif, dan ada pendekatan lain. Pendekatan dimensi dapat melibatkan normalisasi data sampai tingkat tertentu (Kimball, Ralph 2008).

Dalam Bisnis Berbasis Informasi, Robert Hillard mengusulkan pendekatan untuk membandingkan dua pendekatan berdasarkan kebutuhan informasi dari masalah bisnis. Teknik ini menunjukkan bahwa model yang dinormalisasi menyimpan lebih banyak informasi daripada padanan dimensinya (bahkan ketika bidang yang sama digunakan di kedua model) tetapi informasi tambahan ini mengorbankan kegunaan. Teknik mengukur kuantitas informasi dalam kaitannya dengan entropi informasi dan kegunaan dalam hal ukuran transformasi data *Small-Worlds*.

Metode Desain

Desain Bottom-up

Dalam pendekatan bottom-up, data mart pertama kali dibuat untuk menyediakan pelaporan dan kemampuan analitis untuk proses bisnis tertentu. Data mart ini kemudian dapat diintegrasikan untuk membuat gudang data yang komprehensif. Arsitektur bus gudang data pada dasarnya merupakan implementasi dari "bus", kumpulan dimensi yang sesuai dan fakta yang sesuai, yang merupakan dimensi yang dibagi (dengan cara tertentu) antara fakta dalam dua atau lebih data mart.

Desain Top-down

Pendekatan top-down dirancang menggunakan model data perusahaan yang dinormalisasi. Data "atom", yaitu data pada tingkat detail terbesar, disimpan di gudang data. Data mart dimensi yang berisi data yang diperlukan untuk proses bisnis tertentu atau departemen tertentu dibuat dari gudang data.

Desain Hybrid

Gudang data (DW) sering kali menyerupai arsitektur hub dan jari-jari. Sistem warisan yang memberi makan gudang sering kali mencakup manajemen hubungan pelanggan dan perencanaan sumber daya perusahaan, menghasilkan data dalam jumlah besar. Untuk mengkonsolidasikan berbagai model data ini, dan memfasilitasi proses pemuatan transformasi ekstrak, gudang data sering kali menggunakan penyimpanan data operasional, informasi yang darinya diurai menjadi DW yang sebenarnya. Untuk mengurangi redundansi data, sistem yang lebih besar sering kali menyimpan data dengan cara yang dinormalisasi. Data mart untuk laporan tertentu kemudian dapat dibangun di atas DW.

Database DW dalam solusi hybrid disimpan dalam bentuk normal ketiga untuk menghilangkan redundansi data. Database relasional normal, bagaimanapun, tidak efisien untuk laporan intelijen bisnis di mana pemodelan dimensional lazim. Data mart kecil dapat menyimpan data dari gudang terkonsolidasi dan menggunakan data spesifik yang difilter untuk tabel fakta dan dimensi yang diperlukan. DW menyediakan satu sumber informasi dari mana data mart dapat membaca, menyediakan berbagai informasi bisnis. Arsitektur hybrid memungkinkan DW diganti dengan solusi manajemen data master di mana informasi operasional, bukan statis dapat berada.

Komponen Pemodelan Data Vault mengikuti arsitektur hub dan jari-jari. Gaya pemodelan ini adalah desain hibrid, yang terdiri dari praktik terbaik dari bentuk normal ketiga dan skema bintang. Model Data Vault bukanlah bentuk normal ketiga yang sebenarnya, dan melanggar beberapa aturannya, tetapi merupakan arsitektur top-down dengan desain bottom-up. Model Data Vault diarahkan untuk menjadi gudang data. Itu tidak diarahkan untuk dapat diakses pengguna akhir, yang ketika dibangun, masih membutuhkan penggunaan data mart atau area rilis berbasis skema bintang untuk tujuan bisnis.

Sistem Operasional

Sistem operasional dioptimalkan untuk menjaga integritas data dan kecepatan pencatatan transaksi bisnis melalui penggunaan normalisasi basis data dan model hubungan entitas. Perancang sistem operasional biasanya mengikuti aturan Codd tentang normalisasi basis data untuk memastikan integritas data. Codd mendefinisikan lima aturan normalisasi yang semakin ketat. Desain database yang sepenuhnya dinormalisasi (yaitu, yang memenuhi semua lima aturan Codd) sering menghasilkan informasi dari transaksi bisnis yang disimpan dalam lusinan hingga ratusan tabel. Database relasional efisien dalam mengelola hubungan antara tabel-tabel ini. Basis data memiliki kinerja penyisipan / pembaruan yang sangat cepat karena hanya sejumlah kecil data dalam tabel tersebut yang terpengaruh setiap kali transaksi diproses. Akhirnya, untuk meningkatkan kinerja, data lama biasanya dibersihkan secara berkala dari sistem operasional. Gudang data dioptimalkan untuk pola akses analisis.

Pola akses analisis umumnya melibatkan pemilihan bidang tertentu dan jarang jika ‘select*' seperti yang lebih umum dalam database operasional. Karena perbedaan dalam pola akses ini, basis data operasional (secara longgar, OLTP) mendapat manfaat dari penggunaan DBMS berorientasi baris sedangkan basis data analisis (secara longgar, OLAP) mendapat manfaat dari penggunaan DBMS berorientasi kolom. Tidak seperti sistem operasional yang mempertahankan snapshot bisnis, gudang data umumnya mempertahankan sejarah tak terbatas yang diimplementasikan melalui proses ETL yang secara berkala memigrasikan data dari sistem operasional ke gudang data.

Perkembangan dan Evolusi Data Warehouse

Istilah-istilah ini mengacu pada tingkat kecanggihan sebuah gudang data:

Offline operational data warehouse

Gudang data dalam tahap evolusi ini diperbarui pada siklus waktu reguler (biasanya harian, mingguan atau bulanan) dari sistem operasional dan data disimpan dalam data berorientasi pelaporan terintegrasi

Off-line Data Warehouse

Gudang data pada tahap ini diperbarui dari data dalam sistem operasional secara teratur dan data gudang data disimpan dalam struktur data yang dirancang untuk memudahkan pelaporan.

On-Time Data Warehouse

Data Warehousing Terintegrasi Online merepresentasikan data tahap gudang data secara real time di update untuk setiap transaksi yang dilakukan pada sumber data.

Integrated data warehouse

Gudang data ini mengumpulkan data dari berbagai area bisnis, sehingga pengguna dapat mencari informasi yang mereka butuhkan di sistem lain.

4.2 Data Mart

Data mart adalah lapisan akses dari lingkungan data warehouse yang digunakan untuk mengeluarkan data ke pengguna. Data mart adalah bagian dari gudang data dan biasanya berorientasi pada lini bisnis atau tim tertentu. Sedangkan gudang data memiliki kedalaman perusahaan yang luas, informasi di data mart berkaitan dengan satu departemen. Dalam beberapa penerapan, setiap departemen atau unit bisnis dianggap sebagai pemilik data mart termasuk semua perangkat keras, perangkat lunak, dan data. Hal ini memungkinkan setiap departemen untuk mengisolasi penggunaan, manipulasi, dan pengembangan data mereka. Dalam penerapan lain di mana dimensi yang sesuai digunakan, kepemilikan unit bisnis ini tidak akan berlaku untuk dimensi bersama seperti pelanggan, produk, dll.

Organisasi membangun gudang data dan data mart karena informasi dalam database tidak diatur sedemikian rupa sehingga membuatnya mudah diakses, memerlukan kueri yang terlalu rumit atau menghabiskan sumber daya.

Sementara database transaksional dirancang untuk diperbarui, gudang data atau mart hanya dapat dibaca. Gudang data dirancang untuk mengakses kelompok besar catatan terkait. Data mart meningkatkan waktu respons pengguna akhir dengan memungkinkan pengguna memiliki akses ke jenis data tertentu yang paling sering mereka perlukan dengan menyediakan data dengan cara yang mendukung tampilan kolektif sekelompok pengguna.

Data mart pada dasarnya adalah versi data warehouse yang dipadatkan dan lebih fokus yang mencerminkan regulasi dan spesifikasi proses dari setiap unit bisnis dalam suatu organisasi. Setiap data mart didedikasikan untuk fungsi atau wilayah bisnis tertentu. Subkumpulan data ini dapat menjangkau banyak atau semua area subjek fungsional perusahaan. Biasanya beberapa data mart digunakan untuk melayani kebutuhan setiap unit bisnis (data mart yang berbeda dapat digunakan untuk mendapatkan informasi spesifik untuk berbagai departemen perusahaan, seperti akuntansi, pemasaran, penjualan, dll.).

Istilah terkait spreadmart adalah label merendahkan yang menggambarkan situasi yang terjadi ketika satu atau lebih analis bisnis mengembangkan sistem spreadsheet tertaut untuk melakukan analisis bisnis, kemudian mengembangkannya ke ukuran dan tingkat kerumitan yang membuatnya hampir mustahil untuk dipertahankan..

Perbandingan Data Warehouse dengan Data Mart

	Data Warehouse	Data Mart
Lingkupan (Scope)	*Perusahaan*/ Enterprise	*Departemen*
Subjek (Subjects)	Multiple	Single
Sumber data (Data Sources)	*Banyak*	*Sedikit*
Ukuran data	100 GB to > 1 TB	< 100 GB
Waktu implementasi	*Berbulan-bulan bahkan sampai bertahun-tahun.*	*Beberapa bulan*

4.2.1 Skema Desain

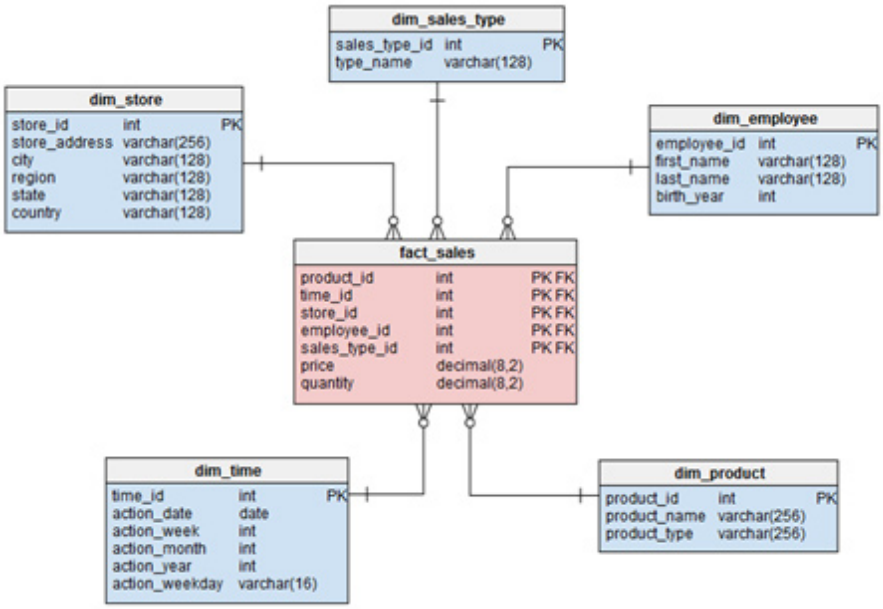
Model yang sering digunakan di dalam data warehouse saat ini adalah skema bintang dan skema snowflake. Masing-masing model tentunya memiliki kelebihan dan kekurangannya

masing-masing. Dalam artikel ini dijelaskan dengan detil mengenai perbedaan kedua skema tersebut. Selain itu dijelaskan pula kondisi-kondisi yang sesuai di dalam mengimplementasikan skema bintang maupun skema snowflake.

Skema bintang dan skema snowflake adalah sarana untuk mengorganisir data mart – data mart atau gudang-gudang data dengan menggunakan basis data relasional. Kedua skema tersebut menggunakan tabel-tabel dimensi untuk mendeskripsikan data-data yang terdapat di dalam tabel fakta.

Setiap perusahaan pada umumnya menjual produk, pengetahuan, maupun jasa. Sehingga sistem penjualan adalah sebuah sistem yang terdapat di sebagian besar perusahaan. Berikut ini dijelaskan mengenai model penjualan baik skema bintang maupun skema snowflake.

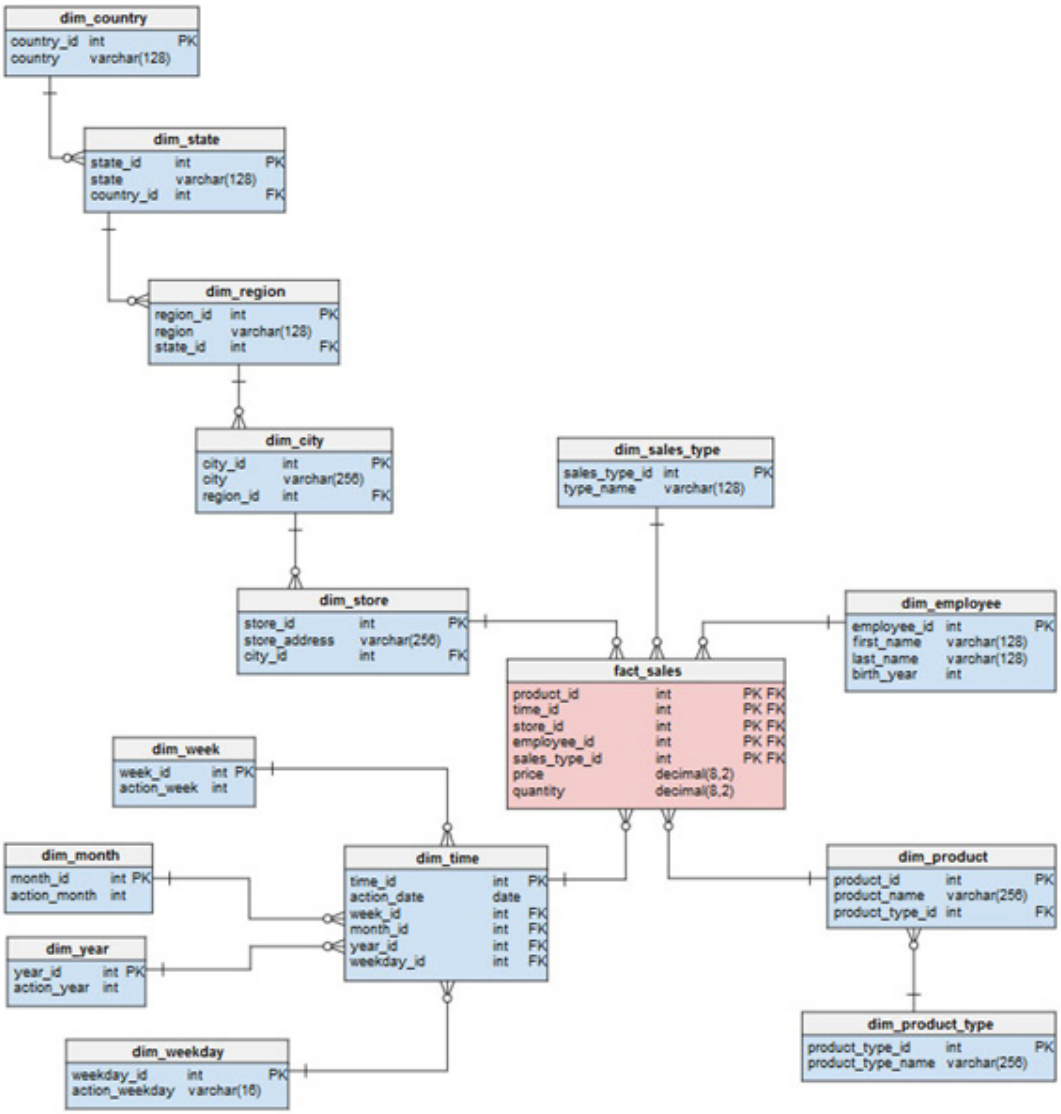
Skema Bintang



Karakteristik utama dari skema bintang adalah bahwa tabel dimensinya tidak dinormalisasi. Pada model di atas, tabel fakta fact_sales (warna merah muda)berisi data-data yang diekstrakdari database operasional. Sedangkan tabel yang berwarna biru muda adalah tabel dimensi. Pada gambar di atas terdapat lima tabel dimensi yaitu dim_sales_type, dim_store, dim_employee, dim_product, dan dim_time.

Dari model ini, kita dapat dengan mudah melihat mengapa skema ini disebut ‘skema bintang’, karena model tersebut terlihat seperti bintang, dengan tabel dimensi yang mengelilingi tabel fakta

Skema Snowflake



Skema snowflake juga menyimpan data yang sama seperti pada skema bintang. Tabel fakta yang digunakan pada skema bintang maupun pada skema snowflake berisi field-field yang sama. Perbedaan utama antara skema bintang dan skema snowflake adalah semua tabel dimensi pada skema snowflake telah dinormalisasi. Proses normalisasi tabel-tabel dimensi pada skema snowflake ini disebut dengan proses snowflaking,sehingga tampilan tabel-tabel pada skema snowflake bentuknya menyerupai snowflake.

Alasan untuk Membuat Mart Data

- Akses mudah ke data yang sering dibutuhkan
- Membuat tampilan kolektif oleh sekelompok pengguna
- Meningkatkan waktu respons pengguna akhir
- Kemudahan pembuatan
- Biaya lebih rendah daripada menerapkan gudang data penuh
- Calon pengguna didefinisikan dengan lebih jelas daripada di gudang data penuh
- Hanya berisi data penting bisnis dan tidak terlalu berantakan.

Data Mart Dependen

Menurut pandangan Bill Inmon, data mart dependen adalah subset logis (tampilan) atau subset fisik (ekstrak) dari gudang data yang lebih besar, diisolasi karena salah satu alasan berikut:

- Penyegaran kebutuhan untuk model atau skema data khusus: misalnya, untuk merestrukturisasi OLAP
- Kinerja: untuk memindahkan data mart ke komputer terpisah untuk efisiensi yang lebih besar atau untuk menghilangkan kebutuhan untuk mengelola beban kerja tersebut di gudang data terpusat.
- Keamanan: untuk memisahkan subset data resmi secara selektif
- Kemanfaatan: untuk melewati tata kelola data dan otorisasi yang diperlukan untuk memasukkan aplikasi baru di Gudang Data Perusahaan
- Proving Ground: untuk mendemonstrasikan viabilitas dan potensi ROI (laba atas investasi) aplikasi sebelum memigrasikannya ke Gudang Data Perusahaan
- Politik: strategi coping untuk TI (Teknologi Informasi) dalam situasi di mana kelompok pengguna memiliki pengaruh yang lebih besar daripada pendanaan atau bukan warga negara yang baik di gudang data terpusat.
- Politik: strategi penanggulangan bagi konsumen data dalam situasi di mana tim data warehouse tidak dapat membuat data warehouse yang dapat digunakan.

Menurut Inmon, pengorbanan yang melekat dengan data mart termasuk skalabilitas terbatas, duplikasi data, ketidakkonsistenan data dengan silo informasi lainnya, dan ketidakmampuan untuk memanfaatkan sumber data perusahaan.

Pengertian alternatif tentang data warehousing adalah dari Ralph Kimball. Dalam pandangannya, data warehouse tidak lebih dari gabungan semua data mart. Tampilan ini membantu mengurangi biaya dan menyediakan pengembangan yang cepat, tetapi dapat

membuat gudang data yang tidak konsisten, terutama di organisasi besar. Oleh karena itu, pendekatan Kimball lebih cocok untuk perusahaan kecil hingga menengah.

4.3 Manajemen Data Master

Pengertian

Manajemen data master (MDM) adalah metode komprehensif yang memungkinkan perusahaan untuk menautkan semua data penting ke satu file, yang disebut file master, yang menyediakan titik acuan umum. Ketika dilakukan dengan benar, manajemen data master merampingkan pembagian data antara personel dan departemen. Selain itu, manajemen data master dapat memfasilitasi komputasi dalam berbagai arsitektur sistem, platform, dan aplikasi.

Pada intinya, Master Data Management (MDM) dapat dilihat sebagai "disiplin untuk peningkatan kualitas khusus" yang ditentukan oleh kebijakan dan prosedur yang diberlakukan oleh organisasi tata kelola data. Tujuan akhirnya adalah untuk menyediakan komunitas pengguna akhir dengan "versi kebenaran yang terpercaya" sebagai dasar keputusan.
Dalam bisnis, *Master Data Management* (MDM) terdiri dari proses, tata kelola, kebijakan, standar, dan alat yang secara konsisten mendefinisikan dan mengelola data penting suatu organisasi untuk memberikan satu titik referensi. MDM menggabungkan aplikasi bisnis, metode manajemen informasi, dan alat manajemen data untuk mengimplementasikan kebijakan, prosedur, infrastruktur yang mendukung penangkapan, integrasi, dan penggunaan bersama data master yang akurat, tepat waktu, konsisten dan lengkap.

Data yang dikuasai mungkin termasuk:
• data referensi - objek bisnis untuk transaksi, dan dimensi untuk analisis
• data analitis - mendukung pengambilan keputusan

Dalam komputasi, alat manajemen data master dapat digunakan untuk mendukung manajemen data master dengan menghapus duplikat, menstandarisasi data (pemeliharaan massal), dan menggabungkan aturan untuk menghilangkan data yang salah dari memasuki sistem untuk membuat sumber otorisasi data master. Data master adalah produk, akun, dan pihak-pihak di mana transaksi bisnis diselesaikan. Akar penyebab masalah berasal dari unit bisnis dan segmentasi lini produk, di mana pelanggan yang sama akan dilayani oleh lini produk yang berbeda, dengan data yang berlebihan dimasukkan tentang pelanggan

(alias pihak dalam peran pelanggan) dan akun untuk memproses transaksi. Redundansi data pesta dan akun diperparah di siklus hidup kantor depan ke belakang, di mana sumber tunggal otoritatif untuk data pesta, akun, dan produk diperlukan, tetapi seringkali sekali lagi dimasukkan atau ditambah secara berlebihan. Manajemen data master memiliki tujuan menyediakan proses untuk mengumpulkan, mengumpulkan, mencocokkan, mengkonsolidasikan, menjamin kualitas, bertahan dan mendistribusikan data tersebut di seluruh organisasi untuk memastikan konsistensi dan kontrol dalam pemeliharaan dan penggunaan aplikasi informasi ini secara berkelanjutan. Tujuan utamanya adalah untuk menyediakan komunitas pengguna akhir dengan "versi tunggal tepercaya dari kebenaran" yang menjadi dasar pengambilan keputusan.

Problematika

Pada tingkat dasar, manajemen data master berupaya untuk memastikan bahwa organisasi tidak menggunakan beberapa versi (yang berpotensi tidak konsisten) dari data master yang sama di berbagai bagian operasinya, yang dapat terjadi di organisasi besar. Contoh tipikal dari manajemen data master yang buruk adalah skenario bank di mana pelanggan telah mengambil hipotek dan bank mulai mengirim permohonan hipotek kepada pelanggan tersebut, mengabaikan fakta bahwa orang tersebut telah memiliki hubungan rekening hipotek dengan bank. . Hal ini terjadi karena informasi pelanggan yang digunakan oleh bagian pemasaran di dalam bank kurang terintegrasi dengan informasi pelanggan yang digunakan oleh bagian layanan pelanggan bank. Dengan demikian, kedua grup tetap tidak menyadari bahwa pelanggan yang ada juga dianggap sebagai prospek penjualan. Proses keterkaitan rekaman digunakan untuk mengaitkan rekaman berbeda yang sesuai dengan entitas yang sama, dalam hal ini orang yang sama.

Masalah lain termasuk (misalnya) masalah dengan kualitas data, klasifikasi dan identifikasi data yang konsisten, dan masalah rekonsiliasi data. Manajemen data master dari sistem data yang berbeda memerlukan transformasi data karena data yang diambil dari sistem data sumber yang berbeda diubah dan dimuat ke dalam hub manajemen data master. Untuk menyinkronkan data master sumber yang berbeda, data master terkelola yang diekstrak dari hub manajemen data master diubah lagi dan dimuat ke sistem data sumber yang berbeda saat data master diperbarui. Seperti gerakan Ekstrak, Transformasi, dan Muat data lainnya, proses ini mahal dan tidak efisien untuk dikembangkan dan dipelihara yang sangat mengurangi laba atas investasi untuk produk manajemen data master.

Salah satu alasan paling umum beberapa perusahaan besar mengalami masalah besar dengan manajemen data master adalah pertumbuhan melalui merger atau akuisisi. Setiap organisasi yang menggabungkan biasanya akan membuat entitas dengan data master duplikat (karena masing-masing kemungkinan memiliki setidaknya satu database master sendiri sebelum merger). Idealnya, administrator database menyelesaikan masalah ini melalui deduplikasi data master sebagai bagian dari penggabungan. Dalam praktiknya, bagaimanapun, rekonsiliasi beberapa sistem data master dapat menimbulkan kesulitan karena ketergantungan aplikasi yang ada pada database master. Akibatnya, lebih sering daripada tidak kedua sistem tidak sepenuhnya bergabung, tetapi tetap terpisah, dengan proses rekonsiliasi khusus yang ditetapkan yang memastikan konsistensi antara data yang disimpan dalam dua sistem. Seiring waktu, bagaimanapun, ketika merger dan akuisisi lebih lanjut terjadi, masalah berlipat ganda, semakin banyak database master muncul, dan proses rekonsiliasi data menjadi sangat kompleks, dan akibatnya tidak dapat dikelola dan tidak dapat diandalkan. Karena tren ini, seseorang dapat menemukan organisasi dengan 10, 15, atau bahkan sebanyak 100 database master yang terpisah dan tidak terintegrasi dengan baik, yang dapat menyebabkan masalah operasional yang serius di bidang kepuasan pelanggan, efisiensi operasional, dukungan keputusan, dan kepatuhan terhadap peraturan.

Solusi

Proses yang biasa terlihat dalam manajemen data master meliputi identifikasi sumber, pengumpulan data, transformasi data, normalisasi, administrasi aturan, deteksi dan koreksi kesalahan, konsolidasi data, penyimpanan data, distribusi data, klasifikasi data, layanan taksonomi, pembuatan master item, pemetaan skema, produk kodifikasi, pengayaan data dan tata kelola data.

Pemilihan entitas yang dipertimbangkan untuk manajemen data master agak bergantung pada sifat organisasi. Dalam kasus umum perusahaan komersial, manajemen data master dapat diterapkan ke entitas seperti pelanggan (integrasi data pelanggan), produk (manajemen informasi produk), karyawan, dan vendor. Proses manajemen data master mengidentifikasi sumber untuk mengumpulkan deskripsi entitas ini. Dalam proses transformasi dan normalisasi, administrator menyesuaikan deskripsi agar sesuai dengan format standar dan domain data, sehingga memungkinkan untuk menghapus

contoh duplikat dari entitas apa pun. Proses tersebut umumnya menghasilkan repositori manajemen data master organisasi, tempat semua permintaan untuk instance entitas tertentu menghasilkan deskripsi yang sama, terlepas dari sumber asal dan tujuan yang meminta.

Alat tersebut meliputi jaringan data, sistem file, gudang data, data mart, penyimpanan data operasional, penambangan data, analisis data, visualisasi data, federasi data, dan virtualisasi data. Salah satu alat terbaru, manajemen data master virtual menggunakan virtualisasi data dan server metadata yang persisten untuk menerapkan hierarki manajemen data master otomatis multi-level

Transmisi Data Master

Ada beberapa cara di mana data master dapat disusun dan didistribusikan ke sistem lain. Ini termasuk:

- Konsolidasi data - Proses pengambilan data master dari berbagai sumber dan diintegrasikan ke dalam satu hub (penyimpanan data operasional) untuk direplikasi ke sistem tujuan lainnya.
- Data federation - Proses menyediakan tampilan virtual tunggal dari data master dari satu atau beberapa sumber ke satu atau lebih sistem tujuan.
- Penyebaran data - Proses penyalinan data master dari satu sistem ke sistem lain, biasanya melalui antarmuka titik-ke-titik dalam sistem lama.

4.4 Dimensi (Gudang Data)

Dimensi adalah struktur yang mengkategorikan fakta dan ukuran untuk memungkinkan pengguna menjawab pertanyaan bisnis. Dimensi yang umum digunakan adalah orang, produk, tempat dan waktu.

Di gudang data, dimensi memberikan informasi pelabelan terstruktur ke ukuran numerik yang tidak berurutan. Dimensi adalah kumpulan data yang terdiri dari elemen data individual yang tidak tumpang tindih. Fungsi utama dimensi ada tiga: menyediakan pemfilteran, pengelompokan, dan pelabelan.

Fungsi ini sering disebut sebagai “slice and dadu”. Mengiris mengacu pada pemfilteran data. Dicing mengacu pada pengelompokan data. Contoh umum data warehouse melibatkan penjualan sebagai ukuran, dengan pelanggan dan produk sebagai dimensi. Dalam setiap penjualan, pelanggan membeli produk. Data dapat diiris dengan menghapus

semua pelanggan kecuali untuk grup yang sedang dipelajari, dan kemudian dipotong dadu dengan mengelompokkan berdasarkan produk.

Elemen data dimensi mirip dengan variabel kategori dalam statistik.
Biasanya dimensi dalam gudang data diatur secara internal ke dalam satu atau beberapa hierarki. "Tanggal" adalah dimensi umum, dengan beberapa kemungkinan hierarki:

- “Hari (dikelompokkan menjadi) Bulan (dikelompokkan menjadi) Tahun”,
- "Hari (dikelompokkan menjadi) Minggu (dikelompokkan menjadi) Tahun"
- “Hari (dikelompokkan ke dalam) Bulan (dikelompokkan ke dalam) Triwulan (dikelompokkan menjadi) Tahun”
- dll.

Jenis-Jenis Dimensi

Confirmed Dimension

Confirmed Dimension adalah sekumpulan atribut data yang telah direferensikan secara fisik dalam beberapa tabel database menggunakan nilai kunci yang sama untuk merujuk ke struktur, atribut, nilai domain, definisi, dan konsep yang sama. Dimensi yang sesuai memotong banyak fakta.

Confirmed Dimension jika keduanya persis sama (termasuk kunci) atau salah satunya adalah bagian sempurna dari yang lain. Yang terpenting, tajuk baris yang dihasilkan dalam dua set jawaban berbeda dari dimensi yang sama harus dapat cocok dengan sempurna.

Confirmed Dimension adalah subkumpulan matematika identik atau ketat dari dimensi yang paling terperinci dan terperinci. Tabel dimensi tidak sesuai jika atribut diberi label berbeda atau berisi nilai yang berbeda. *Confirmed Dimension* datang dalam beberapa rasa yang berbeda. Pada tingkat paling dasar, dimensi yang sesuai memiliki arti yang persis sama dengan setiap tabel fakta yang memungkinkan yang digabungkan. Tabel dimensi tanggal yang terhubung ke fakta penjualan identik dengan dimensi tanggal yang terhubung ke fakta inventaris.

Junk Dimension

Junk Dimension adalah pengelompokan yang mudah dari bendera dan indikator berkardinalitas rendah. Dengan membuat dimensi abstrak, bendera dan indikator ini dihapus dari tabel fakta sambil menempatkannya ke dalam kerangka dimensi yang berguna.

Junk Dimension adalah tabel dimensi yang terdiri dari atribut yang tidak termasuk dalam tabel fakta atau dalam tabel dimensi yang ada. Sifat atribut ini biasanya berupa teks atau berbagai bendera, mis. komentar non-umum atau hanya indikator ya / tidak atau benar / salah. Jenis atribut ini biasanya tetap ada ketika semua dimensi yang jelas dalam proses bisnis telah diidentifikasi dan dengan demikian perancang dihadapkan pada tantangan di mana harus meletakkan atribut ini yang tidak termasuk dalam dimensi lain.

Salah satu solusinya adalah membuat dimensi baru untuk setiap atribut yang tersisa, tetapi karena sifatnya, mungkin perlu membuat sejumlah besar dimensi baru yang menghasilkan tabel fakta dengan kunci asing dalam jumlah yang sangat besar. Perancang juga dapat memutuskan untuk meninggalkan atribut yang tersisa di tabel fakta tetapi ini dapat membuat panjang baris tabel menjadi tidak perlu menjadi besar jika, misalnya, atributnya adalah string teks yang panjang.

Solusi untuk tantangan ini adalah dengan mengidentifikasi semua atribut dan kemudian memasukkannya ke dalam satu atau beberapa *Junk Dimension* . Satu *Junk Dimension* dapat menampung beberapa indikator benar / salah atau ya / tidak yang tidak memiliki korelasi satu sama lain, jadi akan lebih mudah untuk mengubah indikator tersebut menjadi atribut yang lebih menggambarkan. Sebuah contoh akan menjadi indikator apakah sebuah paket telah tiba, alih-alih menunjukkan ini sebagai "ya" atau "tidak", itu akan diubah menjadi "tiba" atau "menunggu" dalam *Junk Dimension* . Perancang dapat memilih untuk membangun tabel dimensi sehingga pada akhirnya memegang semua indikator yang terjadi dengan setiap indikator lainnya sehingga semua kombinasi tercakup. Ini mengatur ukuran tetap untuk tabel itu sendiri yaitu 2x baris, di mana x adalah jumlah indikator. Solusi ini sesuai dalam situasi di mana perancang mengharapkan untuk menemukan banyak kombinasi yang berbeda dan di mana kemungkinan kombinasi terbatas pada tingkat yang dapat diterima. Dalam situasi di mana jumlah indikator besar, sehingga membuat tabel yang sangat besar atau di mana perancang hanya berharap untuk menemukan beberapa kemungkinan kombinasi, akan lebih tepat untuk membangun setiap baris dalam *Junk Dimension* saat kombinasi baru ditemukan. Untuk membatasi ukuran tabel, beberapa *Junk Dimension* mungkin sesuai dalam situasi lain bergantung pada korelasi antara berbagai indikator.

Junk Dimension juga sesuai untuk menempatkan atribut seperti komentar non-umum dari tabel fakta. Atribut seperti itu mungkin terdiri dari data dari kolom komentar opsional saat pelanggan memesan dan akibatnya mungkin akan kosong dalam banyak kasus.

Oleh karena itu, dimensi sampah harus berisi satu baris yang mewakili yang kosong sebagai kunci pengganti yang akan digunakan dalam tabel fakta untuk setiap baris yang dikembalikan dengan kolom komentar kosong.

Degenerate dimension

Degenerate dimension adalah kunci, seperti nomor transaksi, nomor faktur, nomor tiket, atau nomor bill-of-lading, yang tidak memiliki atribut dan karenanya tidak bergabung dengan tabel dimensi yang sebenarnya. Dimensi penurunan sangat umum ketika butir tabel fakta mewakili item transaksi tunggal atau item baris karena dimensi penurunan mewakili pengenal unik dari induk. *Degenerate dimension* sering memainkan peran integral dalam kunci utama tabel fakta.

Role-playing Dimension

Dimensi sering kali didaur ulang untuk beberapa aplikasi dalam database yang sama. Misalnya, dimensi "Tanggal" dapat digunakan untuk "Tanggal Penjualan", serta "Tanggal Pengiriman", atau "Tanggal Sewa". Ini sering disebut sebagai "dimensi bermain peran".

Penggunaan Ketentuan Representasi ISO

Saat mereferensikan data dari registri metadata seperti ISO / IEC 11179, istilah representasi seperti Indikator (nilai boolean benar / salah), Kode (sekumpulan nilai enumerasi yang tidak tumpang tindih) biasanya digunakan sebagai dimensi. Misalnya, menggunakan Model Pertukaran Informasi Nasional (NIEM), nama elemen data adalah PersonGenderCode dan nilai yang disebutkan adalah pria, wanita, dan tidak diketahui.

Tabel Dimensi

Tabel Dimensi

Dalam data warehousing, tabel dimensi adalah salah satu himpunan tabel pendamping ke tabel fakta.

Tabel fakta berisi fakta bisnis (atau ukuran), dan kunci asing yang merujuk ke kunci kandidat (biasanya kunci primer) dalam tabel dimensi.

Berlawanan dengan tabel fakta, tabel dimensi berisi atribut deskriptif (atau bidang) yang biasanya berupa bidang tekstual (atau angka terpisah yang berperilaku seperti teks).

Atribut ini dirancang untuk melayani dua tujuan penting: pembatasan dan / atau pemfilteran kueri, dan pelabelan kumpulan hasil kueri.

Atribut dimensi harus:

- Verbose (label terdiri dari kata-kata lengkap)
- Deskriptif
- Lengkap (tidak memiliki nilai yang hilang)
- Dinilai secara terpisah (hanya memiliki satu nilai per baris tabel dimensi)
- Kualitas terjamin (tidak ada salah eja atau nilai yang tidak mungkin)

Baris tabel dimensi diidentifikasi secara unik oleh satu bidang kunci. Sebaiknya kolom kunci berupa bilangan bulat sederhana karena nilai kunci tidak ada artinya, hanya digunakan untuk menggabungkan kolom antara tabel fakta dan tabel dimensi. Tabel dimensi sering kali menggunakan kunci utama yang juga merupakan kunci pengganti. Kunci pengganti sering dibuat secara otomatis (misalnya "kolom identitas" Sybase atau SQL Server, serial PostgreSQL atau Informix, Oracle SEQUENCE, atau kolom yang ditentukan dengan AUTO_INCRE-MENT di MySQL).

Penggunaan kunci dimensi pengganti membawa beberapa keuntungan, termasuk:

- Performa. Pemrosesan gabungan dibuat jauh lebih efisien dengan menggunakan satu bidang (kunci pengganti)
- Buffering dari praktik manajemen kunci operasional. Ini mencegah situasi di mana baris data yang dipindahkan mungkin muncul kembali ketika kunci aslinya digunakan kembali atau ditetapkan kembali setelah periode dormansi yang lama.
- Pemetaan untuk mengintegrasikan sumber yang berbeda
- Menangani koneksi yang tidak diketahui atau tidak berlaku
- Melacak perubahan dalam nilai atribut dimensi

Meskipun penggunaan kunci pengganti membebani sistem ETL, pemrosesan jalur pipa dapat ditingkatkan, dan alat ETL memiliki pemrosesan kunci pengganti yang lebih baik.

Tujuan tabel dimensi adalah untuk membuat dimensi terstandardisasi dan sesuai yang dapat dibagikan di seluruh lingkungan gudang data perusahaan, dan memungkinkan penggabungan ke beberapa tabel fakta yang mewakili berbagai proses bisnis.

Dimensi yang sesuai penting bagi sifat perusahaan dari sistem DW / BI karena mereka mempromosikan:

- Konsistensi. Setiap tabel fakta difilter secara konsisten, sehingga jawaban kueri diberi label secara konsisten.
- Integrasi. Kueri dapat menelusuri tabel fakta proses yang berbeda secara terpisah untuk setiap tabel fakta individual, lalu menggabungkan hasil pada atribut dimensi umum.
- Mengurangi waktu pengembangan ke pasar. Dimensi umum tersedia tanpa membuatnya kembali.

Seiring waktu, atribut dari baris tertentu dalam tabel dimensi dapat berubah. Misalnya, alamat pengiriman untuk sebuah perusahaan dapat berubah. Kimball menyebut fenomena ini sebagai Dimensi yang Berubah Perlahan. Strategi untuk menghadapi perubahan semacam ini dibagi menjadi tiga kategori:

- Tipe Satu. Cukup timpa nilai lama.
- Tipe Dua. Tambahkan baris baru yang berisi nilai baru, dan bedakan baris tersebut menggunakan teknik pembuatan versi Tuple.
- Tipe Tiga. Tambahkan atribut baru ke baris yang ada.

Pola Umum

Tanggal dan Waktu

Karena banyak tabel fakta di gudang data merupakan rangkaian waktu pengamatan, satu atau lebih dimensi tanggal sering kali diperlukan. Salah satu alasan untuk memiliki dimensi tanggal adalah untuk menempatkan pengetahuan kalender di gudang data alih-alih di-hardcode dalam aplikasi. Meskipun stempel tanggal / waktu SQL sederhana berguna untuk memberikan informasi yang akurat tentang waktu sebuah fakta dicatat, ia tidak dapat memberikan informasi tentang hari libur, periode fiskal, dll. Stempel tanggal / waktu SQL masih dapat berguna untuk disimpan dalam tabel fakta, karena memungkinkan untuk penghitungan yang tepat.

Memiliki tanggal dan waktu dalam dimensi yang sama, dapat dengan mudah menghasilkan dimensi yang sangat besar dengan jutaan baris. Jika diperlukan detail yang tinggi, biasanya ide yang baik untuk membagi tanggal dan waktu menjadi dua atau lebih dimensi yang terpisah. Dimensi waktu dengan butiran detik dalam sehari hanya akan memiliki 86400 baris. Butir yang lebih atau kurang rinci untuk dimensi tanggal / waktu dapat dipilih

tergantung pada kebutuhan. Sebagai contoh, dimensi tanggal dapat akurat untuk tahun, kuartal, bulan atau hari dan dimensi waktu dapat akurat untuk jam, menit atau detik.

Sebagai aturan praktis, dimensi waktu hanya boleh dibuat jika pengelompokan hierarkis diperlukan atau jika ada deskripsi tekstual yang bermakna untuk periode waktu dalam satu hari (mis. "Kesibukan malam" atau "giliran pertama").

Jika baris dalam tabel fakta berasal dari beberapa zona waktu, mungkin berguna untuk menyimpan tanggal dan waktu pada waktu lokal dan waktu standar. Ini dapat dilakukan dengan memiliki dua dimensi untuk setiap dimensi tanggal / waktu yang diperlukan - satu untuk waktu lokal, dan satu untuk waktu standar. Menyimpan tanggal / waktu baik dalam waktu lokal dan standar, akan memungkinkan analisis tentang kapan fakta dibuat dalam pengaturan lokal dan juga dalam pengaturan global. Waktu standar yang dipilih dapat menjadi waktu standar global (mis. UTC), dapat berupa waktu lokal kantor pusat bisnis, atau zona waktu lain yang mungkin digunakan.

4.5 Slowly Changing Dimension

Dimensi dalam manajemen data dan pergudangan data berisi data yang relatif statis tentang entitas seperti lokasi geografis, pelanggan, atau produk. Data yang ditangkap oleh Slowly Changing Dimensions (SCDs) berubah secara perlahan tetapi tidak dapat diprediksi, daripada menurut jadwal reguler.

Beberapa skenario dapat menyebabkan masalah integritas referensial.

Misalnya, database mungkin berisi tabel fakta yang menyimpan catatan penjualan. Tabel fakta ini akan dihubungkan ke dimensi dengan menggunakan kunci asing. Salah satu dimensi berikut mungkin berisi data tentang staf penjualan perusahaan: misalnya, kantor wilayah tempat mereka bekerja. Namun, tenaga penjualan terkadang dipindahkan dari satu kantor regional ke kantor regional lainnya. Untuk tujuan pelaporan penjualan historis, mungkin perlu menyimpan catatan fakta bahwa staf penjualan tertentu telah ditugaskan ke kantor regional tertentu pada tanggal sebelumnya, sedangkan staf penjualan tersebut sekarang ditugaskan ke kantor regional yang berbeda.

Berurusan dengan masalah ini melibatkan metodologi manajemen SCD yang disebut sebagai Tipe 0 sampai 6. SCD Tipe 6 juga kadang-kadang disebut SCD Hibrid.

Tipe 0: Retain Original

Metode Tipe 0 bersifat pasif. Ini mengelola perubahan dimensi dan tidak ada tindakan yang dilakukan. Nilai tetap seperti pada saat rekaman dimensi pertama kali dimasukkan. Dalam keadaan tertentu sejarah dipelihara dengan Tipe 0. Tipe orde tinggi digunakan untuk menjamin pelestarian sejarah sedangkan Tipe 0 memberikan kontrol paling sedikit atau tidak sama sekali. Ini jarang digunakan.

Jenis 1: Overwrite

Metodologi ini menimpa data lama dengan data baru, dan karena itu tidak melacak data historis. Contoh tabel pemasok:

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	CA.

Dalam contoh di atas, Supplier_Code adalah kunci alami dan Supplier_Key adalah kunci pengganti. Secara teknis, kunci pengganti tidak diperlukan, karena baris akan menjadi unik oleh kunci alami (Supplier_Code). Namun, untuk mengoptimalkan kinerja pada gabungan, gunakan integer daripada kunci karakter (kecuali jumlah byte pada kunci karakter kurang dari jumlah byte pada kunci integer).

Jika pemasok memindahkan kantor pusat ke Illinois, catatan tersebut akan ditimpa:

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	IL

Kerugian dari metode Tipe 1 adalah tidak ada riwayat di gudang data. Keuntungannya adalah mudah dirawat.

Jika seseorang telah menghitung tabel agregat yang meringkas fakta berdasarkan negara bagian, itu perlu dihitung ulang saat Supplier_State diubah.

Tipe 2: Add New Row

Metode ini melacak data historis dengan membuat beberapa catatan untuk kunci alami tertentu dalam tabel dimensi dengan kunci pengganti terpisah dan / atau nomor versi yang berbeda. Riwayat tak terbatas disimpan untuk setiap penyisipan.

Misalnya, jika pemasok pindah ke Illinois, nomor versi akan bertambah secara berurutan:

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State	Version
123	ABC	Acme Supply Co	IL	0
124	ABC	Acme Supply Co	IL	1

Metode lainnya adalah dengan menambahkan kolom 'tanggal efektif'.

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State	Start _Date	End_Date
123	ABC	Acme Supply Co	CA	01-Jan-2000	21-Des-2004
124	ABC	Acme Supply Co	IL	22-Des-2004	NULL

End_Date null di baris dua menunjukkan versi tupel saat ini. Dalam beberapa kasus, tanggal tinggi pengganti standar (misalnya 9999-12-31) dapat digunakan sebagai tanggal akhir, sehingga bidang dapat disertakan dalam indeks, dan sehingga substitusi nilai null tidak diperlukan saat membuat kueri.

Transaksi yang mereferensikan kunci pengganti tertentu (Supplier_Key) kemudian secara permanen terikat ke potongan waktu yang ditentukan oleh baris tabel dimensi yang berubah secara perlahan. Tabel agregat yang meringkas fakta menurut negara terus mencerminkan keadaan historis, yaitu keadaan pemasok pada saat transaksi; tidak perlu pembaruan. Untuk mereferensikan entitas melalui kunci alami, perlu untuk menghapus batasan unik yang membuat integritas referensial oleh DBMS tidak mungkin.

Jika ada perubahan retroaktif yang dilakukan pada konten dimensi, atau jika atribut baru ditambahkan ke dimensi (misalnya kolom Sales_Rep) yang memiliki tanggal efektif berbeda dari yang telah ditentukan, hal ini dapat mengakibatkan transaksi yang ada perlu dilakukan. diperbarui untuk mencerminkan situasi baru. Ini bisa menjadi operasi database yang mahal, jadi SCD Tipe 2 bukanlah pilihan yang baik jika model dimensional dapat berubah.

Tipe 3: Add New Attribute

Metode ini melacak perubahan menggunakan kolom terpisah dan mempertahankan riwayat terbatas. Tipe 3 mempertahankan sejarah terbatas karena terbatas pada jumlah kolom yang ditujukan untuk menyimpan data historis. Struktur tabel asli di Tipe 1 dan Tipe 2 sama tetapi Tipe 3 menambahkan kolom tambahan. Dalam contoh berikut, kolom

tambahan telah ditambahkan ke tabel untuk mencatat status asli pemasok - hanya riwayat sebelumnya yang disimpan.

Supplier_ Key	Supplier_ Key	Supplier_Name	Original_ Supplier_State	Effective_ Date	Current_ Supplier_ State
123	ABC	Acme Supply Co	CA	22-Des-2004	IL

Catatan ini berisi kolom untuk keadaan asli dan keadaan saat ini — tidak dapat melacak perubahan jika pemasok pindah untuk kedua kalinya.

Salah satu variasinya adalah membuat bidang Previous_Supplier_State alih-alih Original_ Supplier_State yang hanya akan melacak perubahan historis terbaru.

Tipe 4: Add History Table

Metode Tipe 4 biasanya disebut sebagai menggunakan "tabel riwayat", di mana satu tabel menyimpan data saat ini, dan tabel tambahan digunakan untuk menyimpan catatan dari beberapa atau semua perubahan. Kedua kunci pengganti direferensikan di tabel Fakta untuk meningkatkan kinerja kueri. Untuk contoh di atas nama tabel asli adalah Supplier dan tabel sejarah adalah Supplier_History.

Supplier			
Supplier_Key	Supplier_Key	Supplier_Name	Supplier_State
123	ABC	Acme & Johnson Supply Co	CA

Supplier				
Supplier_Key	Supplier_Key	Supplier_Name	Supplier_State	Create_Date
123	ABC	Acme Supply Co	CA	14-June-2003
124	ABC	Acme & Johnson Supply Co	IL	22-Dec-2004

Metode ini menyerupai bagaimana tabel audit database dan mengubah fungsi teknik pengambilan data.

Tipe 6: Hybrid

Metode Tipe 6 menggabungkan pendekatan tipe 1, 2 dan 3 (1 + 2 + 3 = 6). Salah satu penjelasan yang mungkin tentang asal usul istilah ini adalah bahwa istilah itu diciptakan oleh Ralph Kimball selama percakapan dengan Stephen Pace dari Kalido. Ralph Kimball

menyebut metode ini "Perubahan Tak Terduga dengan Hamparan Versi Tunggal" dalam Perangkat Gudang Data.

Tabel Supplier dimulai dengan satu catatan untuk pemasok contoh kami:

Supplier_ Key	Row_ Key	Supplier_ Code	Supplier_ Name	Current_ State	Historical_ State	Start_ Date	End_Date	Current_ Flag
123	1	ABC	Acme Supply Co	CA	CA	01-Jan-2000	31-Des-2009	Y

Current_State dan Historis_State adalah sama. Atribut Current_Flag opsional menunjukkan bahwa ini adalah rekaman terkini atau terbaru untuk supplier ini.

Ketika Acme Supply Company pindah ke Illinois, kami menambahkan catatan baru, seperti dalam pemrosesan Tipe 2, namun kunci baris disertakan untuk memastikan kami memiliki kunci unik untuk setiap baris:

Supplier_ Key	Row_ Key	Supplier_ Code	Supplier_ Name	Current_ State	Historical_ State	Start_ Date	End_Date	Current_ Flag
123	1	ABC	Acme Supply Co	IL	CA	01-Jan2000	21-Dec-2004	N
123	2	ABC	Acme Supply Co	IL	IL	22-Dec-2004	31-Des-2009	Y

Kami menimpa informasi Current_Flag di catatan pertama (Row_Key = 1) dengan informasi baru, seperti dalam pemrosesan Tipe 1. Kami membuat rekaman baru untuk melacak perubahan, seperti dalam pemrosesan Tipe 2. Dan kami menyimpan riwayat di kolom Status kedua (Historical_State), yang menggabungkan pemrosesan Tipe 3.

Misalnya, jika supplier akan pindah lagi, kami akan menambahkan catatan lain ke dimensi supplier, dan kami akan menimpa konten Current_State column:

Supplier_ Key	Row_ Key	Supplier_ Code	Supplier_ Name	Current_ State	Historical_ State	Start_ Date	End_Date	Current_ Flag
123	1	ABC	Acme Supply Co	NY	CA	01-Jan-2000	21-Dec-2004	N
124	2	ABC	Acme Supply Co	NY	IL	22-Dec-2004	03-Feb-2008	N
125	3	ABC	Acme Supply Co	NY	NY	04-Feb-2008	31-Des-2009	Y

Perhatikan bahwa, untuk record saat ini (Current_Flag = 'Y'), Current_State dan Historical_State selalu sama.

Tipe 2 / tipe 6 Implementasi Fakta

Kunci Pengganti Tipe 2 dengan Atribut Tipe 3

Dalam banyak implementasi SCD Tipe 2 dan Tipe 6, kunci pengganti dari dimensi tersebut dimasukkan ke dalam tabel fakta menggantikan kunci alami ketika data fakta dimuat ke dalam repositori data. Kunci pengganti dipilih untuk rekaman fakta tertentu berdasarkan tanggal efektifnya dan Start_Date dan End_Date dari tabel dimensi. Hal ini memungkinkan data fakta dengan mudah digabungkan ke data dimensi yang benar untuk tanggal efektif yang sesuai.

Berikut adalah tabel Supplier seperti yang kami buat di atas menggunakan metodologi Hybrid Tipe 6:

Supplier_ Key	Supplier_ Code	Supplier_ Name	Current_ State	Historical_ State	Start_ Date	End_Date	Current_ Flag
123	ABC	Acme Supply Co	NY	CA	01-Jan-2000	21-Dec-2004	N
124	ABC	Acme Supply Co	NY	IL	22-Dec-2004	03-Feb-2008	N
123	ABC	Acme Supply Co	NY	NY	04-Feb-2008	31-Des-2009	Y

Setelah tabel Pengiriman berisi Supplier_Key yang benar, tabel tersebut dapat dengan mudah digabungkan ke tabel Supplier menggunakan kunci tersebut. SQL berikut mengambil, untuk setiap rekaman fakta, status pemasok saat ini dan negara tempat pemasok berada pada saat pengiriman:

```
SELECT
  delivery.delivery_cost,
  supplier.supplier_name,
  supplier.historical_state,
  supplier.current_state
FROM delivery
INNER JOIN supplier
  ON delivery.supplier_key = supplier.supplier_key
```

Implementasi tipe 6 murni

Memiliki kunci pengganti Tipe 2 untuk setiap irisan waktu dapat menyebabkan masalah jika dimensinya dapat berubah.

Implementasi Tipe 6 murni tidak menggunakan ini, tetapi menggunakan Kunci Pengganti untuk setiap item data master (mis. Setiap pemasok unik memiliki satu kunci pengganti). Hal ini untuk menghindari perubahan pada data master yang berdampak pada data transaksi yang ada. Ini juga memungkinkan lebih banyak opsi saat menanyakan transaksi.

Berikut adalah tabel Pemasok menggunakan metodologi Tipe 6 murni:

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_ State	Start_ Date	End_Date
456	ABC	Acme Supply Co	CA	01-Jan-2000	21-Dec-2004
456	ABC	Acme Supply Co	IL	22-Dec-2004	03-Feb-2008
456	ABC	Acme Supply Co	NY	04-Feb-2008	31-Des-2009

Contoh berikut menunjukkan bagaimana kueri harus diperluas untuk memastikan satu catatan supplier diambil untuk setiap transaksi.

```
SELECT
  supplier.supplier_code,
  supplier.supplier_state
FROM supplier
INNER JOIN delivery
  ON supplier.supplier_key = delivery.supplier_key
 AND delivery.delivery_date BETWEEN supplier.start_date AND supplier.end_date
```

Catatan fakta dengan tanggal efektif (Delivery_Date) 9 Agustus 2001 akan ditautkan ke Supplier_Code. Dari ABC, dengan Supplier_State dari ‘CA’. Catatan fakta dengan tanggal efektif 11 Oktober 2007 juga akan ditautkan ke Pemasok_ Kode ABC yang sama, tetapi dengan Pemasok_State dari ‘IL’. Meskipun lebih kompleks, ada sejumlah keuntungan dari pendekatan ini, termasuk:

1. Integritas referensial oleh DBMS sekarang mungkin, tetapi orang tidak dapat menggunakan Supplier_Code sebagai kunci asing pada tabel Produk dan menggunakan Supplier_Key sebagai kunci asing setiap produk terikat pada irisan waktu tertentu.

2. Jika ada lebih dari satu tanggal pada kenyataan (mis. Tanggal Pemesanan, Tanggal Pengiriman, Tanggal Pembayaran Faktur) seseorang dapat memilih tanggal mana yang akan digunakan untuk permintaan.
3. Anda bisa melakukan “seperti sekarang”, “seperti pada waktu transaksi” atau “pada titik waktu” kueri dengan mengubah logika filter tanggal.
4. Anda tidak perlu memproses ulang tabel Fakta jika ada perubahan dalam tabel dimensi (misalnya menambahkan bidang tambahan secara retrospektif yang mengubah irisan waktu, atau jika seseorang membuat kesalahan dalam tanggal pada tabel dimensi, seseorang dapat memperbaikinya mudah).
5. Anda dapat memperkenalkan tanggal dua-temporal dalam tabel dimensi.
6. Anda dapat menggabungkan fakta ke beberapa versi tabel dimensi untuk memungkinkan pelaporan informasi yang sama dengan tanggal efektif yang berbeda, dalam kueri yang sama.

Contoh berikut menunjukkan bagaimana tanggal tertentu seperti '2012-01-01 00:00:00' (yang bisa menjadi tanggal waktu saat ini) dapat digunakan

```
SELECT
  supplier.supplier_code,
  supplier.supplier_state
FROM supplier
INNER JOIN delivery
  ON supplier.supplier_key = delivery.supplier_key
 AND ‘2012-01-01 00:00:00’ BETWEEN supplier.start_date AND supplier.end_date
```

Menggunakan Surrogate and Natural Key

Implementasi alternatif adalah menempatkan kunci pengganti dan kunci alami ke dalam tabel fakta. Ini memungkinkan pengguna untuk memilih catatan dimensi yang sesuai berdasarkan:

- tanggal efektif utama pada catatan fakta (di atas),
- informasi terbaru atau terkini,
- tanggal lain yang terkait dengan catatan fakta.

Metode ini memungkinkan tautan yang lebih fleksibel ke dimensi, bahkan jika seseorang telah menggunakan pendekatan Tipe 2 daripada Tipe 6.
Ini adalah tabel Pemasok karena kami mungkin telah membuatnya menggunakan metodologi Tipe 2:

Supplier_Key	Supplier_ Code	Supplier_Name	Supplier_ State	Start_ Date	End_Date	Current_ Flag
123	ABC	Acme Supply Co	CA	01-Jan-2000	21-Dec-2004	N
124	ABC	Acme Supply Co	IL	22-Dec-2004	03-Feb-2008	N
125	ABC	Acme Supply Co	NY	04-Feb-2008	31-Des-2009	Y

SQL berikut mengambil Supplier_Name dan Supplier_State terbaru untuk setiap rekaman fakta:

```
SELECT
  delivery.delivery_cost,
  supplier.supplier_name,
  supplier.supplier_state
FROM delivery ‘Y’
INNER JOIN supplier
  ON delivery.supplier_code = supplier.supplier_code
WHERE supplier.current_flag = ‘Y’
```

Jika ada beberapa tanggal pada catatan fakta, fakta dapat digabungkan ke dimensi menggunakan tanggal lain dan bukan tanggal efektif utama. Misalnya, tabel Pengiriman mungkin memiliki tanggal efektif utama Delivery_Date, tetapi mungkin juga memiliki Order_Date yang terkait dengan setiap catatan.

SQL berikut mengambil Supplier_Name dan Supplier_State yang benar untuk setiap catatan fakta berdasarkan Order_Date:

```
SELECT
  delivery.delivery_cost,
  supplier.supplier_name,
  supplier.supplier_state
FROM delivery
INNER JOIN supplier
  ON delivery.supplier_code = supplier.supplier_code
 AND delivery.order_date BETWEEN supplier.start_date AND supplier.end_date
```

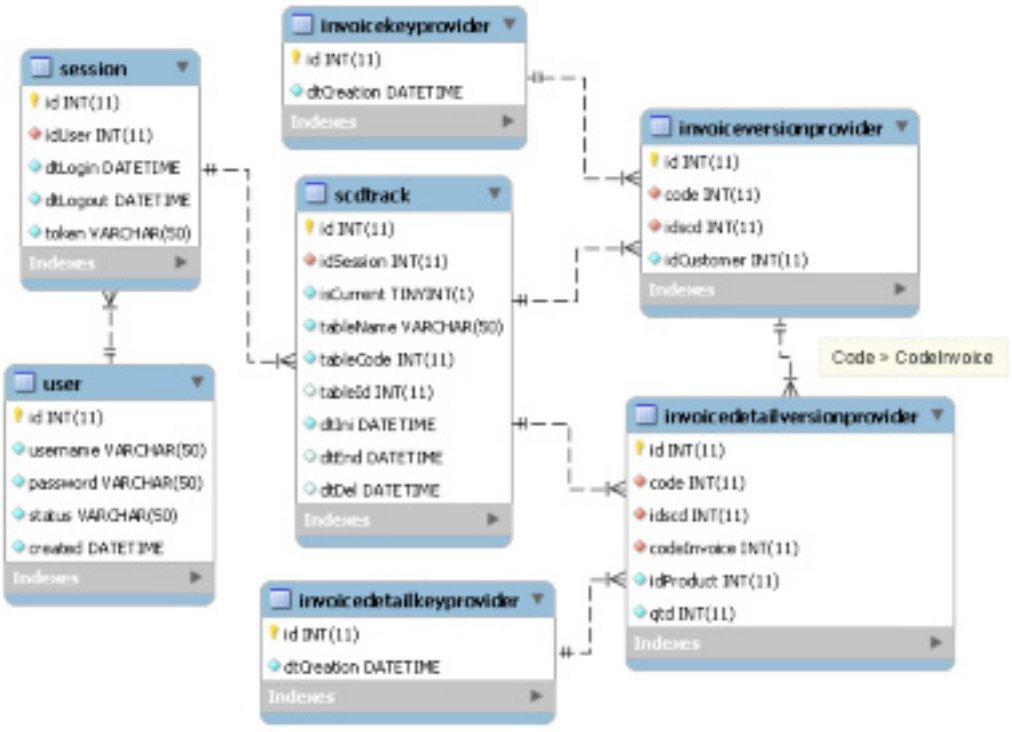
Catatan penting:

- Integritas referensial oleh DBMS tidak dimungkinkan karena tidak ada yang unik untuk membuat hubungan.

- Jika hubungan dibuat dengan pengganti untuk menyelesaikan masalah di atas maka yang satu diakhiri dengan entitas yang terikat pada irisan waktu tertentu.
- Jika kueri gabungan tidak ditulis dengan benar, ini mungkin mengembalikan baris duplikat dan / atau memberikan jawaban yang salah.
- Perbandingan tanggal mungkin tidak bekerja dengan baik.
- Beberapa alat Business Intelligence tidak menangani pembuatan gabungan yang rumit dengan baik.
- Proses ETL yang diperlukan untuk membuat tabel dimensi perlu dirancang dengan cermat untuk memastikan bahwa tidak ada tumpang tindih dalam periode waktu untuk setiap item data referensi yang berbeda.
- Banyak masalah di atas dapat diselesaikan dengan menggunakan diagram campuran model scd di bawah ini.

Gabungan Type

Jenis SCD yang berbeda dapat diterapkan ke kolom tabel yang berbeda. Misalnya, kita dapat menerapkan Tipe 1 ke kolom Supplier_Name dan Tipe 2 ke kolom Supplier_State pada tabel yang sama.



Model scd

4.6 Pemodelan Data Vault

Pemodelan data vault adalah metode pemodelan basis data yang dirancang untuk menyediakan penyimpanan historis jangka panjang dari data yang berasal dari berbagai sistem operasional. Ini juga merupakan metode untuk melihat data historis yang berhubungan dengan masalah seperti audit, penelusuran data, kecepatan pemuatan, dan ketahanan untuk berubah serta menekankan perlunya melacak dari mana semua data dalam basis data berasal. Ini berarti bahwa setiap baris dalam ruang data harus disertai oleh atribut sumber catatan dan memuat tanggal, memungkinkan auditor untuk melacak nilai kembali ke sumber.

Pemodelan data vault tidak membedakan antara data yang baik dan buruk ("buruk" artinya tidak sesuai dengan aturan bisnis). Ini dirangkum dalam pernyataan bahwa penyimpanan data menyimpan "versi tunggal fakta" (juga dinyatakan oleh Dan Linstedt sebagai "semua data, sepanjang waktu") yang bertentangan dengan praktik dalam metode penyimpanan data gudang lainnya " versi tunggal dari kebenaran "di mana data yang tidak sesuai dengan definisi dihapus atau" dibersihkan ".

Metode pemodelan dirancang agar tahan terhadap perubahan dalam lingkungan bisnis tempat data disimpan, dengan secara eksplisit memisahkan informasi struktural dari atribut deskriptif. Data vault dirancang untuk memungkinkan pemuatan paralel sebanyak mungkin, sehingga implementasi yang sangat besar dapat keluar tanpa perlu desain ulang besar.

Sejarah dan Filsafat

Dalam pemodelan data warehouse ada dua opsi bersaing terkenal untuk pemodelan lapisan tempat data disimpan. Baik Anda memodelkan menurut Ralph Kimball, dengan dimensi yang sesuai dan bus data perusahaan, atau Anda memodelkan sesuai dengan Bill Inmon dengan database dinormalisasi. Kedua teknik memiliki masalah ketika berhadapan dengan perubahan dalam sistem yang memberi makan gudang data. Untuk dimensi yang sesuai Anda juga harus membersihkan data (untuk menyesuaikannya) dan ini tidak diinginkan dalam sejumlah kasus karena hal ini pasti akan kehilangan informasi. Data vault dirancang untuk menghindari atau meminimalkan dampak dari masalah-masalah tersebut, dengan memindahkannya ke area gudang data yang berada di luar area penyimpanan historis (pembersihan dilakukan di data mart) dan dengan memisahkan item struktural (kunci bisnis dan asosiasi antara kunci bisnis) dari atribut deskriptif. Dan

Linstedt, pencipta metode ini, menjelaskan basis data yang dihasilkan sebagai berikut: Data Vault Model adalah perincian berorientasi, pelacakan historis, dan serangkaian tabel yang dinormalisasi secara unik yang mendukung satu atau lebih area fungsional bisnis. Ini adalah pendekatan hibrida yang mencakup breed terbaik antara bentuk normal ke-3 (3NF) dan skema bintang. Desainnya fleksibel, dapat diukur, konsisten, dan dapat disesuaikan dengan kebutuhan filosofi Data vault perusahaan adalah bahwa semua data adalah data yang relevan, bahkan jika tidak sejalan dengan definisi dan aturan bisnis yang ditetapkan. Jika data tidak sesuai dengan definisi dan aturan ini maka itu merupakan masalah bagi bisnis, bukan gudang data. Penentuan data yang "salah" adalah interpretasi data yang berasal dari sudut pandang tertentu yang mungkin tidak berlaku untuk semua orang, atau pada setiap titik waktu. Oleh karena itu vault data harus menangkap semua data dan hanya ketika melaporkan atau mengekstraksi data dari vault data adalah data yang ditafsirkan.

Masalah lain yang menjadi respons data loncatan adalah semakin banyak kebutuhan akan auditabilitas dan keterlacakan semua data di gudang data. Karena persyaratan Sarbanes Oxley di AS dan tindakan serupa di Eropa, ini adalah topik yang relevan untuk banyak implementasi intelijen bisnis, oleh karena itu fokus dari implementasi penyimpanan data adalah penelusuran lengkap dan kemampuan audit semua informasi.

Data Vault 2.0 adalah spesifikasi baru, ini merupakan standar terbuka. Spesifikasi baru berisi komponen yang menentukan praktik terbaik implementasi, metodologi (SEI / CMMI, Six Sigma, SDLC, dll.), Arsitektur, dan model. Data Vault 2.0 memiliki fokus termasuk komponen-komponen baru seperti Big Data, NoSQL - dan juga berfokus pada kinerja model yang ada. Spesifikasi lama (sebagian besar didokumentasikan di sini) sangat fokus pada pemodelan data vault. Itu didokumentasikan dalam buku: Membangun Gudang Data yang Dapat diskala dengan Data Vault 2.0. Penting untuk mengembangkan spesifikasi untuk memasukkan komponen-komponen baru, bersama dengan praktik terbaik untuk menjaga sistem EDW dan BI tetap terkini dengan kebutuhan dan keinginan bisnis saat ini.

Sejarah

Pemodelan data vault awalnya dikandung oleh Dan Linstedt pada tahun 1990 dan dirilis pada tahun 2000 sebagai metode pemodelan domain publik. Dalam serangkaian lima artikel tentang Data Administration Newsletter, aturan dasar metode Data Vault diperluas dan dijelaskan. Ini berisi gambaran umum, gambaran umum komponen, diskusi tentang

tanggal dan penggabungan akhir, tabel tautan, dan artikel tentang praktik pemuatan. Nama alternatif (dan jarang digunakan) untuk metode ini adalah “Arsitektur Pemodelan Integrasi Dasar yang Umum.” Data Vault 2.0 telah hadir pada 2013 dan membawa ke tabel Big Data, NoSQL, integrasi tanpa-terstruktur semi-terstruktur, bersama dengan metodologi, arsitektur, dan penerapan praktik terbaik.

Interpretasi Alternatif

Menurut Dan Linstedt, Model Data diilhami oleh (atau terpola) pandangan sederhana tentang neuron, dendrit, dan sinapsis - di mana neuron dikaitkan dengan Hub dan Satelit Hub, Tautan adalah dendrit (vektor informasi), dan Tautan lainnya adalah sinapsis (vektor dalam arah yang berlawanan). Dengan menggunakan serangkaian algoritma penambangan data, tautan dapat dinilai dengan peringkat kepercayaan dan kekuatan. Mereka dapat dibuat dan dijatuhkan dengan cepat sesuai dengan belajar tentang hubungan yang saat ini tidak ada. Model dapat secara otomatis berubah, disesuaikan, dan disesuaikan saat digunakan dan diberi makan struktur baru. Pandangan lain adalah bahwa model data vault memberikan ontologi Perusahaan dalam arti menggambarkan istilah dalam domain perusahaan (Hub) dan hubungan di antara mereka (Tautan), menambahkan atribut deskriptif (Satelit) jika diperlukan. Cara lain untuk memikirkan model data vault adalah sebagai model grafik. Model data vault sebenarnya menyediakan model “berbasis grafik” dengan hub dan hubungan dalam dunia basis data relasional. Dengan cara ini, pengembang dapat menggunakan SQL untuk mendapatkan hubungan berbasis grafik dengan respons sub-detik.

Pengertian Dasar

Data vault berupaya memecahkan masalah berurusan dengan perubahan lingkungan dengan memisahkan kunci bisnis (yang tidak bermutasi sesering mungkin, karena mereka secara unik mengidentifikasi entitas bisnis) dan hubungan antara kunci-kunci bisnis tersebut, dari atribut deskriptif kunci-kunci tersebut. . Kunci bisnis dan asosiasinya adalah atribut struktural, membentuk kerangka model data. Metode data vault memiliki sebagai salah satu aksioma utama bahwa kunci bisnis nyata hanya berubah ketika bisnis berubah dan karena itu merupakan elemen yang paling stabil dari mana untuk memperoleh struktur database historis. Jika Anda menggunakan kunci ini sebagai tulang punggung gudang data, Anda dapat mengatur sisa data di sekitarnya. Ini berarti bahwa memilih kunci yang tepat untuk hub adalah sangat penting untuk stabilitas model Anda. Kunci disimpan dalam tabel dengan beberapa kendala pada struktur. Tabel-kunci ini disebut hub.

Hub

Hub berisi daftar kunci bisnis unik dengan kecenderungan berubah yang rendah. Hub juga berisi kunci pengganti untuk setiap item Hub dan metadata yang menjelaskan asal dari kunci bisnis. Atribut deskriptif untuk informasi pada Hub (seperti deskripsi untuk kunci, mungkin dalam beberapa bahasa) disimpan dalam struktur yang disebut tabel satelit yang akan dibahas di bawah ini. Hub berisi setidaknya bidang-bidang berikut:

• kunci pengganti, digunakan untuk menghubungkan struktur lain ke tabel ini.

• kunci bisnis, driver untuk hub ini. Kunci bisnis dapat terdiri dari beberapa bidang.

• sumber rekaman, yang dapat digunakan untuk melihat sistem apa yang memuat setiap kunci bisnis terlebih dahulu. opsional, Anda juga dapat memiliki bidang metadata dengan informasi tentang pembaruan manual (pengguna / waktu) dan tanggal ekstraksi.

Hub tidak diperbolehkan mengandung beberapa kunci bisnis, kecuali ketika dua sistem memberikan kunci bisnis yang sama tetapi dengan tabrakan yang memiliki arti berbeda. Hub biasanya harus memiliki setidaknya satu satelit.

Contoh Hub

Ini adalah contoh untuk tabel hub yang berisi mobil, yang disebut “Mobil” (H_CAR). Kunci penggeraknya adalah nomor identifikasi kendaraan.

Fieldname	Deskripsi	Wajib?	Catatan
H_CAR_ID	ID urutan dan kunci pengganti untuk hub	Tidak	Direkomendasikan tapi boleh memilih
VEHICLE_ID_NR	Kunci bisnis yang menggerakkan hub ini. Bisa lebih dari satu bidang untuk kunci bisnis komposit	Ya	
H_RSRC	Sumber data kunci ini saat pertama kali dimuat	Ya	
LOAD_AUDIT_ID	ID ke dalam tabel dengan informasi audit, seperti waktu buka, durasi pemuatan, jumlah baris, dll	Tidak	

Tautan / Link

Asosiasi atau transaksi antara kunci bisnis (terkait misalnya hub untuk pelanggan dan produk satu sama lain melalui transaksi pembelian) dimodelkan menggunakan tabel tautan. Tabel-tabel ini pada dasarnya banyak-ke-banyak tabel bergabung, dengan beberapa metadata. Tautan dapat menautkan ke tautan lain, untuk menangani perubahan granularity (misalnya, menambahkan kunci baru ke tabel database akan mengubah butir dari tabel database). Misalnya, jika Anda memiliki hubungan antara pelanggan dan alamat, Anda dapat menambahkan referensi ke tautan antara hub untuk produk dan perusahaan transportasi. Ini bisa berupa tautan yang disebut “Pengiriman”. Merujuk tautan di tautan

lain dianggap praktik yang buruk, karena memperkenalkan ketergantungan antar tautan yang membuat pemuatan paralel menjadi lebih sulit. Karena tautan ke tautan lain sama dengan tautan baru dengan hub dari tautan lain, dalam hal ini membuat tautan tanpa merujuk tautan lain adalah solusi yang lebih disukai. Tautan terkadang menghubungkan hub ke informasi yang tidak dengan sendirinya cukup untuk membangun hub. Ini terjadi ketika salah satu kunci bisnis yang terkait dengan tautan tersebut bukan kunci bisnis yang nyata. Sebagai contoh, ambil formulir pemesanan dengan “nomor pesanan” sebagai kunci, dan baris pesanan yang dikunci dengan nomor semi-acak untuk membuatnya unik. Katakanlah, “nomor unik”.

Kunci yang terakhir bukanlah kunci bisnis yang nyata, jadi bukan hub. Namun, kami perlu menggunakannya untuk menjamin rincian yang benar untuk tautan tersebut. Dalam hal ini, kami tidak menggunakan hub dengan kunci pengganti, tetapi menambahkan “nomor unik” kunci bisnis itu sendiri ke tautan. Ini dilakukan hanya ketika tidak ada kemungkinan pernah menggunakan kunci bisnis untuk tautan lain atau sebagai kunci untuk atribut dalam satelit. Konstruk ini disebut ‘tautan pasak-berkaki’ oleh Dan Linstedt di forumnya (yang sekarang sudah tidak ada).

Tautan berisi kunci pengganti untuk hub yang ditautkan, kunci pengganti mereka sendiri untuk tautan dan metadata yang menjelaskan asal dari asosiasi. Atribut deskriptif untuk informasi mengenai asosiasi (seperti waktu, harga atau jumlah) disimpan dalam struktur yang disebut tabel satelit yang dibahas di bawah ini.

Contoh Tautan

Ini adalah contoh untuk tabel tautan antara dua hub untuk mobil (H_CAR) dan orang (H_PERSON). Tautan ini disebut “Driver” (L_DRIVER).

Fieldname	Deskripsi	Wajib?	Catatan
L_DRIVER_ID	ID urutan dan kunci pengganti untuk Link	Tidak	Direkomendasikan tapi boleh memilih
H_CAR_ID	kunci pengganti untuk hub mobil, jangkar pertama dari tautan	Ya	
H_PERSON_ID	Sumber data kunci ini saat pertama kali dimuat	Ya	
L_RSR	Sumber data pengaitan ini saat pertama kali dimuat	Ya	
LOAD_AUDIT_ID	ID ke dalam tabel dengan informasi audit, seperti waktu buka, durasi pemuatan, jumlah baris, dll.	Tidak	

Satellite

Hub dan tautan membentuk struktur model, tetapi tidak memiliki atribut temporal dan tidak memiliki atribut deskriptif. Ini disimpan dalam tabel terpisah yang disebut satelit. Ini terdiri dari metadata yang menautkannya ke hub atau tautan induknya, metadata menjelaskan asal asosiasi dan atribut, serta garis waktu dengan tanggal mulai dan berakhir untuk atribut. Jika hub dan tautan menyediakan struktur model, satelit menyediakan "daging" model, konteks untuk proses bisnis yang ditangkap dalam hub dan tautan. Atribut-atribut ini disimpan baik berkenaan dengan detail masalah serta garis waktu dan dapat berkisar dari cukup kompleks (semua bidang yang menggambarkan profil lengkap klien) hingga cukup sederhana (satelit pada tautan dengan hanya indikator yang valid dan garis waktu). Biasanya atribut dikelompokkan dalam satelit berdasarkan sistem sumber. Namun, atribut deskriptif seperti ukuran, biaya, kecepatan, jumlah atau warna dapat berubah pada tingkat yang berbeda, sehingga Anda juga dapat membagi atribut ini dalam satelit yang berbeda berdasarkan tingkat perubahannya. Semua tabel berisi metadata, minimal menggambarkan setidaknya sistem sumber dan tanggal entri ini menjadi valid, memberikan pandangan historis lengkap data saat memasuki gudang data.

Contoh Satelit

Ini adalah contoh untuk satelit pada tautan driver antara hub untuk mobil dan orang, yang disebut "Asuransi Pengemudi" (S_DRIVER_INSURANCE). Satelit ini berisi atribut yang spesifik untuk asuransi hubungan antara mobil dan orang yang mengendarainya, misalnya indikator apakah ini adalah pengemudi utama, nama perusahaan asuransi untuk mobil dan orang ini (bisa juga terpisah hub) dan ringkasan jumlah kecelakaan yang melibatkan kombinasi kendaraan dan pengemudi ini. Juga termasuk referensi ke tabel pencarian-atau referensi yang disebut R_RISK_CATEGORY yang berisi kode untuk kategori risiko di mana hubungan ini dianggap jatuh.

Fieldname	Deskripsi	Wajib?	Catatan
S_DRIVER_INSURANCE_ID	ID urutan dan kunci pengganti untuk satelit di link	Tidak	Direkomendasikan tapi boleh memilih
L_DRIVER_ID	(pengganti) kunci utama untuk tautan pengemudi, induk dari satelit	Ya	
S_SEQ_NR	Urutan atau nomor urut, untuk menerapkan keunikan jika ada beberapa satelit yang valid untuk satu kunci induk	Ya	Ini bisa terjadi jika, misalnya, Anda memiliki KURSUS hub dan nama kursusnya adalah atribut tetapi dalam beberapa bahasa berbeda. Tanggal Muat (tanggal mulai) untuk validitas kombinasi ini

S_LDTS	Muat Tanggal Akhir (tanggal akhir) untuk validitas kombinasi nilai atribut ini untuk kunci induk L_DRIVER_ID	Tidak(**)	
S _LEDTS	Muat Tanggal Akhir (tanggal akhir) untuk validitas kombinasi nilai atribut ini untuk kunci induk L_DRIVER_ID	YA	
I N D _ P R I M A R Y _ DRIVER	Indikator apakah pengemudi adalah pengemudi utama mobil ini	Tidak (*)	
I N S U R A N C E _ COMPANy	Nama perusahaan asuransi untuk kendaraan ini dan pengemudi ini	Tidak (*)	
NR_OF_ACCIDENTS	Jumlah kecelakaan pengemudi di kendaraan ini	Tidak (*)	
R_RISK_CATEGORY_ CD	Kategori risiko untuk pengemudi. Ini adalah referensi ke R_RISK_ KATEGORI	Tidak (*)	
S_RSRC	Sumber catatan informasi di satelit ini saat pertama kali dimuat	Ya	
LOAD_AUDIT_ID	ID ke dalam tabel dengan informasi audit, seperti waktu buka, durasi pemuatan, jumlah baris, dll.	Tidak	

(*) setidaknya satu atribut adalah wajib. (**) nomor urut menjadi wajib jika diperlukan untuk menegakkan keunikan untuk beberapa satelit yang valid pada hub atau tautan yang sama.

Tabel Referensi

Tabel referensi adalah bagian normal dari model brankas data yang sehat. Mereka ada untuk mencegah penyimpanan berlebihan data referensi sederhana yang banyak direferensikan. Secara lebih formal, Dan Linstedt mendefinisikan data referensi sebagai berikut: Setiap informasi yang dianggap perlu untuk menyelesaikan deskripsi dari kode, atau untuk menerjemahkan kunci ke (sic) secara konsisten. Banyak dari bidang ini bersifat "deskriptif" dan menggambarkan keadaan spesifik dari informasi lain yang lebih penting. Dengan demikian, data referensi tinggal di tabel terpisah dari tabel Vault data mentah. Tabel referensi dirujuk dari Satelit, tetapi tidak pernah diikat dengan kunci asing fisik. Tidak ada struktur yang ditentukan untuk tabel referensi: gunakan yang berfungsi paling baik dalam kasus spesifik Anda, mulai dari tabel pencarian sederhana hingga brankas data kecil atau bahkan bintang. Mereka bisa bersifat historis atau tidak memiliki sejarah,

tetapi Anda disarankan untuk tetap menggunakan kunci alami dan tidak membuat kunci pengganti dalam kasus itu. Biasanya, brankas data memiliki banyak tabel referensi, sama seperti Gudang Data lainnya.

Contoh Referensi

Ini adalah contoh tabel referensi dengan kategori risiko untuk pengemudi kendaraan. Itu bisa direferensikan dari satelit apa pun di ruang data. Untuk saat ini kami mereferensikannya dari satelit S_DRIVER_INSURANCE.
Tabel referensi adalah R_RISK_CATEGORY.

Fieldname	Deskripsi	Wajib?
R_RISK_CATEGORY_CD	Kode untuk kategori risiko	Ya
RISK_CATEGORY_DESC	Penjelasan tentang kategori risiko	Tidak (*)

(*) setidaknya satu atribut adalah wajib.

Latihan Memasukkan/Memuat Data

ETL untuk memperbarui model data vault cukup mudah. Pertama, Anda harus memuat semua hub, membuat ID pengganti untuk setiap kunci bisnis baru. Setelah melakukan itu, Anda sekarang dapat menyelesaikan semua kunci bisnis untuk mengganti ID jika Anda meminta hub. Langkah kedua adalah menyelesaikan tautan antara hub dan membuat ID pengganti untuk setiap asosiasi baru. Pada saat yang sama, Anda juga dapat membuat semua satelit yang dilampirkan ke hub, karena Anda dapat menyelesaikan kunci ke ID pengganti. Setelah Anda membuat semua tautan baru dengan kunci pengganti, Anda dapat menambahkan satelit ke semua tautan. Karena hub tidak bergabung satu sama lain kecuali melalui tautan, Anda dapat memuat semua hub secara paralel. Karena tautan tidak dilampirkan secara langsung satu sama lain, Anda dapat memuat semua tautan secara paralel. Karena satelit hanya dapat dilampirkan ke hub dan tautan, Anda juga dapat memuatnya secara paralel.

ETL cukup mudah dan cocok untuk otomatisasi atau templating yang mudah. Masalah hanya terjadi pada tautan yang terkait dengan tautan lain, karena menyelesaikan kunci bisnis di tautan hanya mengarah ke tautan lain yang harus dipecahkan juga. Karena kesetaraan situasi ini dengan tautan ke banyak hub, kesulitan ini dapat dihindari dengan merombak kasus-kasus seperti itu dan ini sebenarnya praktik yang disarankan. Data tidak pernah dihapus dari brankas data, kecuali jika Anda memiliki kesalahan teknis saat memuat data.

Data Vault dan Pemodelan Dimensi

Layer pada model data vault biasanya digunakan untuk menyimpan data. Ini tidak dioptimalkan untuk kinerja kueri, juga tidak mudah untuk melakukan kueri dengan alat kueri terkenal seperti Cognos, SAP Business Objects, Pentaho et al. Karena alat komputasi pengguna akhir ini mengharapkan atau lebih suka datanya terdapat dalam model dimensi, biasanya diperlukan konversi.

Untuk tujuan ini, hub dan satelit terkait pada hub tersebut dapat dianggap sebagai dimensi dan tautan serta satelit terkait pada tautan tersebut dapat dilihat sebagai tabel fakta dalam model dimensi. Ini memungkinkan Anda dengan cepat membuat prototipe model dimensi dari model brankas data menggunakan tampilan. Untuk alasan kinerja, model dimensi biasanya akan diimplementasikan dalam tabel relasional, setelah disetujui.

Perhatikan bahwa meskipun memindahkan data dari model data vault ke model dimensi (bersih) relatif mudah, sebaliknya tidak semudah itu.

Metodologi Gudang Data

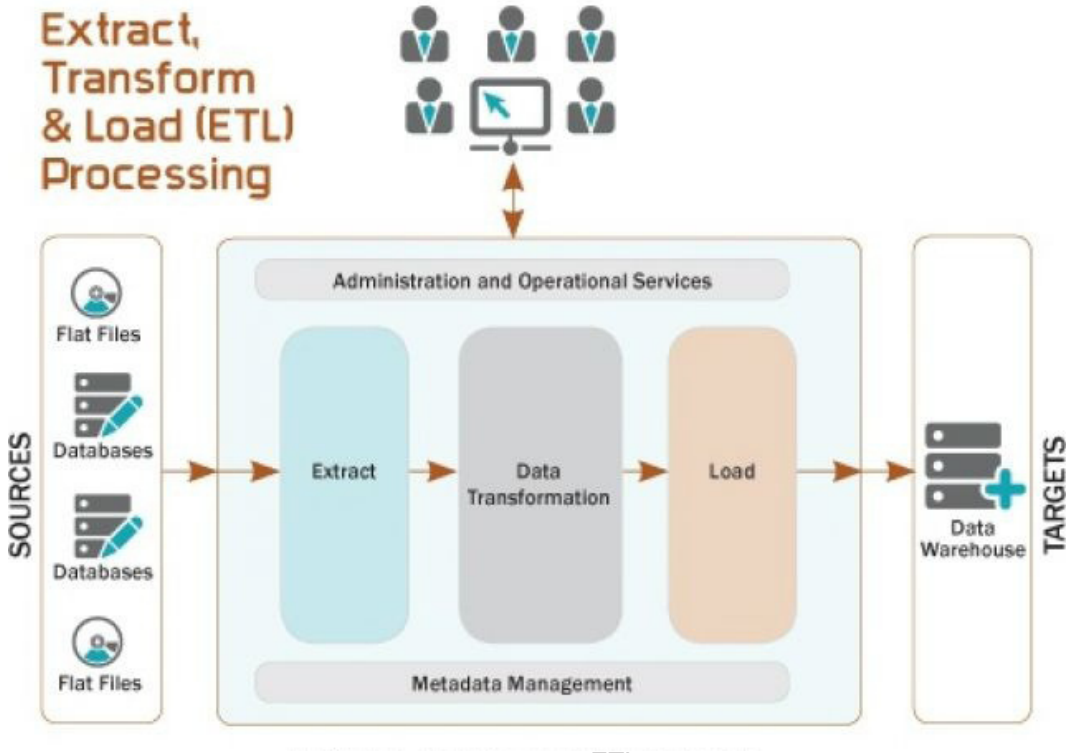
Metodologi data vault didasarkan pada praktik terbaik SEI / CMMI Level 5. Ini mencakup beberapa komponen CMMI Level 5, dan menggabungkannya dengan praktik terbaik dari Six Sigma, TQM, dan SDLC. Terutama, ini difokuskan pada metodologi tangkas Scott Ambler untuk membangun dan menerapkan. Proyek data vault memiliki siklus rilis pendek yang dikendalikan ruang lingkup dan harus terdiri dari rilis produksi setiap 2 hingga 3 minggu.

Tim yang menggunakan metodologi data vault akan secara otomatis mengadopsi proyek yang berulang, konsisten, dan terukur yang diharapkan pada CMMI Level 5. Data yang mengalir melalui sistem brankas data EDW akan mulai mengikuti siklus hidup TQM (manajemen kualitas total) yang sudah lama

4.7 Sistem Extract, Transform, Load (ELT)

Dalam komputasi, Extract, Transform, Load (ETL) mengacu pada proses dalam penggunaan basis data dan khususnya dalam pergudangan data. Ekstraksi data adalah ketika data diekstraksi dari sumber data homogen atau heterogen; transformasi data di mana data ditransformasikan untuk menyimpan dalam format atau struktur yang tepat untuk keperluan query dan analisis; pemuatan data tempat data dimuat ke basis data target akhir, lebih

khusus lagi, penyimpanan data operasional, data mart, atau data warehouse. Karena ekstraksi data membutuhkan waktu, itu biasa untuk mengeksekusi tiga fase secara paralel. Sementara data sedang diekstraksi, proses transformasi lain dijalankan. Ini memproses data yang sudah diterima dan mempersiapkannya untuk memuat. Segera setelah ada beberapa data yang siap untuk dimuat ke target, pemuatan data dimulai tanpa menunggu selesainya fase sebelumnya. Sistem ETL umumnya mengintegrasikan data dari berbagai aplikasi (sistem), biasanya dikembangkan dan didukung oleh vendor yang berbeda atau di-host pada perangkat keras komputer yang terpisah. Sistem yang berbeda yang berisi data asli sering dikelola dan dioperasikan oleh karyawan yang berbeda. Misalnya, sistem akuntansi biaya dapat menggabungkan data dari daftar gaji, penjualan, dan pembelian.



Fungsi Dasar *ETL*

Extraction

Bagian pertama dari proses ETL melibatkan penggalian data dari sistem sumber. Dalam banyak kasus ini merupakan aspek yang paling penting dari ETL, karena mengekstraksi data dengan benar menetapkan tahapan untuk keberhasilan proses selanjutnya. Sebagian besar proyek pergudangan data menggabungkan data dari sistem sumber yang berbeda. Setiap sistem yang terpisah juga dapat menggunakan organisasi dan / atau format data yang berbeda. Format sumber data umum termasuk basis data relasional, XML dan

file datar, tetapi juga dapat mencakup struktur basis data non-relasional seperti Sistem Manajemen Informasi (IMS) atau struktur data lainnya seperti Metode Akses Penyimpanan Virtual (VSAM) atau Metode Akses Berurutan Terindeks (Indexed Sequential Access Method) ISAM), atau bahkan format yang diambil dari sumber luar dengan cara seperti spidering web atau memo layar. Streaming sumber data yang diekstraksi dan memuat on-the-fly ke database tujuan adalah cara lain untuk melakukan ETL ketika tidak ada penyimpanan data antara yang diperlukan.

Secara umum, fase ekstraksi bertujuan untuk mengubah data menjadi format tunggal yang sesuai untuk pemrosesan transformasi. Bagian intrinsik dari ekstraksi melibatkan validasi data untuk mengkonfirmasi apakah data yang diambil dari sumber memiliki nilai yang benar / diharapkan dalam domain yang diberikan (seperti pola / default atau daftar nilai). Jika data gagal aturan validasi itu ditolak seluruhnya atau sebagian. Data yang ditolak idealnya dilaporkan kembali ke sistem sumber untuk analisis lebih lanjut untuk mengidentifikasi dan memperbaiki catatan yang salah. Dalam beberapa kasus, proses ekstraksi itu sendiri mungkin harus melakukan aturan validasi data untuk menerima data dan mengalir ke fase berikutnya.

Transformation

Proses transformasi data merupakan proses mengubah data dari format operasional menjadi format data warehouse. Proses transformasi berupa tugas-tugas seperti mengkonversi tipe data, melakukan beberapa perhitungan, penyaringan data yang tidak relevan, dan meringkasnya. Proses transformasi dibutuhkan untuk memenuhi kebutuhan bisnis suatu perusahaan. Langkah-langkah dalam transformasi data adalah sebagai berikut :

- Memetakan data input dari skema data aslinya ke skema data warehouse.
- Melakukan konversi tipe data atau format data.
- Pembersihan serta pembuangan duplikasi dan kesalahan data.
- Penghitungan nilai-nilai derivat atau mula-mula.
- Penghitungan nilai-nilai agregat atau rangkuman.
- Pemerikasaan integritas referensi data.
- Pengisian nilai-nilai kosong dengan nilai default.
- Penggabungan data.

Loading

Fase load merupakan tahapan yang berfungsi untuk memasukkan data ke dalam target akhir, yaitu ke dalam suatu data warehouse. Waktu dan jangkauan untuk mengganti atau menambah data tergantung pada perancangan data warehouse pada waktu menganalisa keperluan informasi. Fase load berinteraksi dengan suatu database, constraint didefinisikan dalam skema database sebagai suatu trigger yang diaktifkan pada waktu melakukan load data (contohnya : uniqueness,referential, integrity, mandatory fields), yang juga berkontribusi untuk keseluruhan tampilan dan kualitas data dari proses ETL.

Fase pemuatan memuat data ke target akhir yang mungkin berupa flat file sederhana atau gudang data. Bergantung pada persyaratan organisasi, proses ini sangat bervariasi. Beberapa gudang data dapat menimpa informasi yang ada dengan informasi kumulatif; memperbarui data yang diekstraksi sering dilakukan setiap hari, mingguan, atau bulanan. Gudang data lain (atau bahkan bagian lain dari gudang data yang sama) dapat menambahkan data baru dalam bentuk historis secara berkala — misalnya setiap jam. Untuk memahami hal ini, pertimbangkan gudang data yang diperlukan untuk mengelola catatan penjualan tahun lalu. Gudang data ini menimpa data yang lebih lama dari satu tahun dengan data yang lebih baru. Namun, entri data untuk jendela satu tahun dibuat secara historis. Waktu dan ruang lingkup untuk mengganti atau menambahkan adalah pilihan desain strategis tergantung pada waktu yang tersedia dan kebutuhan bisnis. Sistem yang lebih kompleks dapat mempertahankan sejarah dan jejak audit dari semua perubahan pada data yang dimuat di gudang data.

Saat fase pemuatan berinteraksi dengan database, batasan yang ditentukan dalam skema database - serta pemicu yang diaktifkan saat pemuatan data - berlaku (misalnya, keunikan, integritas referensial, bidang wajib), yang juga berkontribusi pada kinerja kualitas data secara keseluruhan dari proses ETL.

- Misalnya, lembaga keuangan mungkin memiliki informasi tentang pelanggan di beberapa departemen dan setiap departemen mungkin memiliki informasi pelanggan yang dicantumkan dengan cara yang berbeda. Departemen keanggotaan mungkin mencantumkan pelanggan berdasarkan nama, sedangkan departemen akuntansi mungkin mencantumkan pelanggan berdasarkan nomor. ETL dapat menggabungkan semua elemen data ini dan menggabungkannya menjadi presentasi yang seragam, seperti untuk disimpan dalam database atau gudang data.

- Cara lain perusahaan menggunakan ETL adalah dengan memindahkan informasi ke aplikasi lain secara permanen. Misalnya, aplikasi baru mungkin menggunakan vendor database lain dan kemungkinan besar skema database yang sangat berbeda. ETL dapat digunakan untuk mengubah data menjadi format yang sesuai untuk digunakan aplikasi baru.
- Contohnya adalah Expense and Cost Recovery System (ECRS) seperti yang digunakan oleh akuntans, konsultan, dan firma hukum. Data biasanya berakhir di sistem waktu dan penagihan, meskipun beberapa bisnis juga dapat memanfaatkan data mentah untuk laporan produktivitas karyawan ke Manusia

Tahapan dan Proses Real-life ETL

Real-life ETL Cycle yang khas terdiri dari langkah-langkah eksekusi berikut:

1. Cycle initiation
2. Mendapatkan referensi data
3. Extract (dari sumber)
4. Validasi
5. Transform ((membersihkan, menerapkan aturan bisnis, memeriksa integritas data, membuat agregat atau disagregasi)
6. Stage (memuat ke dalam tabel pementasan, jika diperlukan)
7. Audit reports (misalnya, tentang kepatuhan terhadap aturan bisnis. Juga, jika terjadi kegagalan, membantu mendiagnosis / memperbaiki)
8. Publish (ke target tabel)
9. Archive

Tantangan dan Hambatan

Proses ETL dapat melibatkan kompleksitas yang cukup besar, dan masalah operasional yang signifikan dapat terjadi dengan sistem ETL yang dirancang dengan tidak tepat.

Kisaran nilai data atau kualitas data dalam sistem operasional dapat melebihi ekspektasi desainer pada saat aturan validasi dan transformasi ditentukan. Pembuatan profil data suatu sumber selama analisis data dapat mengidentifikasi kondisi data yang harus dikelola oleh spesifikasi aturan transformasi, yang mengarah ke amandemen aturan validasi yang diterapkan secara eksplisit dan implisit dalam proses ETL.

Gudang data biasanya dirakit dari berbagai sumber data dengan format dan tujuan berbeda. Dengan demikian, ETL adalah proses utama untuk menyatukan semua data dalam lingkungan standar yang homogen.

Analis desain harus menetapkan skalabilitas sistem ETL selama masa penggunaannya - termasuk memahami volume data yang harus diproses dalam perjanjian tingkat layanan. Waktu yang tersedia untuk mengekstrak dari sistem sumber dapat berubah, yang berarti jumlah data yang sama mungkin harus diproses dalam waktu yang lebih singkat. Beberapa sistem ETL harus melakukan penskalaan untuk memproses terabyte data untuk memperbarui gudang data dengan data puluhan terabyte. Peningkatan volume data mungkin memerlukan desain yang dapat diskalakan dari batch harian ke beberapa hari mikro batch untuk integrasi dengan antrian pesan atau pengambilan data perubahan waktu nyata untuk transformasi dan pembaruan berkelanjutan.

Hasil Kinerja ETL

Vendor ETL mengukur sistem rekaman mereka pada beberapa TB (terabyte) per jam (atau ~ 1 GB per detik) menggunakan server yang kuat dengan banyak CPU, banyak hard drive, beberapa koneksi gigabit-jaringan, dan banyak memori. Rekor ETL tercepat saat ini dipegang oleh Syncsort, Vertica, dan HP sebesar 5,4TB dalam waktu kurang dari satu jam, yang lebih dari dua kali lebih cepat dari rekor sebelumnya yang dipegang oleh Microsoft dan Unisys.

Dalam kehidupan nyata, bagian paling lambat dari proses ETL biasanya terjadi pada fase pemuatan database. Basis data dapat bekerja lambat karena harus menjaga konkurensi, pemeliharaan integritas, dan indeks. Jadi, untuk kinerja yang lebih baik, mungkin masuk akal untuk menggunakan:

- Metode Ekstrak Jalur Langsung atau pembongkaran massal jika memungkinkan (alih-alih meminta basis data) untuk mengurangi beban pada sistem sumber sekaligus mendapatkan ekstrak kecepatan tinggi
- Sebagian besar pemrosesan transformasi di luar database
- Operasi pemuatan massal jika memungkinkan

Namun, meskipun menggunakan operasi massal, akses database biasanya menjadi penghambat dalam proses ETL. Beberapa metode umum yang digunakan untuk meningkatkan kinerja adalah:

- Tabel partisi (dan indeks): coba pertahankan ukuran partisi yang sama (perhatikan nilai null yang dapat mengubah partisi)
- Lakukan semua validasi di lapisan ETL sebelum pemuatan: nonaktifkan pemeriksaan integritas (nonaktifkan batasan ...) di tabel database target selama pemuatan
- Nonaktifkan pemicu (nonaktifkan pemicu ...) dalam tabel database target selama pemuatan: simulasikan efeknya sebagai langkah terpisah
- Hasilkan ID di lapisan ETL (bukan di database)
- Letakkan indeks (pada tabel atau partisi) sebelum pemuatan - dan buat kembali setelah pemuatan (SQL: jatuhkan indeks ...; buat indeks ...)
- Gunakan pemuatan massal paralel bila memungkinkan - berfungsi dengan baik saat tabel dipartisi atau tidak ada indeks (Catatan: upaya untuk melakukan pemuatan paralel ke dalam tabel yang sama (partisi) biasanya menyebabkan kunci - jika tidak pada baris data, maka pada indeks)
- Jika ada persyaratan untuk melakukan penyisipan, pembaruan, atau penghapusan, cari tahu baris mana yang harus diproses dengan cara apa di lapisan ETL, dan kemudian proses ketiga operasi ini dalam database secara terpisah; Anda sering dapat melakukan pemuatan massal untuk penyisipan, tetapi pembaruan dan penghapusan biasanya melalui API (menggunakan SQL)

Apakah akan melakukan operasi tertentu dalam database atau di luar mungkin melibatkan trade-off. Misalnya, menghapus duplikat menggunakan perbedaan mungkin lambat dalam database; jadi, masuk akal untuk melakukannya di luar. Di sisi lain, jika menggunakan different secara signifikan (x100) mengurangi jumlah baris yang akan diekstraksi, maka masuk akal untuk menghapus duplikasi sedini mungkin di database sebelum membongkar data.

Sumber masalah yang umum di ETL adalah banyaknya ketergantungan di antara pekerjaan ETL. Misalnya, pekerjaan "B" tidak dapat dimulai sementara pekerjaan "A" belum selesai. Seseorang biasanya dapat mencapai kinerja yang lebih baik dengan memvisualisasikan semua proses pada grafik, dan mencoba untuk mengurangi grafik dengan memanfaatkan paralelisme secara maksimal, dan membuat "rantai" pemrosesan berurutan sesingkat mungkin. Sekali lagi, mempartisi tabel besar dan indeksnya dapat sangat membantu.

Masalah umum lainnya terjadi ketika data tersebar di antara beberapa database, dan pemrosesan dilakukan di database tersebut secara berurutan. Terkadang replikasi database mungkin terlibat sebagai metode menyalin data antar database - ini dapat

memperlambat keseluruhan proses secara signifikan. Solusi umum adalah mengurangi grafik pemrosesan menjadi hanya tiga lapisan:

- Sumber
- layer/Lapisan Pusat ETL
- Target

Pendekatan ini memungkinkan pemrosesan untuk memanfaatkan paralelisme secara maksimal. Misalnya, jika Anda perlu memuat data ke dalam dua database, Anda dapat menjalankan beban secara paralel (alih-alih memuat ke database pertama - dan kemudian mereplikasi ke database kedua).

Terkadang pemrosesan harus dilakukan secara berurutan. Misalnya, data dimensional (referensi) diperlukan sebelum seseorang bisa mendapatkan dan memvalidasi baris untuk tabel "Facts" utama.

Pemrosesan Paralel

Perkembangan terbaru dalam perangkat lunak ETL adalah penerapan pemrosesan paralel (*parallel processing*). Ini telah mengaktifkan sejumlah metode untuk meningkatkan kinerja ETL secara keseluruhan saat menangani data dalam volume besar.

Aplikasi ETL menerapkan tiga jenis utama paralelisme:

- Data: Dengan membagi satu file berurutan menjadi file data yang lebih kecil untuk menyediakan akses paralel
- Pipeline: memungkinkan pengoperasian beberapa komponen secara bersamaan pada aliran data yang sama, mis. mencari nilai pada catatan 1 sekaligus menambahkan dua bidang pada catatan 2
- Komponen: Menjalankan beberapa proses secara bersamaan pada aliran data yang berbeda dalam pekerjaan yang sama, mis. mengurutkan satu file input sambil menghapus duplikat di file lain

Ketiga jenis paralelisme biasanya beroperasi digabungkan dalam satu pekerjaan.

Kesulitan tambahan muncul dengan memastikan bahwa data yang diunggah relatif konsisten. Karena beberapa database sumber mungkin memiliki siklus pembaruan yang berbeda (beberapa mungkin diperbarui setiap beberapa menit, sementara yang lain mungkin memerlukan beberapa hari atau minggu), sistem ETL mungkin diperlukan untuk

menahan data tertentu hingga semua sumber disinkronkan. Demikian juga, di mana gudang mungkin harus direkonsiliasi dengan konten dalam sistem sumber atau dengan buku besar, menetapkan titik sinkronisasi dan rekonsiliasi menjadi perlu.

Rerunnability dan Recoverability

Prosedur data warehousing biasanya membagi proses ETL yang besar menjadi bagian-bagian kecil yang berjalan secara berurutan atau paralel. Untuk melacak aliran data, masuk akal untuk memberi tag pada setiap baris data dengan "row_id", dan menandai setiap bagian dari proses dengan "run_id". Jika terjadi kegagalan, memiliki ID ini membantu untuk memutar kembali dan menjalankan kembali bagian yang gagal. Praktik terbaik juga membutuhkan checkpoint, yang merupakan status ketika fase tertentu dari proses selesai. Setelah berada di pos pemeriksaan, ada baiknya untuk menulis semuanya ke disk, membersihkan beberapa file sementara, mencatat status, dan sebagainya.

ETL Virtual

Sejak 2010, virtualisasi data telah mulai memajukan pemrosesan ETL. Penerapan virtualisasi data ke ETL memungkinkan penyelesaian tugas ETL yang paling umum dari migrasi data dan integrasi aplikasi untuk beberapa sumber data yang tersebar. ETL virtual beroperasi dengan representasi abstrak dari objek atau entitas yang dikumpulkan dari berbagai sumber data relasional, semi-terstruktur, dan tidak terstruktur. Alat ETL dapat memanfaatkan pemodelan berorientasi objek dan bekerja dengan representasi entitas yang disimpan secara terus-menerus dalam arsitektur hub-and-spoke yang terletak di pusat. Koleksi semacam itu yang berisi representasi entitas atau objek yang dikumpulkan dari sumber data untuk pemrosesan ETL disebut repositori metadata dan dapat berada di memori atau dibuat tetap. Dengan menggunakan repositori metadata persisten, alat ETL dapat bertransisi dari proyek satu kali ke middleware persisten, melakukan harmonisasi data dan pembuatan profil data secara konsisten dan hampir secara real-time.

4.8 Skema Bintang (Star Schema)

Dalam komputasi, skema bintang adalah gaya skema data mart yang paling sederhana dan merupakan pendekatan yang paling banyak digunakan untuk mengembangkan gudang data dan data mart dimensi. Skema bintang terdiri dari satu atau lebih tabel fakta yang mereferensikan sejumlah tabel dimensi. Skema bintang adalah kasus khusus yang penting dari skema kepingan salju, dan lebih efektif untuk menangani kueri yang lebih sederhana.

Skema bintang mendapatkan namanya dari kemiripan model fisik dengan bentuk bintang dengan tabel fakta di tengahnya dan tabel dimensi yang mengelilinginya mewakili titik-titik bintang.

Model

Skema bintang memisahkan data proses bisnis menjadi fakta, yang menyimpan data kuantitatif terukur tentang suatu bisnis, dan dimensi yang merupakan atribut deskriptif yang terkait dengan data fakta. Contoh data fakta antara lain harga jual, kuantitas penjualan, dan waktu, jarak, kecepatan, dan pengukuran berat. Contoh atribut dimensi terkait mencakup model produk, warna produk, ukuran produk, lokasi geografis, dan nama staf penjualan.

Skema bintang yang memiliki banyak dimensi terkadang disebut skema kelabang. Memiliki dimensi yang hanya terdiri dari beberapa atribut, meskipun lebih sederhana untuk dikelola, menghasilkan kueri dengan banyak tabel bergabung dan membuat skema bintang kurang mudah digunakan.

Tabel Fakta

Tabel fakta mencatat pengukuran atau metrik untuk peristiwa tertentu. Tabel fakta umumnya terdiri dari nilai numerik, dan kunci asing ke data dimensi tempat informasi deskriptif disimpan. Tabel fakta dirancang untuk tingkat detail seragam yang rendah (disebut sebagai "perincian" atau "butir"), yang berarti fakta dapat merekam peristiwa pada tingkat yang sangat atomik. Hal ini dapat mengakibatkan akumulasi sejumlah besar rekaman dalam tabel fakta seiring waktu. Tabel fakta didefinisikan sebagai salah satu dari tiga jenis:

- Tabel fakta transaksi mencatat fakta tentang peristiwa tertentu (misalnya, peristiwa penjualan)
- Tabel fakta snapshot mencatat fakta pada waktu tertentu (misalnya, detail akun di akhir bulan)
- Tabel ringkasan yang terakumulasi mencatat fakta agregat pada titik waktu tertentu (misalnya, total penjualan bulanan untuk suatu produk)

Tabel fakta umumnya diberi kunci pengganti untuk memastikan setiap baris dapat diidentifikasi secara unik. Kunci ini adalah kunci utama sederhana.

Tabel Dimensi

Tabel dimensi biasanya memiliki jumlah record yang relatif kecil dibandingkan dengan tabel fakta, tetapi setiap record mungkin memiliki jumlah atribut yang sangat besar untuk menggambarkan data fakta. Dimensi dapat menentukan berbagai macam karakteristik, tetapi beberapa atribut paling umum yang ditentukan oleh tabel dimensi meliputi:

- Tabel dimensi waktu mendeskripsikan waktu pada level terendah dari perincian waktu yang kejadiannya dicatat dalam skema bintang
- Tabel dimensi geografi menggambarkan data lokasi, seperti negara, negara bagian, atau kota
- Tabel dimensi produk menjelaskan produk
- Tabel dimensi karyawan menggambarkan karyawan, seperti staf penjualan
- Tabel dimensi rentang menjelaskan rentang waktu, nilai dolar, atau jumlah terukur lainnya untuk menyederhanakan pelaporan

Tabel dimensi umumnya diberi kunci primer pengganti, biasanya tipe data bilangan bulat kolom tunggal, yang dipetakan ke kombinasi atribut dimensi yang membentuk kunci alami.

Kelebihan Skema Bintang

Skema bintang didenormalisasi, artinya aturan normalisasi normal yang diterapkan pada database relasional transaksional dilonggarkan selama desain dan implementasi skema bintang. Manfaat denormalisasi skema bintang adalah:

- Kueri yang lebih sederhana - logika bergabung skema bintang umumnya lebih sederhana daripada logika bergabung yang diperlukan
- mengambil data dari skema transaksional yang sangat dinormalisasi.
- Logika pelaporan bisnis yang disederhanakan - jika dibandingkan dengan skema yang sangat dinormalisasi, skema bintang menyederhanakan logika pelaporan bisnis umum, seperti periode-demi-periode dan sebagai-pelaporan.
- Keuntungan kinerja kueri - skema bintang dapat memberikan peningkatan kinerja untuk aplikasi pelaporan hanya-baca jika dibandingkan dengan skema yang sangat dinormalisasi.
- Agregasi cepat - kueri yang lebih sederhana terhadap skema bintang dapat menghasilkan peningkatan kinerja untuk operasi agregasi.
- Data Cubes- skema bintang digunakan oleh semua sistem OLAP untuk membangun kubus OLAP milik sendiri secara efisien; pada kenyataannya, sebagian besar

sistem OLAP menyediakan mode operasi ROLAP yang dapat menggunakan skema bintang secara langsung sebagai sumber tanpa membangun struktur kubus berpemilik.

Kelemahan Skema Bintang

Kerugian utama dari skema bintang adalah integritas data tidak ditegakkan seperti halnya dalam database yang sangat dinormalisasi. Penyisipan dan pembaruan satu kali dapat menghasilkan anomali data yang dirancang untuk dihindari oleh skema yang dinormalisasi. Secara umum, skema bintang dimuat dengan cara yang sangat terkontrol melalui pemrosesan batch atau "umpan tetesan" yang mendekati waktu nyata, untuk mengimbangi kurangnya perlindungan yang diberikan oleh normalisasi.

Skema bintang juga tidak sefleksibel dalam hal kebutuhan analitis seperti model data yang dinormalisasi. Model yang dinormalkan memungkinkan semua jenis kueri analitik dieksekusi selama mereka mengikuti logika bisnis yang ditentukan dalam model. Skema bintang cenderung lebih dibuat khusus untuk tampilan data tertentu, sehingga tidak memungkinkan analisis yang lebih kompleks. Skema bintang tidak mendukung hubungan banyak-ke-banyak antara entitas bisnis - setidaknya tidak secara alami. Biasanya hubungan ini disederhanakan dalam skema bintang agar sesuai dengan model dimensi sederhana.

Contoh Penggunaan

Mengambil contoh tentang sebuah database penjualan, mungkin dari rantai toko, diklasifikasikan menurut tanggal, toko, dan produk. Gambar skema di sebelah kanan adalah versi skema bintang dari skema sampel yang diberikan di artikel skema kepingan salju.

Fact_Sales adalah tabel fakta dan ada tiga tabel dimensi Dim_Date, Dim_Store dan Dim_ Product.

Setiap tabel dimensi memiliki kunci utama pada kolom Id-nya, yang berkaitan dengan salah satu kolom (dilihat sebagai baris dalam skema contoh) dari kunci utama tiga kolom (gabungan) tabel Fact_Sales (Date_Id, Store_Id, Product_Id). Kolom non-primary key Units_Sold dari tabel fakta dalam contoh ini menunjukkan ukuran atau metrik yang dapat

digunakan dalam penghitungan dan analisis. Kolom non-kunci utama dari tabel dimensi mewakili atribut tambahan dari dimensi (seperti dimensi Tahun Dim_Date).

Misalnya, kueri berikut menjawab berapa banyak TV yang telah terjual, untuk setiap merek dan negara, pada tahun 1997:

```
SELECT
 P.Brand,
S.Country AS Countries,
 SUM(F.Units_Sold)
FROM Fact_Sales F
INNER JOIN Dim_Date D     ON (F.Date_Id = D.Id)
INNER JOIN Dim_Store S    ON (F.Store_Id = S.Id)
INNER JOIN Dim_Product P ON (F.Product_Id = P.Id)
WHERE D.Year = 1997 AND  P.Product_Category = 'tv'
GROUP BY
 P.Brand,
 S.Country
```

Bab 5 : Studi Terpadu Tentang Riset Pasar

Upaya yang dilakukan untuk mengumpulkan informasi yang terkait dengan pelanggan atau pasar dikenal sebagai riset pasar. Riset pasar adalah bagian penting dari strategi bisnis. Segmentasi pasar, tren pasar, analisis SWOT dan riset pasar adalah beberapa topik yang dijelaskan dalam bab ini.

5.1 Riset Pasar

Riset pasar adalah segala upaya terorganisir untuk mengumpulkan informasi tentang pasar sasaran atau pelanggan. Ini adalah komponen yang sangat penting dari strategi bisnis. Istilah ini biasanya dipertukarkan dengan riset pemasaran; namun, praktisi ahli mungkin ingin menarik perbedaan, karena riset pemasaran berkaitan secara khusus tentang proses pemasaran, sedangkan riset pasar berkaitan secara khusus dengan pasar.

Riset pasar adalah salah satu faktor kunci yang digunakan dalam mempertahankan daya saing atas pesaing. Riset pasar memberikan informasi penting untuk mengidentifikasi dan menganalisis kebutuhan pasar, ukuran pasar, dan persaingan. Teknik riset pasar mencakup teknik kualitatif seperti kelompok fokus, wawancara mendalam, dan etnografi, serta teknik kuantitatif seperti survei pelanggan, dan analisis data sekunder.

Riset pasar, yang mencakup riset sosial dan opini, adalah pengumpulan dan interpretasi sistematis informasi tentang individu atau organisasi menggunakan metode dan teknik statistik dan analitis dari ilmu sosial terapan untuk mendapatkan wawasan atau mendukung pengambilan keputusan.

Sejarah Perkembangan Riset Pasar

Riset pasar mulai dikonseptualisasikan dan dipraktikkan secara formal selama tahun 1920-an, sebagai cabang dari ledakan periklanan Zaman Keemasan radio di Amerika Serikat. Pengiklan mulai menyadari pentingnya demografi yang diungkapkan dengan mensponsori berbagai program radio. Untuk lebih memahami riset pemasaran kita akan membahas secara ringkas riwayat perkembangannya. Perkembangan riset pemasaran di awal abad 20 sejajar dengan tumbuhnya konsep pemasaran. Mulai periode ini, filsafat manajemen yang memberi arah pada organisasi secara perlahan-lahan berubah ke arah orientasi

konsumen seperti apa yang kita alami sekarang. Selama periode tahun 1900-1930, perhatian manajemen terutama difokuskan pada masalah-masalah dan peluang-peluang yang berkaitan dengan produksi ; antara tahun 1930 sampai akhir tahun 1940-an, orientasi berubah ke masalah dan peluang yang berkaitan dengan distribusi; sejak akhir tahun 1940-an perhatian mulai terarah pada kebutuhan dan keinginan konsumen. Perubahan filsafat manajemen dalam periode-periode tersebut tercermin dalam kegiatan pemasaran dari pelbagai organisasi. Meskipun sudah ada beberapa orang dan lembaga yang terlibat dalam kegiatan riset pemasaran insidentil sebelum tahun 1910, periode 1910-1920 dikenal sebagai periode penggunaan formal riset pemasaran. Di tahun 1911, J. George Frederick mendirikan perusahaan riset dengan nama The Business Bourse. Charles Coolidge Parlin, pada tahun yang sama, ditunjuk sebagai manajer Divisi Riset Komersiil dari Curtis Publishing Company. Pemakaian nama "riset komersiil" mempunyai makna khusus karena kebanyakan orang-orang bisnis menganggap istilah riset terlalu "tinggi" untuk jasa bisnis. Parlin mengelola salah satu organisasi riset terkemuka dalam periode ini. Keberhasilan pekerjaan Parlin, memberi insiparai kepada beberapa perusahaan industri dan media periklanan untuk mendirikan divisi riset. Pada tahun 1915, United States Rubber Company mengontrak Dr. Paul H. Nystrom dengan tugas utama mengelola sebuah departemen baru, Departemen Riset Komersiil. Di tahun 1917, Swiff dan Company mengontrak Dr. Louis D.H. Weld dari Yale University untuk menjadi manajer dari Departemen Riset Komersiil. Di tahun 1919, Profesor C.S. Duncan dari University of Chicago menerbitkan Commercial Research: An Outline of Working Principles. Karya ini dianggap sebagai buku utama yang pertama kali diterbutkan dalam bidang riset komersiil.

Di tahun 1921, Market Analysis karya Percival White diterbitkan dan merupakan buku tentang riset yang untuk pertama kali memperoleh sambutan hangat. Buku ini mengalami cetak ulang beberapa kali. Ditahun 1937 Lyndon O. Brown menerbitkan Market Research and Analysis. Buku ini menjadi salah satu buku teks perguruan tinggi yang terkenal dan fenomena ini mencerminkan minat yang makin berkembang terhadap riset pemasaran di kalangan kampus. Penerimaan konsep pemasaran mendorong perubahan, dari tekanan ke "riset pasar" menuju ke "riset pemasaran". Riset pasar mengimplikasikan bahwa fokus riset adalah pada analisis pasar. Pergeseran ke riset pemasaran memperluas hakikat dan peranan riset, dengan tekanan pada hubungan antara peneliti dengan proses manajemen pemasaran. Publikasi riset pemasaran karangan Harper Boyd dan Ralph Westfall di tahun 1956 mencerminkan perubahan dalam orientasi ini.

Riset Pasar untuk Bisnis / Perencanaan

Riset pasar adalah cara mendapatkan gambaran umum tentang keinginan, kebutuhan, dan kepercayaan konsumen. Ini juga dapat melibatkan menemukan bagaimana mereka bertindak. Penelitian ini dapat digunakan untuk menentukan bagaimana suatu produk dapat dipasarkan. Peter Drucker percaya penelitian pasar menjadi intisari dari pemasaran.

Ada dua jenis utama riset pasar. Penelitian Utama dibagi menjadi penelitian Kuantitatif dan Kualitatif dan penelitian Sekunder.

Faktor-faktor yang dapat diselidiki melalui riset pasar meliputi :

Informasi pasar

Melalui informasi Pasar, seseorang dapat mengetahui harga berbagai komoditas di pasar, serta situasi penawaran dan permintaan. Peneliti pasar memiliki peran yang lebih luas daripada yang sebelumnya diakui dengan membantu klien mereka memahami aspek sosial, teknis, dan bahkan hukum pasar.

Segmentasi pasar

Segmentasi pasar adalah pembagian pasar atau populasi ke dalam subkelompok dengan motivasi yang sama. Ini banyak digunakan untuk segmentasi pada perbedaan geografis, perbedaan kepribadian, perbedaan demografis, perbedaan teknis, penggunaan perbedaan produk, perbedaan psikografis dan perbedaan gender. Untuk segmentasi B2B, firmografi biasanya digunakan.

Trend pasar

Tren pasar adalah pergerakan pasar ke atas atau ke bawah, selama periode waktu tertentu. Menentukan ukuran pasar mungkin lebih sulit jika seseorang memulai dengan inovasi baru. Dalam hal ini, Anda harus menghitung angka dari jumlah pelanggan potensial, atau segmen pelanggan. [Ilar 1998]

Analisis SWOT

SWOT adalah analisis tertulis tentang Kekuatan, Kelemahan, Peluang, dan Ancaman terhadap entitas bisnis. SWOT tidak hanya digunakan pada tahap penciptaan perusahaan tetapi juga dapat digunakan sepanjang masa perusahaan. SWOT juga dapat ditulis untuk kompetisi untuk memahami bagaimana mengembangkan bauran pemasaran dan produk.

Faktor lain yang dapat diukur adalah efektivitas pemasaran. Ini termasuk

- Analisis pelanggan
- Pemodelan pilihan
- Analisis pesaing
- Analisis resiko
- Penelitian produk
- Mengiklankan penelitian
- Pemodelan bauran pemasaran
- Simulasi Pemasaran Uji

Manfaat Riset Pasar

Riset Pasar adalah alat penting yang membantu dalam membuat keputusan strategis. Ini mengurangi risiko yang terlibat dalam pengambilan keputusan serta strategi. Perusahaan melakukan penelitian ini di rumah atau mengalihdayakan proses ini kepada pakar bisnis atau organisasi yang telah mendedikasikan dan melatih sumber daya untuk melakukan ini. Dalam beberapa tahun terakhir, tren yang meningkat dari perusahaan riset pasar yang membantu ahli strategi bisnis telah muncul. Beberapa manfaat utamanya adalah:

a. Riset pemasaran membantu dalam memberikan tren yang akurat dan terbaru terkait dengan permintaan, perilaku konsumen, penjualan, peluang pertumbuhan, dll.
b. Membantu dalam pemahaman yang lebih baik tentang pasar, sehingga membantu dalam desain produk, fitur dan prakiraan permintaan
c. Membantu dalam mempelajari dan memahami pesaing, sehingga mengidentifikasi proposisi penjualan unik untuk bisnis.

Riset pasar memberikan banyak manfaat khususnya pada bidang Industri dan bisnis serta memiliki ROI yang tinggi.

Riset Pasar untuk Industri Film

Penting untuk menguji materi pemasaran film untuk melihat bagaimana penonton akan menerimanya. Ada beberapa praktik riset pasar yang dapat digunakan:

(1) pengujian konsep, yang mengevaluasi reaksi terhadap ide film dan cukup jarang;
(2) studio positioning, yang menganalisis naskah untuk peluang pemasaran;
(3) kelompok fokus, yang menyelidiki pendapat penonton tentang sebuah film dalam kelompok kecil sebelum dirilis;

(4) uji pemutaran, yang melibatkan pratinjau film sebelum rilis teater;
(5) studi pelacakan, yang mengukur (seringkali melalui polling telepon) kesadaran penonton terhadap sebuah film setiap minggu sebelum dan selama rilis teater;
(6) pengujian periklanan, yang mengukur tanggapan terhadap materi pemasaran seperti trailer dan iklan televisi; dan terakhir
(7) exit survey, yang mengukur reaksi penonton setelah menonton film di bioskop.

Pengaruh Internet

Ketersediaan penelitian melalui Internet telah mempengaruhi sejumlah besar konsumen yang menggunakan media ini; untuk mendapatkan pengetahuan yang berkaitan dengan hampir semua jenis produk dan layanan yang tersedia. Ini telah ditambahkan oleh faktor pertumbuhan pasar global yang baru muncul, seperti Cina, Indonesia dan Rusia, yang secara signifikan melebihi dari pasar e-commerce B2B yang mapan dan lebih maju. Berbagai statistik menunjukkan bahwa meningkatnya permintaan konsumen tercermin tidak hanya dalam jangkauan luas dan beragam aplikasi riset Internet umum, tetapi juga dalam penetrasi riset belanja online.

Hal ini didorong oleh situs web, grafik, dan konten yang menyempurnakan produk yang dirancang untuk menarik pembeli biasa yang “berselancar”, meneliti kebutuhan khusus mereka, harga dan kualitas yang kompetitif. Menurut Small Business Administration (SBA), bisnis yang sukses dikontribusikan secara signifikan dengan mendapatkan pengetahuan tentang pelanggan, pesaing, dan industri terkait. Riset pasar tidak hanya menciptakan pemahaman ini, tetapi juga proses analisis data mengenai produk dan layanan mana yang diminati.

Kemudahan dan kemudahan akses Internet telah menciptakan fasilitas penelitian B2C E-commerce global, untuk jaringan belanja online yang luas yang telah memotivasi pasar ritel di negara-negara maju. Pada tahun 2010, antara $ 400 miliar dan $ 600 miliar pendapatan dihasilkan oleh media ini juga, diantisipasi bahwa pada tahun 2015, pasar online ini akan menghasilkan pendapatan antara $ 700 miliar dan $ 950 miliar. Pengaruh riset pasar, apa pun bentuknya, merupakan insentif yang sangat kuat untuk semua jenis konsumen dan penyedia mereka!

Selain aktivitas riset pasar berbasis web online, Internet juga telah memengaruhi mode pengumpulan data jalan raya dengan, misalnya, mengganti papan klip kertas tradisional dengan penyedia survei online. Selama 5 tahun terakhir, survei seluler menjadi semakin

populer. Seluler telah membuka pintu ke metode baru yang inovatif untuk melibatkan responden, seperti komunitas pemungutan suara sosial.

Penelitian dan Aplikasi Media Sosial

Riset UK Market Research Society (MRS) menunjukkan bahwa rata-rata, tiga platform media sosial yang terutama digunakan oleh kaum Milenial adalah LinkedIn, Facebook, dan YouTube. Aplikasi Media Sosial, menurut T-Systems, membantu menghasilkan pasar B2B E-commerce dan mengembangkan efisiensi proses bisnis elektronik. Aplikasi ini adalah sarana yang sangat efektif untuk riset pasar, yang dikombinasikan dengan E-commerce, sekarang dianggap sebagai bidang bisnis global yang terpisah dan sangat menguntungkan. Sementara banyak model bisnis B2B sedang diperbarui, berbagai keuntungan dan manfaat yang ditawarkan oleh platform Media Sosial diintegrasikan di dalamnya.

Organisasi intelijen bisnis telah menyusun laporan komprehensif terkait dengan penjualan ritel online global, yang menjelaskan pola dan tren pertumbuhan berkelanjutan di industri. Dengan judul "Global B2C E-Commerce dan Pasar Pembayaran Online 2014", laporan tersebut melihat penurunan tingkat pertumbuhan secara keseluruhan di Amerika Utara dan Eropa Barat, karena pertumbuhan yang diharapkan dalam penjualan pasar online, diserap ke pasar negara berkembang. Kawasan Asia-Pasifik diperkirakan akan mengalami pertumbuhan tercepat di pasar B2C E-Commerce dan menggantikan Amerika Utara sebagai pemimpin kawasan penjualan B2C E-Commerce, dalam beberapa tahun. Ini secara efektif, menawarkan platform motivasi yang signifikan untuk layanan Internet baru, untuk mempromosikan aplikasi ramah penelitian pasar pengguna.

Sektor Penelitian dan Pasar

Penyedia penjualan online utama di B2C E-Commerce, di seluruh dunia, termasuk Am- azon.com Inc. yang berbasis di AS yang tetap menjadi pendapatan E-Commerce, pemimpin global. Pemimpin pertumbuhan sepuluh besar dunia adalah dua perusahaan online asal China, yang keduanya melakukan Initial Public Offering (IPO) tahun ini; Alibaba Group Holding Ltd. dan JD Inc. Perusahaan lain dari sepuluh besar adalah Cnova N.V., anak perusahaan E-Commerce dari French Group Casino yang baru dibentuk, dengan berbagai pengecer toko yang mengembangkan dan memperluas fasilitas E-Commerce mereka di seluruh dunia. Ini merupakan indikasi lebih lanjut tentang bagaimana konsumen semakin tertarik pada peluang pencarian ulang online dan memperluas kesadaran mereka tentang apa yang tersedia bagi mereka.

Penyedia jasa; misalnya yang terkait dengan keuangan, perdagangan pasar luar negeri dan investasi mempromosikan berbagai informasi dan peluang penelitian kepada pengguna online. Selain itu, mereka memberikan strategi yang komprehensif dan kompetitif dengan alat penelitian pasar, yang dirancang untuk mempromosikan peluang bisnis di seluruh dunia bagi wirausahawan dan penyedia mapan. Akses umum, ke fasilitas riset pasar yang akurat dan didukung, merupakan aspek penting dari perkembangan dan kesuksesan bisnis saat ini. Asosiasi Riset Pemasaran didirikan pada tahun 1957 dan diakui sebagai salah satu asosiasi terkemuka dan terkemuka dalam profesi riset opini dan pemasaran. Ini melayani tujuan memberikan wawasan dan kecerdasan yang membantu bisnis membuat keputusan terkait penyediaan produk dan layanan kepada konsumen dan industri.

Pengetahuan organisasi tentang kondisi pasar dan persaingan diperoleh dengan meneliti sektor terkait, yang memberikan keuntungan untuk masuk ke industri baru dan mapan. Ini memungkinkan strategi yang efektif untuk diterapkan; penilaian lingkungan global di sektor jasa, serta perdagangan pasar luar negeri dan hambatan investasi! Penelitian, digunakan untuk mempromosikan peluang ekspor dan investasi masuk, membantu menentukan bagaimana melaksanakan strategi kompetitif, fokus pada kebijakan yang objektif dan memperkuat peluang global. Ini adalah media yang memengaruhi, mengatur, dan menegakkan perjanjian, preferensi, lingkungan perdagangan yang merata, dan daya saing di pasar internasional.

Aspek industri ritel dari riset pasar online, sedang diubah di seluruh dunia oleh M-Commerce dengan audiens selulernya, yang meningkat pesat seiring dengan peningkatan volume dan variasi produk yang dibeli di media seluler. Riset yang dilakukan di pasar Amerika Utara dan Eropa, menunjukkan bahwa penetrasi M-Commerce terhadap total perdagangan retail online telah mencapai 10% atau lebih. Hal ini juga menunjukkan bahwa di pasar negara berkembang, penetrasi ponsel pintar dan tablet meningkat pesat dan berkontribusi signifikan terhadap pertumbuhan belanja online

5.2 Segmentasi Pasar

Segmentasi pasar adalah proses membagi konsumen atau pasar bisnis yang luas, biasanya terdiri dari pelanggan yang ada dan pelanggan potensial, menjadi sub-kelompok konsumen (dikenal sebagai segmen) berdasarkan beberapa jenis karakteristik bersama. Dalam membagi atau mengelompokkan pasar, peneliti biasanya mencari karakteristik yang sama seperti kebutuhan umum, minat yang sama, gaya hidup yang mirip, atau bahkan profil demografis yang serupa. Tujuan keseluruhan dari segmentasi adalah

untuk mengidentifikasi segmen hasil tinggi - yaitu, segmen yang kemungkinan paling menguntungkan atau yang memiliki potensi pertumbuhan - sehingga dapat dipilih untuk perhatian khusus (yaitu menjadi pasar sasaran). Banyak cara berbeda untuk mensegmentasi pasar telah diidentifikasi. Penjual Business-to-business (B2B) mungkin membagi pasar menjadi berbagai jenis bisnis atau negara. Sedangkan penjual bisnis ke konsumen (B2C) mungkin membagi pasar menjadi segmen demografis, segmen gaya hidup, segmen perilaku atau segmen bermakna lainnya.

Segmentasi pasar mengasumsikan bahwa segmen pasar yang berbeda memerlukan program pemasaran yang berbeda - yaitu, penawaran, harga, promosi, distribusi yang berbeda atau beberapa kombinasi variabel pemasaran. Segmentasi pasar tidak hanya dirancang untuk mengidentifikasi segmen yang paling menguntungkan, tetapi juga untuk mengembangkan profil segmen utama agar lebih memahami kebutuhan dan motivasi pembelian mereka. Wawasan dari analisis segmentasi selanjutnya digunakan untuk mendukung pengembangan dan perencanaan strategi pemasaran. Banyak pemasar menggunakan pendekatan S-T-P; Segmentasi → Penargetan → Positioning untuk memberikan kerangka kerja untuk tujuan perencanaan pemasaran. Artinya, pasar tersegmentasi, satu atau lebih segmen dipilih untuk penargetan, dan produk atau layanan diposisikan dengan cara yang selaras dengan pasar sasaran atau pasar yang dipilih.

Tinjauan Sejarah Segmentasi Pasar

Sejarawan bisnis, Richard S. Tedlow, mengidentifikasi empat tahapan dalam evolusi segmentasi pasar:

- Fragmentasi (sebelum 1880-an): Perekonomian dicirikan oleh pemasok regional kecil yang menjual barang secara lokal atau regional
- Unifikasi atau Pemasaran Massal (1880s-1920s): Ketika sistem transportasi membaik, ekonomi menjadi bersatu. Barang standar dan bermerek didistribusikan di tingkat nasional. Produsen cenderung menuntut standarisasi yang ketat untuk mencapai skala ekonomi dengan maksud untuk menembus pasar pada tahap awal siklus hidup produk, mis. Model T Ford
- Segmentasi (1920-an-1980-an): Ketika ukuran pasar meningkat, pabrikan mampu menghasilkan model berbeda yang dipasang pada titik kualitas berbeda untuk memenuhi kebutuhan berbagai segmen pasar demografis dan psikografis. Ini adalah era diferensiasi pasar berdasarkan faktor demografis, sosial ekonomi dan gaya hidup.

- Hyper-segmentation (1980-an +): pergeseran ke arah definisi segmen pasar yang semakin sempit. Kemajuan teknologi, terutama di bidang komunikasi digital, memungkinkan pemasar untuk berkomunikasi dengan konsumen individu atau kelompok yang sangat kecil.

Model T *Ford Henry Ford terkenal mengatakan "Setiap pelanggan dapat memiliki mobil yang dicat dengan warna apa pun yang dia inginkan selama itu hitam"*

Segmentasi pasar kontemporer muncul pada abad ke-20 karena para pemasar menanggapi dua masalah yang mendesak. Data demografis dan pembelian tersedia untuk kelompok tetapi jarang untuk individu dan kedua, saluran periklanan dan distribusi tersedia untuk kelompok, tetapi jarang untuk konsumen tunggal dan pemasar merek mendekati tugas dari sudut pandang taktis. Jadi, segmentasi pada dasarnya adalah proses yang digerakkan oleh merek. Sampai saat ini, sebagian besar pendekatan segmentasi mempertahankan perspektif taktis ini karena mereka menjawab pertanyaan-pertanyaan langsung jangka pendek; biasanya keputusan tentang "pasar yang dilayani" saat ini dan berkaitan dengan menginformasikan keputusan bauran pemasaran.

Evaluasi dalam Segmentasi Pasar

Keterbatasan segmentasi konvensional telah didokumentasikan dengan baik dalam literatur. Pemahaman yang sudah berlangsung lama ini meliputi:

- bahwa tidak ada yang lebih baik daripada pemasaran massal dalam membangun merek
- bahwa dalam pasar yang kompetitif, segmen jarang menunjukkan perbedaan besar dalam cara mereka menggunakan merek
- gagal mengidentifikasi cluster yang cukup sempit
- Segmentasi geografis / demografis terlalu deskriptif dan kurang memiliki wawasan yang cukup tentang motivasi yang diperlukan untuk mendorong strategi komunikasi

- kesulitan dengan dinamika pasar, terutama ketidakstabilan segmen dari waktu ke waktu dan perubahan struktural yang mengarah pada merayapnya segmen dan migrasi keanggotaan saat individu berpindah dari satu segmen ke segmen lainnya

Segmentasi pasar mendapat banyak kritik. Namun terlepas dari keterbatasannya, segmentasi pasar tetap menjadi salah satu konsep yang bertahan lama dalam pemasaran dan terus digunakan secara luas dalam praktik. Sebuah penelitian di Amerika, misalnya, menunjukkan bahwa hampir 60 persen eksekutif senior telah menggunakan segmentasi pasar dalam dua tahun terakhir.

Strategi Segmentasi Pasar

Pertimbangan utama bagi pemasar adalah apakah melakukan segmentasi atau tidak. Bergantung pada filosofi perusahaan, sumber daya, jenis produk atau karakteristik pasar, suatu bisnis dapat mengembangkan pendekatan yang tidak berbeda atau pendekatan yang berbeda. Dalam pendekatan yang tidak berbeda (juga dikenal sebagai pemasaran massal), pemasar mengabaikan segmentasi dan mengembangkan produk yang memenuhi kebutuhan pembeli dengan jumlah terbesar. Dalam pendekatan yang berbeda, perusahaan menargetkan satu atau lebih segmen pasar, dan mengembangkan penawaran terpisah untuk setiap segmen.

Dalam pemasaran konsumen, sulit untuk menemukan contoh pendekatan yang tidak berbeda. Bahkan barang-barang seperti garam dan gula, yang dulunya diperlakukan sebagai komoditas, sekarang sangat dibedakan. Konsumen dapat membeli berbagai produk garam; garam dapur, garam dapur, garam laut, garam batu, garam halal, garam mineral, garam herbal atau sayuran, garam beryodium, pengganti garam dan jika itu tidak cukup pilihan, di tingkat merek, koki gourmet cenderung menjadi yang utama perbedaan antara garam Maldon dan merek pesaing lainnya. Tabel berikut menguraikan pendekatan strategis utama.

Tabel 1: Pendekatan Strategis Utama untuk Segmentasi

Jumlah Segmen	Segmentasi Strategi	Catatan
Nol	Strategi yang tidak terdiferensiasi	Pemasaran secara massal
Satu	Strategi Fokus	Pemasaran khusus
Dua atau Lebih	Strategi berbeda	Banyak ceruk
Ribuan	Hipersegmentasi	Pemasaran one-to-one

Sejumlah faktor cenderung memengaruhi strategi segmentasi perusahaan:

- Sumber daya perusahaan: Ketika sumber daya dibatasi, strategi terkonsentrasi mungkin lebih efektif.
- Variabilitas produk: Untuk produk yang sangat seragam (seperti gula atau baja), pemasaran yang tidak terdiferensiasi mungkin lebih tepat. Untuk produk yang dapat dibedakan, (seperti mobil), maka pendekatan yang dibedakan atau terkonsentrasi diindikasikan.
- Siklus hidup produk: Untuk produk baru, satu versi dapat digunakan pada tahap peluncuran, tetapi ini dapat diperluas ke pendekatan yang lebih tersegmentasi dari waktu ke waktu. Dengan semakin banyaknya pesaing yang memasuki pasar, mungkin perlu dibedakan.
- Karakteristik pasar: Ketika semua pembeli memiliki selera yang sama, atau tidak bersedia membayar premium untuk kualitas yang berbeda, maka pemasaran yang tidak terdiferensiasi diindikasikan.
- Aktivitas kompetitif: Ketika pesaing menerapkan segmentasi pasar yang terdiferensiasi atau terkonsentrasi, menggunakan pemasaran yang tidak terdiferensiasi dapat berakibat fatal. Sebuah perusahaan harus mempertimbangkan apakah dapat menggunakan pendekatan segmentasi pasar yang berbeda.

Proses Segmentasi Pasar: S-T-P

Proses segmentasi pasar ternyata sederhana. Tujuh langkah dasar menjelaskan keseluruhan proses termasuk segmentasi, penargetan, dan pemosisian. Namun, dalam praktiknya, tugas tersebut bisa sangat melelahkan karena melibatkan penelitian terhadap banyak data, dan membutuhkan banyak keterampilan dalam analisis, interpretasi, dan beberapa penilaian. Meskipun banyak analisis perlu dilakukan, dan banyak keputusan perlu dibuat, pemasar cenderung menggunakan apa yang disebut proses S-T-P, yaitu Segmentation → Targeting→ Positioning, sebagai kerangka kerja yang luas untuk menyederhanakan proses dan diuraikan di sini:

Segmentasi pasar / segmentation

1. Mengidentifikasi pasar (juga dikenal sebagai alam semesta) yang akan disegmentasi.
2. Mengidentifikasi, pilih dan terapkan basis atau basis yang akan digunakan dalam segmentasi
3. Mengembangkan profil segmen

Menentukan target / Targeting

4. Mengevaluasi daya tarik setiap segmen
5. Memiilih segmen atau pasar yang akan ditargetkan

Memposisikan produk / Positioning

6. Mengidentifikasi posisi optimal untuk setiap segmen
7. Mengembangkan program pemasaran untuk setiap segmen

Dasar untuk Segmentasi Pasar Konsumen

Langkah utama dalam proses segmentasi adalah pemilihan alas yang sesuai. Pada langkah ini, pemasar mencari cara untuk mencapai homogenitas internal (kesamaan dalam segmen), dan heterogenitas eksternal (perbedaan antar segmen). Dengan kata lain, mereka mencari proses yang meminimalkan perbedaan antara anggota segmen dan memaksimalkan perbedaan antara setiap segmen. Selain itu, pendekatan segmentasi harus menghasilkan segmen yang bermakna untuk masalah atau situasi pemasaran tertentu. Misalnya, warna rambut seseorang mungkin menjadi dasar yang relevan untuk produsen sampo, tetapi tidak akan relevan untuk penjual jasa keuangan. Memilih basis yang tepat membutuhkan pemikiran yang matang dan pemahaman dasar tentang pasar yang akan disegmentasi.

Pada kenyataannya, pemasar dapat menyegmentasikan pasar menggunakan basis atau variabel apa pun asalkan dapat diidentifikasi, terukur, dapat ditindaklanjuti, dan stabil. Misalnya, beberapa rumah mode telah mengelompokkan pasar menggunakan ukuran pakaian wanita sebagai variabel. Namun, dasar yang paling umum untuk melakukan segmentasi pasar konsumen meliputi: geografi, demografi, psikografis, dan perilaku. Pemasar biasanya memilih satu basis untuk analisis segmentasi, meskipun, beberapa basis dapat digabungkan menjadi satu segmentasi dengan hati-hati. Misalnya, geografi dan demografi sering digabungkan, tetapi basis lain jarang digabungkan. Mengingat bahwa psikografik mencakup variabel demografis seperti usia, jenis kelamin dan pendapatan serta variabel sikap dan perilaku, tidak masuk akal untuk menggabungkan psikografis dengan demografi atau basis lain. Setiap upaya untuk menggunakan basis gabungan membutuhkan pertimbangan yang cermat dan landasan logis.

Bagian berikut ini memberikan penjelasan rinci tentang bentuk paling umum dari segmentasi pasar konsumen.

Segmentasi Geografis

Segmentasi geografis membagi pasar sesuai dengan kriteria geografis. Dalam praktiknya, pasar dapat disegmentasi seluas benua dan sesempit kawasan atau kode pos. Variabel geografis yang khas meliputi:

- Negara mis. Amerika Serikat, Inggris, Cina, Jepang, Korea Selatan, Malaysia, Singapura, Australia, Selandia Baru
- Wilayah mis. Utara, Barat laut, Barat tengah, Selatan, Tengah
- Kepadatan populasi: mis. kawasan pusat bisnis (CBD), perkotaan, pinggiran kota, pedesaan, regional
- Ukuran kota atau kota: mis. di bawah 1.000; 1.000 - 5.000; 5.000 - 10.000 ... 1.000.000 - 3.000.000 dan lebih dari 3.000.000
- Zona iklim: mis. Mediterania, Beriklim sedang, Sub-Tropis, Tropis, Kutub,

Pendekatan geo-cluster (juga disebut segmentasi geo-demografis) menggabungkan data demografis dengan data geografis untuk membuat profil yang lebih kaya dan lebih rinci. Model geo-cluster memecah suatu wilayah menjadi kelompok rumah tangga berdasarkan asumsi bahwa orang-orang di lingkungan tertentu cenderung menunjukkan kesamaan dalam komposisi demografis dan karakteristik rumah tangga lainnya.

Segmentasi geografis dapat dianggap sebagai langkah pertama dalam pemasaran internasional, di mana pemasar harus memutuskan apakah akan menyesuaikan produk yang ada dan program pemasaran untuk kebutuhan unik pasar geografis yang berbeda. Dewan Pemasaran Pariwisata sering mensegmentasi pengunjung internasional berdasarkan negara asalnya. Sebagai contoh, Tourism Australia melakukan pemasaran di 16 pasar geografis inti; di mana Cina, Inggris, AS, Selandia Baru, dan Jepang telah diidentifikasi sebagai segmen prioritas karena mereka memiliki potensi terbesar untuk pertumbuhan dan merupakan segmen yang sangat menguntungkan dengan pengeluaran yang lebih tinggi dari rata-rata per kunjungan. Tourism Australia melakukan penelitian ekstensif pada masing-masing segmen ini dan mengembangkan profil kaya segmen prioritas tinggi untuk lebih memahami kebutuhan mereka dan bagaimana mereka membuat keputusan perjalanan. Wawasan dari analisis ini digunakan dalam pengembangan produk perjalanan, alokasi anggaran promosi, strategi periklanan dan dalam keputusan perencanaan kota yang lebih luas. Misalnya, mengingat jumlah pengunjung Jepang, kota Melbourne telah membuat papan nama Jepang di kawasan wisata.

Sejumlah paket geo-demografis eksklusif tersedia untuk penggunaan komersial. Contohnya termasuk Acorn di Inggris, Segmen Mosaik (geodemografi) Experian (aktif di Amerika Utara, Amerika Selatan, Inggris, Eropa, Afrika Selatan dan sebagian Asia-Pasifik) atau Helix Personas (Australia, Selandia Baru, dan Indonesia). Perlu dicatat bahwa semua paket komersial ini menggabungkan geografi dengan data perilaku, demografis, dan sikap dan menghasilkan sejumlah segmen yang sangat besar. Misalnya, NZ Helix Personas mengelompokkan populasi Selandia Baru yang relatif kecil menjadi 51 profil kepribadian terpisah di tujuh komunitas geografis (terlalu banyak untuk diperinci di halaman ini). Database komersial ini biasanya memungkinkan calon klien akses ke demonstrasi skala terbatas, dan pembaca yang tertarik untuk mempelajari lebih lanjut tentang bagaimana segmentasi geografi-demografis dapat menguntungkan pemasar dan perencana, disarankan untuk bereksperimen dengan perangkat lunak demonstrasi online melalui situs web perusahaan-perusahaan ini.

Segmentasi geografis secara luas digunakan dalam kampanye pemasaran langsung untuk mengidentifikasi bidang-bidang yang berpotensi untuk penjualan pribadi, distribusi kotak surat atau surat langsung. Segmentasi Geocluster banyak digunakan oleh Pemerintah dan departemen sektor publik seperti perencanaan kota, otoritas kesehatan, polisi, departemen peradilan pidana, telekomunikasi dan organisasi utilitas publik seperti dewan air.

Segmentasi Demografis

Segmentasi menurut demografi didasarkan pada variabel demografis konsumen seperti usia, pendapatan, ukuran keluarga, status sosial-ekonomi, dll. Segmentasi demografis mengasumsikan bahwa konsumen dengan profil demografis yang sama akan menunjukkan pola pembelian, motivasi, minat, dan gaya hidup yang serupa dan bahwa ini karakteristik akan diterjemahkan ke dalam preferensi produk / merek serupa. Dalam praktiknya, segmentasi demografis berpotensi menggunakan variabel apa pun yang digunakan oleh pengumpul sensus negara. Variabel demografis yang khas dan deskriptornya adalah sebagai berikut:

- Usia: mis. Di bawah 5, 5-8 tahun, 9-12 tahun, 13-17 tahun, 18-24, 25-29, 30-39, 40-49, 50-59, 60+
- Jenis kelamin Laki-laki Perempuan
- Pekerjaan: Profesional, wiraswasta, semi-profesional, klerikal / admin, penjualan, perdagangan, pertambangan, produsen utama, mahasiswa, tugas rumah, pengangguran, pensiunan
- Kelas sosial (atau status sosial ekonomi):

- Status Perkawinan: Lajang, menikah, bercerai, janda
- Tahap Kehidupan Keluarga: Lajang muda; Muda menikah tanpa anak; Keluarga muda dengan anak di bawah 5 tahun; Lansia menikah dengan anak-anak; Lansia menikah tanpa anak yang tinggal di rumah, Lansia hidup sendiri
- Ukuran keluarga / jumlah tanggungan: 0, 1-2, 3-4, 5+
- Penghasilan: Di bawah $ 10.000; 10.000-20.000; 20.001-30.000; 30.001-40.000, 40.001- 50.000 dll
- Tingkat pendidikan: Sekolah dasar; Beberapa sekolah menengah, Sekolah menengah pertama, Beberapa universitas, Gelar; Pascasarjana atau lebih tinggi
- Kepemilikan rumah: Menyewa, Memiliki rumah dengan hipotek, Rumah yang dimiliki langsung
- Etnis: Asia, Afrika, Aborigin, Polinesia, Melanesia, Amerika Latin, Afrika-Amerika, Indian Amerika dll
- Agama: Katolik, Protestan, Muslim, Yahudi, Budha, Hindu, Lainnya

Gerakan Kepanduan/Pramuka menawarkan contoh yang sangat baik dari segmentasi demografis dalam praktiknya. Pramuka mengembangkan berbagai produk berdasarkan kelompok umur yang relevan. Kelompok umur adalah sebuah tingkatan dalam kepramukaan yang ditentukan oleh umur anggotanya. Kelompok dibagi menjadi 4 : Kelompok umur 7-10 tahun disebut dengan Pramuka Siaga Kelompok umur 11-15 tahun disebut dengan Pramuka Penggalang Kelompok umur 16-20 tahun disebut dengan Pramuka Penegak Kelompok umur 21 – 25 tahun disebut dengan Pramuka Pandega Ada juga Kelompok Khusus, yaitu Kelompok yang ditujukan untuk orang yang memiliki kedudukan dalam kepramukaan. Misalnya Pramuka Pembina, adalah sebutan untuk orang dewasa yang memimpin Pramuka.
Gerakan Kepanduan Pramuka memberikan seragam yang berbeda kepada setiap anggota kelompok dan mengembangkan program kegiatan yang berbeda untuk setiap segmen. Pramuka bahkan melayani kebutuhan lebih dari 25 yang menawarkan mereka peran sebagai pemimpin atau sukarelawan pramuka. Segmentasi Pramuka adalah contoh analisis segmentasi demografis sederhana yang hanya menggunakan satu variabel, yaitu usia.

Dalam praktiknya, sebagian besar segmentasi demografis menggunakan kombinasi variabel demografis. Misalnya, pendekatan segmentasi yang dikembangkan untuk Selandia Baru oleh Nielsen Research menggabungkan beberapa variabel demografis termasuk usia, tahap kehidupan, dan status sosial ekonomi. Produk segmentasi berpemilik, yang

dikenal sebagai geoTribes, membagi pasar NZ menjadi 15 suku, yaitu: Rockafellas - keluarga dewasa yang makmur; Keluarga muda paruh baya dan ambisius yang berambisi; Fortunats- Pensiunan yang aman secara finansial dan pra-pensiunan; Single dan pasangan yang berorientasi Karir-Karier; Preppies- Anak-anak dewasa dari orang tua yang kaya; Independen - Lajang dan pasangan muda; Suburban Splendour - keluarga dewasa kelas menengah; Twixters- Anak dewasa tinggal di rumah; Debstars-Keluarga muda yang diperpanjang secara finansial; Boomer - Pra-pensiunan kerah putih pos keluarga; True Blues - Keluarga dewasa kerah biru dan single atau pasangan pra-pensiunan; Struggleville - Berjuang keluarga muda dan setengah baya; Gray Power-Better off pensiunan; Korban-Pensiunan yang hidup dengan pendapatan minimal dan Orang-Orang Langsing yang hidup dalam keadaan kurang mampu.

Penggunaan beberapa variabel segmentasi biasanya memerlukan analisis database menggunakan teknik statistik canggih seperti analisis klaster atau analisis komponen utama. Perlu dicatat bahwa jenis analisis ini membutuhkan ukuran sampel yang sangat besar. Namun, pengumpulan data mahal untuk masing-masing perusahaan. Untuk alasan ini, banyak perusahaan membeli data dari perusahaan riset pasar komersial, banyak di antaranya mengembangkan perangkat lunak berpemilik untuk menginterogasi data. Paket eksklusif, seperti yang disebutkan dalam geoTribes sebelumnya oleh Nielsen misalnya, menawarkan klien akses ke database yang luas bersama dengan program yang memungkinkan pengguna untuk menginterogasi data melalui antarmuka ‘ramah pengguna’. Dengan kata lain, pengguna tidak memerlukan pemahaman terperinci tentang prosedur statistik ‘back-end’ yang digunakan untuk menganalisis data dan mendapatkan segmen pasar. Namun, pengguna masih harus terampil dalam menafsirkan temuan untuk digunakan dalam pengambilan keputusan pemasaran.

Segmentasi Psikografis

Segmentasi psikografis, yang kadang-kadang disebut segmentasi gaya hidup, diukur dengan mempelajari aktivitas, minat, dan pendapat (AIO) pelanggan. Ini mempertimbangkan bagaimana orang menghabiskan waktu luang mereka, dan pengaruh luar mana yang paling mereka responsif dan pengaruhi. Psikografi adalah dasar yang sangat banyak digunakan untuk segmentasi, karena memungkinkan pemasar untuk mengidentifikasi segmen pasar yang didefinisikan secara ketat dan lebih memahami motivasi untuk memilih produk tertentu.

Salah satu analisis segmentasi psikografis yang paling terkenal adalah yang disebut Values And Lifestyles Segments (VALS), metode psikometrik eksklusif yang mengukur sikap, perilaku, dan demografi yang selaras dengan preferensi merek dan adopsi produk baru. Pendekatan ini awalnya dikembangkan oleh SBI International pada tahun 1970-an, dan tipologi atau segmen telah mengalami beberapa revisi selama bertahun-tahun. Seperti saat ini berdiri, segmen VALs yang menggambarkan populasi orang dewasa Amerika adalah: Orang-orang yang berprestasi (26%), Strivers (24%), Experiencer (21%), Inovator, Pemikir, Percaya, Pencipta, dan Penyintas. Segmen VALs adalah spesifik negara dan pengembang menawarkan tipologi segmentasi VALs untuk digunakan di Cina, Jepang, Republik Dominika, Nigeria, Inggris, dan Venezuela.

Di luar Amerika Serikat, negara-negara lain telah mengembangkan merek mereka sendiri dari segmentasi psikografis. Di Australia dan Selandia Baru, Roy Morgan Research telah mengembangkan Segmen Nilai yang menggambarkan sepuluh segmen berdasarkan pola pikir, demografi, dan perilaku. Segmen Nilai adalah: Pencapai Visible (20%); Kehidupan Keluarga Tradisional (19%); Sadar Sosial (14%); Look-At-Me (11%); Kehidupan Keluarga Konvensional (10%); Sesuatu yang Lebih Baik (9%); Optimis Muda (7%); Kesepakatan yang Lebih Adil (5%); Konservatif Nyata (3%) dan Kebutuhan Dasar (2%) Perusahaan riset pasar, Nielsen menawarkan PALS (Personal Aspirational Lifestyle Segments), segmentasi berbasis nilai yang mengurutkan prioritas gaya hidup masa depan, seperti pentingnya karier, keluarga, ingin memiliki kesenangan, atau keinginan untuk gaya hidup seimbang. PALS membagi pasar Australia menjadi enam kelompok, yaitu Success Driven (25%); Pencari Saldo (23%); Sadar Kesehatan (22%), Pencari Harmoni (11%), Individualis (11%) dan Pencari Menyenangkan (8%).
Di Inggris, segmen-segmen berikut berdasarkan gaya hidup Inggris dikembangkan oleh McCann-Erickson: Avant-Guardians (tertarik pada perubahan); Pontifikator (tradisionalis); Bunglon (ikuti kerumunan) dan Sleepwalker (berprestasi kurang puas).

Sementara banyak dari analisis segmentasi psikografis eksklusif ini terkenal, sebagian besar studi berdasarkan psikografis dirancang khusus. Itu adalah segmen yang dikembangkan untuk produk individual pada waktu tertentu. Salah satu utas umum di antara studi segmentasi psikografis adalah bahwa mereka menggunakan nama unik untuk menggambarkan segmen.

Segmentasi Perilaku

Segmentasi perilaku membagi konsumen ke dalam kelompok sesuai dengan perilaku yang diamati. Banyak pemasar percaya bahwa variabel perilaku lebih unggul daripada demografi dan geografi untuk membangun segmen pasar. Variabel perilaku khas dan deskriptornya meliputi:

- Kesempatan Pembelian / Penggunaan: mis. acara reguler, acara khusus, acara meriah, pemberian hadiah
- Berusaha Manfaat: mis. ekonomi, kualitas, tingkat layanan, kenyamanan, akses
- Status Pengguna: mis. Pengguna pertama kali, Pengguna biasa, Bukan pengguna
- Tingkat Penggunaan / Frekuensi Pembelian: mis. Pengguna ringan, pengguna berat, pengguna sedang
- Status Loyalitas: mis. Loyal, switcher, tidak loyal, murtad
- Kesiapan Pembeli: mis. Tidak sadar, sadar, niat untuk membeli
- Sikap terhadap Produk atau Layanan: mis. Penggemar, Biasa saja, Bermusuhan; Sadar Harga, Sadar Berkualitas
- Status Adopter: mis. Pengadopsi awal, pengadopsi terlambat, lamban

Perhatikan bahwa deskriptor ini hanyalah contoh yang biasa digunakan. Pemasar harus memastikan bahwa mereka menyesuaikan variabel dan deskriptor untuk kondisi lokal dan untuk aplikasi tertentu. Sebagai contoh, dalam industri kesehatan, perencana sering membagi pasar yang luas sesuai dengan ‘kesadaran kesehatan’ dan mengidentifikasi segmen yang sadar rendah, sedang, dan sangat sehat. Ini adalah contoh segmentasi perilaku, menggunakan sikap terhadap produk atau layanan sebagai deskriptor atau variabel utama.

Segmentasi Kejadian Pembelian / Penggunaan

Segmentasi peristiwa pembelian atau penggunaan berfokus pada analisis kesempatan ketika konsumen mungkin membeli atau mengkonsumsi suatu produk. Ini mendekati model segmentasi tingkat pelanggan dan tingkat kejadian dan memberikan pemahaman tentang kebutuhan, perilaku, dan nilai pelanggan individu dalam berbagai kesempatan penggunaan dan waktu. Tidak seperti model segmentasi tradisional, pendekatan ini memberikan lebih dari satu segmen untuk setiap pelanggan unik, tergantung pada keadaan mereka saat ini.

Sebagai contoh, Cadbury membagi pasar menjadi lima segmen berdasarkan penggunaan dan perilaku:

Cadbury mengelompokkan konsumen dari berbagai cokelat mereka sesuai dengan kesempatan penggunaan

- Makan Langsung (34%): Didorong oleh kebutuhan untuk ngemil, memanjakan atau meningkatkan energi. Produk yang memenuhi kebutuhan ini termasuk merek seperti Kit-Kat, Mars Bars
- Stok Rumah (25%): Didorong oleh kebutuhan untuk memiliki sesuatu di dapur untuk dibagikan kepada keluarga di depan TV atau untuk ngemil di rumah. Produk yang memenuhi kebutuhan ini adalah balok atau banyak kemasan.
- Anak-anak (17%): Didorong oleh kebutuhan akan makanan ringan setelah sekolah, pesta, camilan. Produk yang memenuhi kebutuhan ini termasuk Smarties dan Milky Bars.
- Pemberian hadiah (15%): Produk yang dibeli sebagai hadiah, kebutuhan termasuk tanda penghargaan, gerakan romantis atau acara khusus. Produk yang memenuhi kebutuhan ini termasuk cokelat kotak seperti Cabury's Roses atau Quality Street
- Musiman (3,4%): Didorong oleh kebutuhan untuk memberikan hadiah atau menciptakan suasana meriah. Produk ini terutama dibeli pada hari-hari raya seperti Natal, Paskah, Advent. Produk yang memenuhi kebutuhan ini termasuk telur Paskah dan hiasan pohon Natal.

Segmentasi Pasar Lainnya

Segmentasi Pasar berdasarkan hubungan secara ekstrim

Merupakan bentuk efektif segmentasi bagi penggunaan merek, seperti:

- Tingkat penggunaan: beda segmentasi terletak pada pengguna berat, pengguna sedang, dan pengguna ringan. Bukan pengguna sebuah produk, jasa, atau merek khusus.
- Tingkat kesadaran: kesadaran konsumen pada produk, kesiapan membeli produk, atau apakah konsumen membutuhkan informasi tentang produk tersebut.
- Loyalitas merek: Loyalitas konsumen pada merek dijadikan perusahaan sebagai identifikasi karakteristik konsumen di mana mereka bisa langsung menjadi pendukung promosi ke orang dengan karakteristik yang sama namun dengan populasi yang lebih besar.

Segmentasi berdasarkan situasi penggunaan

Kesempatan atau situasi bisa menentukan apakah konsumen akan membeli atau mengkonsumsi. Segmentasi ini dibuat untuk membantu perusahaan memperluas penggunaan produk.

Segmentasi berdasarkan benefit

Bentuk segmentasi yang mengklasifikasikan pembeli sesuai dengan menfaat berbeda yang mereka cari dari produk merupakan bentuk segmentasi yang kuat. Sebuah studi yang melakukan pengujian apakah yang mengendalikan preferensi konsumen terhadap micro atau craftbeer, teridentifikasi lima keuntungan strategic brand, yaitu:

- Fungsional (contoh kualitas)
- Nilai uang
- Manfaat sosial
- Manfaat emosi positif
- Manfaat emosi negatif

Segmentasi hybrid

Segmen ini dibentuk berdasarkan kombinasi beberapa variabel segmen yang membentuk sebuah segmen tunggal. Sebagai contoh segmen geodemografis, sangat berguna untuk menemukan prospek terbaik bagi seorang pengiklan atau pemasar dalam menemukan kepribadian, tujuan, dan ketertarikan dan diisolasikan di mana mereka hidup.

Memilih Pasar Sasaran

Keputusan besar lain dalam mengembangkan strategi segmentasi adalah pemilihan segmen pasar yang akan menjadi fokus perhatian khusus (dikenal sebagai target pasar). Pemasar menghadapi sejumlah keputusan penting:

- Kriteria apa yang harus digunakan untuk mengevaluasi pasar?
- Berapa banyak pasar yang akan dimasuki (satu, dua atau lebih)?
- Segmen pasar mana yang paling berharga?

Ketika seorang pemasar memasuki lebih dari satu pasar, segmen tersebut sering diberi label pasar target primer, pasar target sekunder. Pasar primer adalah target pasar yang dipilih sebagai fokus utama kegiatan pemasaran. Target pasar sekunder kemungkinan adalah segmen yang tidak sebesar pasar primer, tetapi memiliki potensi pertumbuhan. Atau, kelompok sasaran sekunder dapat terdiri dari sejumlah kecil pembeli yang menyumbang proporsi volume penjualan yang relatif tinggi mungkin karena nilai pembelian atau frekuensi pembelian.

Dalam hal mengevaluasi pasar, tiga pertimbangan utama sangat penting:

- Ukuran dan pertumbuhan segmen
- Segmen daya tarik struktural
- Tujuan dan sumber daya perusahaan.

Tidak ada formula untuk mengevaluasi daya tarik segmen pasar dan penilaian yang baik harus dilakukan. Namun demikian, sejumlah pertimbangan dapat digunakan untuk mengevaluasi segmen pasar untuk daya tarik.

Ukuran dan Pertumbuhan Segmen:

- Seberapa besar ceruk pasar yang ada?
- Apakah segmen pasar cukup substansial untuk menguntungkan?
- Ukuran segmen dapat diukur dalam jumlah pelanggan, tetapi langkah-langkah superior cenderung mencakup nilai atau volume penjualan
- Apakah segmen pasar tumbuh atau menyusut?
- Apa indikasi bahwa pertumbuhan akan berkelanjutan dalam jangka panjang? Apakah pertumbuhan yang diamati berkelanjutan?
- Apakah segmen stabil dari waktu ke waktu? (Segmen harus memiliki waktu yang cukup untuk mencapai tingkat kinerja yang diinginkan)

Daya Tarik Struktural Segmen:

- Sejauh mana pesaing menargetkan segmen pasar ini?
- Bisakah kita mengukir posisi yang layak untuk dibedakan dari pesaing?
- Seberapa responsifkah anggota segmen pasar terhadap program pemasaran?
- Apakah segmen pasar ini dapat dijangkau dan diakses? (mis., berkenaan dengan distribusi dan promosi)

Tujuan dan Sumber Daya Perusahaan:

- Apakah segmen pasar ini selaras dengan filosofi operasi perusahaan kami?
- Apakah kita memiliki sumber daya yang diperlukan untuk memasuki segmen pasar ini?
- Apakah kita memiliki pengalaman sebelumnya dengan segmen pasar ini atau segmen pasar serupa?
- Apakah kita memiliki keterampilan dan / atau pengetahuan untuk memasuki segmen pasar ini dengan sukses?

Segmentasi Pasar dan Program Pemasaran

Ketika segmen telah ditentukan dan tawaran terpisah dikembangkan untuk masing-masing segmen inti, tugas pemasar berikutnya adalah merancang program pemasaran (juga dikenal sebagai bauran pemasaran) yang akan beresonansi dengan target pasar atau pasar. Mengembangkan program pemasaran membutuhkan pengetahuan mendalam

tentang kebiasaan pembelian segmen pasar utama, outlet ritel pilihan mereka, kebiasaan media mereka dan sensitivitas harga mereka. Program pemasaran untuk setiap merek atau produk harus didasarkan pada pemahaman tentang target pasar (atau target pasar) yang terungkap dalam profil pasar.

Tabel 3 memberikan ringkasan singkat dari bauran pemasaran yang digunakan oleh Black and Decker, produsen tiga merek alat-alat listrik dan peralatan rumah tangga. Setiap merek ditempatkan pada titik kualitas yang berbeda dan menargetkan segmen pasar yang berbeda. Tabel ini dirancang untuk menggambarkan bagaimana program pemasaran harus disesuaikan untuk setiap segmen pasar.

Segmen	Produk / Merek	Strategi Produk	Strategi Harga	Strategi Promosi	Strategi Tempat
Rumahan/ DIY	Black & Decker	Kualitas Memadai untuk digunakan sesekali	Harga lebih rendah	Iklan TV selama liburan	Distribusi massal (toko tingkat bawah)
Weekend Warior	Firestorn	Kualitas Memadai untuk penggunaan biasa	Harga lebih tinggi	Iklan di majalah D-I-Y, road show	Distribusi selektif (toko perangkat keras tingkat atas)
Kaum profesional	De Walt	Kualitas memadai untuk pemakaian sehari-hari	Harga tertinggi	Penjualan pribadi (perwakilan penjualan menelepon di lokasi kerja)	Distribusi selektif (toko perangkat keras tingkat atas)

Basis untuk Segmentasi Pasar Bisnis

Bisnis dapat dikelompokkan menurut industri, ukuran bisnis, omset, jumlah karyawan atau variabel lain yang relevan.

Firmografi

Firmografi (juga dikenal sebagai segmentasi berbasis fitur atau emporografi) adalah jawaban komunitas bisnis terhadap segmentasi demografis. Ini biasanya digunakan di pasar bisnis-ke-bisnis (diperkirakan 81% pemasar B2B menggunakan teknik ini). Di bawah pendekatan ini, target pasar tersegmentasi berdasarkan fitur seperti ukuran perusahaan (baik dalam hal pendapatan atau jumlah karyawan), sektor industri atau lokasi (negara dan / atau wilayah).

Segmentasi Akun multi-variabel

Dalam Manajemen Wilayah Penjualan, menggunakan lebih dari satu kriteria untuk mengkarakterisasi akun organisasi, seperti mengelompokkan akun penjualan oleh pemerintah, bisnis, pelanggan, dll. Dan ukuran atau durasi akun, dalam upaya meningkatkan efisiensi waktu dan volume penjualan.

Menggunakan Segmentasi dalam Retensi Pelanggan

Pendekatan dasar untuk segmentasi berbasis retensi adalah bahwa perusahaan menandai setiap pelanggan aktifnya dengan empat nilai:

Apakah pelanggan ini berisiko tinggi membatalkan layanan perusahaan?

Salah satu indikator paling umum dari pelanggan berisiko tinggi adalah penurunan dalam penggunaan layanan perusahaan. Misalnya, dalam industri kartu kredit, ini dapat ditandai melalui penurunan pembelanjaan pelanggan untuk kartunya.

Apakah pelanggan ini berisiko tinggi beralih ke pesaing untuk membeli produk?

Banyak kali pelanggan memindahkan preferensi pembelian ke merek pesaing. Hal ini dapat terjadi karena berbagai alasan yang sulit untuk diukur. Sering kali bermanfaat bagi perusahaan terdahulu untuk mendapatkan wawasan yang bermakna, melalui analisis data, mengapa perubahan preferensi ini terjadi. Wawasan semacam itu dapat mengarah pada strategi yang efektif untuk memenangkan kembali pelanggan atau bagaimana tidak kehilangan target pelanggan di tempat pertama.

Apakah pelanggan ini layak dipertahankan?

Penentuan ini bermuara pada apakah laba pasca-retensi yang dihasilkan dari pelanggan diperkirakan lebih besar dari biaya yang dikeluarkan untuk mempertahankan pelanggan, dan termasuk evaluasi siklus hidup pelanggan.

Taktik retensi apa yang harus digunakan untuk mempertahankan pelanggan ini?

Analisis siklus hidup pelanggan ini biasanya dimasukkan dalam rencana pertumbuhan bisnis untuk menentukan taktik mana yang harus diterapkan untuk mempertahankan atau melepaskan pelanggan. Taktik yang biasa digunakan berkisar dari memberikan diskon pelanggan khusus hingga mengirim komunikasi pelanggan yang memperkuat proposisi nilai layanan yang diberikan.

Segmentasi: Algoritma dan Pendekatan

Pilihan metode statistik yang sesuai untuk segmentasi, tergantung pada sejumlah faktor termasuk, pendekatan luas (a-priori atau post-hoc), ketersediaan data, kendala waktu, tingkat keterampilan dan sumber daya pemasar.

Segmentasi A-priori

Menurut Asosiasi Riset Pasar (*Market Research Association*-MRA), penelitian apriori terjadi ketika "kerangka teori dikembangkan sebelum penelitian dilakukan". Dengan kata lain, pemasar memiliki gagasan tentang apakah melakukan segmentasi pasar secara geografis, demografis, psikografis, atau perilaku sebelum melakukan penelitian apa pun. Misalnya, seorang pemasar mungkin ingin mempelajari lebih lanjut tentang motivasi dan demografi pengguna yang ringan dan sedang dalam upaya memahami taktik apa yang dapat digunakan untuk meningkatkan tingkat penggunaan. Dalam hal ini, variabel target diketahui - pemasar telah tersegmentasi menggunakan variabel perilaku - status pengguna. Langkah selanjutnya adalah mengumpulkan dan menganalisis data sikap untuk pengguna yang ringan dan sedang. Analisis khas meliputi tabulasi silang sederhana, distribusi frekuensi dan kadang-kadang regresi logistik atau analisis CHAID.

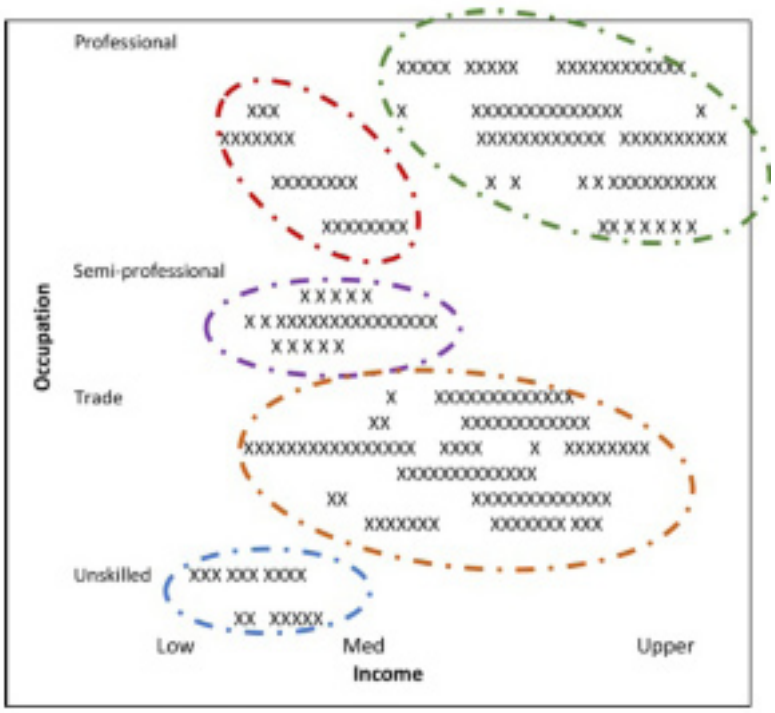
Kerugian utama dari segmentasi a-priori adalah tidak mengeksplorasi peluang lain untuk mengidentifikasi segmen pasar yang bisa lebih bermakna.

Segmentasi Post-hoc

Sebaliknya, segmentasi post-hoc tidak membuat asumsi tentang kerangka kerja teoritis yang optimal. Sebaliknya, peran analis adalah untuk menentukan segmen yang paling berarti untuk masalah atau situasi pemasaran tertentu. Dalam pendekatan ini, data empiris mendorong pemilihan segmentasi. Analis biasanya menggunakan beberapa jenis analisis pengelompokan atau pemodelan persamaan struktural untuk mengidentifikasi segmen dalam data. Gambar di samping menggambarkan bagaimana segmen dapat dibentuk menggunakan pengelompokan, namun perhatikan bahwa diagram ini hanya menggunakan dua variabel, sedangkan dalam praktiknya pengelompokan mempekerjakan sejumlah besar variabel. Segmentasi pasca-hoc bergantung pada akses ke set data yang kaya, biasanya dengan jumlah kasus yang sangat besar.

Teknik Statistik Untuk Segmentasi

Pemasar sering melibatkan perusahaan riset komersial atau konsultan untuk melakukan analisis segmentasi, terutama jika mereka tidak memiliki keterampilan statistik untuk melakukan analisis. Beberapa segmentasi, terutama analisis post-hoc, bergantung pada analisis statistik yang canggih.



Visualisasi segmen *pasar yang dibentuk menggunakan metode clustering.*

Teknik statistik umum untuk analisis segmentasi meliputi:

- Algoritme pengelompokan seperti K-means atau analisis Klaster lainnya
- Analisis Konjoin
- Regresi Logistik (juga dikenal sebagai Regresi Logit)
- CHAID Detektor Interaksi Otomatis Chi-square; sejenis pohon keputusan
- Structural Equation Modeling (SEM)
- Multidimensional scaling and Canonical Analysis
- Model campuran statistik seperti Latent Class Analysis
- Pendekatan ensemble seperti Random Forests
- Algoritme lain seperti Neural Networks

Sumber Data digunakan untuk Segmentasi

Pemasar menggunakan berbagai sumber data untuk studi segmentasi dan profil pasar. Sumber informasi yang umum termasuk:

Database Internal

- Catatan transaksi pelanggan mis. nilai jual per transaksi, frekuensi pembelian
- Catatan keanggotaan pelindung, mis. anggota aktif, anggota lama, lama keanggotaan
- Database manajemen hubungan pelanggan (CRM)
- Survei internal

Sumber Eksternal

- Survei Komersial atau studi pelacakan (tersedia dari perusahaan riset besar seperti Nielsen dan Roy Morgan)
- Instansi pemerintah atau asosiasi Profesional / Industri
- Data sensus
- Perilaku pembelian yang diamati - dikumpulkan dari agen online seperti Google
- Teknik penambangan data

5.3 Tren Pasar

Tren pasar adalah kecenderungan yang dirasakan pasar keuangan untuk bergerak ke arah tertentu dari waktu ke waktu. Tren ini diklasifikasikan sebagai kerangka waktu yang lama, primer untuk kerangka waktu menengah, dan sekunder untuk kerangka waktu pendek. Trader berusaha mengidentifikasi tren pasar menggunakan analisis teknis, suatu kerangka kerja yang mencirikan tren pasar sebagai kecenderungan harga yang dapat diprediksi di dalam pasar ketika harga mencapai level support dan resistance, bervariasi dari waktu ke waktu.

Tren hanya dapat ditentukan di belakang, karena kapan pun harga di masa depan tidak diketahui.

Nomenklatur Pasar

Istilah “bull market” dan “bear market” masing-masing menggambarkan tren pasar naik dan turun, dan dapat digunakan untuk menggambarkan pasar sebagai keseluruhan atau sektor tertentu dan sekuritas. Nama-nama itu mungkin sesuai dengan fakta bahwa

seekor banteng menyerang dengan mengangkat tanduknya ke atas, sementara beruang menyerang dengan cakarnya dalam gerakan ke bawah.

Definisi Menurut Bahasa

Gaya bertarung kedua hewan mungkin memiliki dampak besar pada nama. Satu etimologi hipotetis menunjuk ke London "pekerja kulit" (pembuat pasar) London, yang akan menjual kulit beruang sebelum beruang benar-benar terperangkap dalam kontradiksi pepatah *ne vendez pas la peau de l'our avant de l'avoir tué* ("don ' "Jangan menjual kulit beruang sebelum Anda membunuh beruang") - sebuah peringatan terhadap optimisme berlebihan. Pada saat peristiwan yang terkenal *South Sea Bubble* tahun 1721, beruang itu juga dikaitkan dengan penjualan pendek; pekerja akan menjual kulit beruang yang tidak mereka miliki untuk mengantisipasi jatuhnya harga, yang akan memungkinkan mereka membelinya nanti untuk mendapatkan keuntungan tambahan.

Beberapa analogi yang telah digunakan sebagai perangkat mnemonik:

- Bull adalah kependekan dari "bully", dalam arti "luar biasa" yang sekarang agak kuno.
- Berkaitan dengan kecepatan hewan: Bulls biasanya menyerang dengan kecepatan yang sangat tinggi, sedangkan beruang biasanya dianggap sebagai penggerak yang malas dan berhati-hati — kesalahpahaman, karena beruang, dalam kondisi yang tepat, dapat berlari lebih cepat dari seekor kuda.
- Mereka pada awalnya digunakan mengacu pada dua keluarga perbankan pedagang lama, Barings dan Bulstrodes.
- Kata "bull" memainkan pengembalian pasar menjadi "penuh", sedangkan "bear" menyinggung pengembalian pasar menjadi "telanjang".
- "Bull" melambangkan pengisian ke depan dengan keyakinan berlebihan, sedangkan "beruang" melambangkan mempersiapkan diri untuk musim dingin dan hibernasi dalam keraguan.

Tren Sekuler

Tren pasar sekuler adalah tren jangka panjang yang berlangsung 5 hingga 25 tahun dan terdiri dari serangkaian tren primer. Pasar beruang sekuler terdiri dari pasar bull yang lebih kecil dan pasar bear yang lebih besar; pasar bull sekuler terdiri dari pasar bull yang lebih besar dan pasar bear yang lebih kecil.

Dalam pasar bullish sekuler tren yang berlaku adalah “bullish” atau bergerak ke atas. Pasar saham Amerika Serikat digambarkan berada di pasar bullish sekuler dari sekitar tahun 1983 hingga 2000 (atau 2007), dengan gangguan singkat termasuk jatuhnya tahun 1987 dan jatuhnya pasar tahun 2000-2002 yang dipicu oleh gelembung dot-com.

Di pasar beruang sekuler, tren yang berlaku adalah “bearish” atau bergerak ke bawah. Contoh pasar beruang sekuler terjadi dalam emas antara Januari 1980 hingga Juni 1999, yang berpuncak dengan Brown Bottom. Selama periode ini, harga emas nominal turun dari tertinggi $ 850 / oz ($ 30 / g) ke level terendah $ 253 / oz ($ 9 / g), dan menjadi bagian dari Depresi Komoditas Hebat.

Tren Utama

Patung dua *binatang simbolis keuangan, beruang dan banteng, di depan Bursa Efek Frankfurt.*

Tren utama memiliki dukungan luas di seluruh pasar (sebagian besar sektor) dan bertahan selama satu tahun atau lebih.

Pasar Banteng

Pasar banteng (*Bull Market*) adalah periode kenaikan harga secara umum. Awal mula pasar bullish ditandai oleh pesimisme yang meluas. Poin ini adalah ketika “kerumunan” adalah yang paling “bearish”. Perasaan putus asa berubah menjadi harapan, “optimisme”, dan akhirnya euforia, saat banteng menjalankan perjalanannya. Ini sering memimpin siklus ekonomi, misalnya dalam resesi penuh, atau sebelumnya.

Sebuah kartun *tahun 1901 yang menggambarkan pemodal J. P. Morgan sebagai banteng* dengan investor *yang bersemangat*

Analisis data pasar saham Morningstar, Inc. dari tahun 1926 hingga 2014 menemukan bahwa pasar banteng tipikal "bertahan 8,5 tahun dengan pengembalian total kumulatif rata-rata 458%", sementara keuntungan tahunan untuk pasar banteng berkisar antara 14,9% hingga 34,1%.

Contohnya :

India Stock Exchange Index India, BSE SENSEX, berada dalam tren pasar bullish selama sekitar lima tahun dari April 2003 hingga Januari 2008 karena meningkat dari 2.900 poin menjadi 21.000 poin. Pasar banteng terkemuka menandai periode 1925-1929, 1953–1957 dan periode 1993-1997 ketika AS dan banyak pasar saham lainnya naik; sementara periode pertama berakhir dengan tiba-tiba dengan dimulainya Depresi Hebat, akhir periode waktu berikutnya sebagian besar periode soft landing, yang menjadi pasar beruang besar.

Pasar Beruang

Bear market (Pasar beruang) adalah penurunan umum di pasar saham selama periode waktu tertentu. Ini adalah transisi dari optimisme investor tinggi ke ketakutan dan pesimisme investor yang meluas. Menurut The Vanguard Group, "Meskipun tidak ada definisi yang disepakati tentang pasar beruang, satu ukuran yang diterima secara umum adalah penurunan harga 20% atau lebih selama setidaknya periode dua bulan."

Analisis data pasar saham Morningstar, Inc. dari tahun 1926 hingga 2014 menemukan bahwa pasar beruang biasa "bertahan 1,3 tahun dengan rata-rata kerugian kumulatif

-41%", sementara penurunan tahunan untuk pasar beruang berkisar antara -19,7% hingga -47% .

Contohnya :

Pasar beruang mengikuti Wall Street Crash tahun 1929 dan menghapus 89% (dari 386 hingga 40) dari kapitalisasi pasar Dow Jones Industrial Average pada Juli 1932, menandai dimulainya Depresi Hebat. Setelah mendapatkan kembali hampir 50% dari kerugiannya, pasar beruang yang lebih lama dari tahun 1937 hingga 1942 terjadi di mana pasar kembali dipotong setengah. Pasar beruang jangka panjang lainnya terjadi dari sekitar tahun 1973 hingga 1982, meliputi krisis energi tahun 1970-an dan pengangguran yang tinggi pada awal 1980-an. Namun pasar beruang lain terjadi antara Maret 2000 Oktober 2002. Contoh terbaru terjadi antara Oktober 2007 dan Maret 2009, sebagai akibat dari krisis keuangan 2007-08.

Market Top

Market top (atau market high) biasanya bukan peristiwa dramatis. Pasar telah mencapai titik tertinggi, selama beberapa waktu (biasanya beberapa tahun). Ini didefinisikan surut karena pelaku pasar tidak menyadarinya saat itu terjadi. Penurunan kemudian mengikuti, biasanya secara bertahap pada awalnya dan kemudian dengan lebih cepat. William J. O'Neil dan perusahaan melaporkan bahwa sejak 1950-an suatu pasar teratas dicirikan oleh tiga hingga lima hari distribusi dalam indeks pasar utama yang terjadi dalam periode waktu yang relatif singkat. Distribusi adalah penurunan harga dengan volume lebih tinggi dari sesi sebelumnya.

Contohnya

Puncak gelembung dot-com (yang diukur oleh NASDAQ-100) terjadi pada 24 Maret 2000. Indeks ditutup pada 4,704.73. Nasdaq memuncak pada 5.132,50 dan S&P 500 pada 1525.20.

Market Bottom

Market bottom adalah pembalikan tren, akhir dari penurunan pasar, dan awal dari tren bergerak ke atas (bull market).

Sangat sulit untuk mengidentifikasi bottom (disebut oleh investor sebagai "bottom picking") ketika sedang terjadi. Kenaikan setelah penurunan seringkali berumur pendek dan harga

mungkin melanjutkan penurunannya. Ini akan membawa kerugian bagi investor yang membeli saham selama pasar yang salah dipahami atau "salah".

Baron Rothschild dikatakan telah menyarankan bahwa waktu terbaik untuk membeli adalah ketika ada "darah di jalanan", yaitu, ketika pasar telah jatuh secara drastis dan sentimen investor sangat negatif.

Contohnya :

Beberapa contoh dasar pasar, dalam hal nilai penutupan Dow Jones Industrial Average (DJIA) meliputi:

- Dow Jones Industrial Average mencapai titik terendah di 1738,74 pada 19 Oktober 1987, sebagai akibat dari penurunan dari 2722,41 pada 25 Agustus 1987. Hari ini disebut Black Monday (grafik).
- Nilai terendah 7286.27 dicapai pada DJIA pada tanggal 9 Oktober 2002 sebagai hasil dari penurunan dari 11722,98 pada tanggal 14 Januari 2000. Ini termasuk jumlah perantara antara 8235,81 pada 21 September 2001 (perubahan 14% dari 10 September) yang menyebabkan top intermediate dari 10635.25 pada 19 Maret 2002 (grafik). Nasdaq "tech-heavy" turun 79% dari puncak 5132 (10 Maret 2000) menjadi 1108 bawahnya (10 Oktober 2002).
- Nilai terbawah 6.440.08 (DJIA) pada 9 Maret 2009 tercapai setelah penurunan terkait dengan krisis subprime mortgage mulai 14164.41 pada 9 Oktober 2007 (grafik).

Tren Sekunder

Tren sekunder adalah perubahan jangka pendek dalam arah harga dalam tren primer. Durasi adalah beberapa minggu atau beberapa bulan.

Salah satu jenis tren pasar sekunder disebut koreksi pasar. Koreksi adalah penurunan harga jangka pendek dari 5% hingga 20% atau lebih. Sebuah contoh terjadi dari April hingga Juni 2010, ketika S&P 500 naik dari di atas 1.200 menjadi hampir 1000; ini dielu-elukan sebagai akhir dari pasar bullish dan mulai dari pasar beruang, tetapi tidak, dan pasar kembali naik. Koreksi adalah gerakan ke bawah yang tidak cukup besar untuk menjadi pasar beruang (ex post).

Jenis lain dari tren sekunder disebut reli pasar beruang (kadang-kadang disebut "pengisap reli" atau "bouncing kucing mati") yang terdiri dari kenaikan harga pasar hanya 10% atau

20% dan kemudian tren pasar beruang yang berlaku berlanjut. Reli pasar beruang terjadi di indeks Dow Jones setelah jatuhnya pasar saham 1929 yang mengarah ke bawah pasar pada tahun 1932, dan sepanjang akhir 1960-an dan awal 1970-an. Nikkei 225 Jepang telah ditandai oleh sejumlah pasar beruang
reli sejak akhir 1980-an sambil mengalami tren penurunan jangka panjang secara keseluruhan.

Pasar Australia pada awal 2015 telah digambarkan sebagai “pasar meerkat”, yang malu-malu dengan sentimen konsumen dan bisnis yang rendah.

Penyebab

Harga aset seperti saham ditentukan oleh penawaran dan permintaan. Menurut definisi, pasar menyeimbangkan pembeli dan penjual, sehingga tidak mungkin memiliki ‘pembeli lebih banyak dari penjual’ atau sebaliknya, meskipun itu adalah ungkapan yang umum. Untuk lonjakan permintaan, pembeli akan menaikkan harga yang bersedia mereka bayar, sementara penjual akan menaikkan harga yang ingin mereka terima. Untuk lonjakan pasokan, yang terjadi adalah sebaliknya.

Penawaran dan permintaan dibuat ketika investor menggeser alokasi investasi di antara jenis aset. Sebagai contoh, pada suatu waktu, investor dapat memindahkan uang dari obligasi pemerintah ke saham “teknologi”; di lain waktu, mereka dapat memindahkan uang dari saham “teknologi” ke obligasi pemerintah. Dalam setiap kasus, ini akan mempengaruhi harga kedua jenis aset.

Umumnya, investor mencoba mengikuti strategi beli-rendah, jual-tinggi tetapi sering keliru akhirnya membeli tinggi dan menjual rendah. Investor dan pedagang pelawan berusaha untuk “memudar” tindakan investor (beli saat mereka menjual, jual saat mereka membeli). Waktu ketika sebagian besar investor menjual saham dikenal sebagai distribusi, sementara waktu ketika sebagian besar investor membeli saham dikenal sebagai akumulasi.

Menurut teori standar, penurunan harga akan menghasilkan lebih sedikit penawaran dan lebih banyak permintaan, sedangkan kenaikan harga akan melakukan sebaliknya. Ini bekerja dengan baik untuk sebagian besar aset tetapi sering bekerja terbalik untuk saham karena kesalahan banyak investor membuat membeli tinggi dalam keadaan euforia dan menjual rendah dalam keadaan ketakutan atau panik akibat naluri menggiring. Jika kenaikan harga menyebabkan peningkatan permintaan, atau penurunan harga

menyebabkan peningkatan pasokan, ini menghancurkan loop umpan balik negatif yang diharapkan dan harga akan tidak stabil. Ini bisa dilihat dalam gelembung atau crash.

Sentimen Investor

Sentimen investor adalah indikator pasar saham pelawan. Ketika sebagian besar investor menyatakan sentimen bearish (negatif), beberapa analis menganggapnya sebagai sinyal kuat bahwa pasar bawah mungkin dekat. Kemampuan prediksi sinyal semacam itu dianggap tertinggi ketika sentimen investor mencapai nilai ekstrem. Indikator yang mengukur sentimen investor dapat meliputi:

David Hirshleifer melihat dalam fenomena tren sebuah jalan dimulai dengan reaksi rendah dan berakhir dengan reaksi berlebihan oleh investor / pedagang.

- Indeks Sentimen Kecerdasan Investor: Jika spread Bull-Bear (% Bulls -% Bears) dekat dengan terendah historis, ini mungkin menandakan bottom. Biasanya, jumlah beruang yang disurvei akan melebihi jumlah sapi jantan. Namun, jika jumlah sapi jantan berada pada tingkat yang sangat tinggi dan jumlah beruang berada pada tingkat yang sangat rendah, secara historis, puncak pasar mungkin telah terjadi atau hampir terjadi. Ukuran pelawan ini lebih dapat diandalkan untuk waktu yang kebetulan di posisi terendah pasar daripada puncak.
- Indikator sentimen American Association of Individual Investors (AAII): Banyak yang merasa bahwa sebagian besar penurunan sudah terjadi begitu indikator ini memberikan pembacaan minus 15% atau di bawahnya.
- Indikator sentimen lainnya termasuk rasio Nova-Ursa, Bunga Pendek / Total Market Float, dan rasio put / call.

5.4 Analisis SWOT

Analisis SWOT (sebagai alternatif, matriks SWOT) adalah singkatan dari kekuatan, kelemahan, peluang, dan ancaman dan merupakan metode perencanaan terstruktur yang mengevaluasi keempat elemen proyek atau usaha bisnis. Analisis SWOT dapat dilakukan untuk perusahaan, produk, tempat, industri, atau orang. Ini melibatkan menentukan tujuan dari usaha bisnis atau proyek dan mengidentifikasi faktor-faktor internal dan eksternal yang menguntungkan dan tidak menguntungkan untuk mencapai tujuan itu. Beberapa penulis memuji SWOT untuk Albert Humphrey, yang memimpin konvensi di Stanford Research Institute (sekarang SRI International) pada 1960-an dan 1970-an menggunakan data

dari perusahaan-perusahaan Fortune 500. Namun, Humphrey sendiri tidak mengklaim penciptaan SWOT, dan asal-usulnya tetap tidak jelas. Sejauh mana lingkungan internal perusahaan sesuai dengan lingkungan eksternal diungkapkan oleh konsep kecocokan strategis.

- Kekuatan/*Strenghts*: karakteristik bisnis atau proyek yang memberikan keunggulan dibandingkan yang lain
- Kelemahan/*Weakness*: karakteristik yang menempatkan bisnis atau proyek pada posisi yang kurang menguntungkan dibandingkan dengan yang lain
- Peluang/*Opportunities*: elemen-elemen di lingkungan yang dapat dieksploitasi oleh bisnis atau proyek untuk keuntungannya
- Ancaman/*Threats*: elemen-elemen di lingkungan yang dapat menyebabkan masalah

Analisis SWOT, *dengan empat elemennya dalam matriks 2 × 2.*

Identifikasi SWOT penting karena mereka dapat menginformasikan langkah-langkah selanjutnya dalam perencanaan untuk mencapai tujuan. Pertama, pembuat keputusan harus mempertimbangkan apakah tujuan tersebut dapat dicapai, mengingat SWOT. Jika tujuannya tidak dapat dicapai, mereka harus memilih tujuan yang berbeda dan ulangi prosesnya.

Pengguna analisis SWOT harus mengajukan dan menjawab pertanyaan yang menghasilkan informasi yang bermakna untuk setiap kategori (kekuatan, kelemahan, peluang, dan ancaman) untuk membuat analisis ini berguna dan menemukan keunggulan kompetitif mereka.

Faktor Internal dan Eksternal

Jadi dikatakan bahwa jika Anda tahu musuh Anda dan mengenal diri sendiri, Anda bisa memenangkan seratus pertempuran tanpa kehilangan satu pun. Jika Anda hanya mengenal diri sendiri, tetapi bukan lawan Anda, Anda mungkin menang atau kalah. Jika Anda tidak mengenal diri sendiri atau musuh Anda, Anda akan selalu membahayakan diri sendiri.

Seni Perang oleh Sun Tzu

Analisis SWOT bertujuan untuk mengidentifikasi faktor-faktor kunci internal dan eksternal yang dipandang penting untuk mencapai tujuan. Analisis SWOT mengelompokkan informasi penting ke dalam dua kategori utama:

1. Faktor internal - kekuatan dan kelemahan internal organisasi
2. Faktor eksternal - peluang dan ancaman yang disajikan oleh lingkungan di luar organisasi

Analisis dapat melihat faktor internal sebagai kekuatan atau kelemahan tergantung pada pengaruhnya pada tujuan organisasi. Apa yang mungkin mewakili kekuatan sehubungan dengan satu tujuan mungkin kelemahan (gangguan, persaingan) untuk tujuan lain. Faktor-faktor dapat mencakup semua 4P serta personil, keuangan, kemampuan manufaktur, dan sebagainya.

Faktor-faktor eksternal dapat mencakup masalah ekonomi makro, perubahan teknologi, perundang-undangan, dan perubahan sosial budaya, serta perubahan di pasar atau dalam posisi kompetitif. Hasilnya sering disajikan dalam bentuk matriks.

Analisis SWOT hanyalah salah satu metode kategorisasi dan memiliki kelemahannya sendiri. Sebagai contoh, ia mungkin cenderung membujuk para penggunanya untuk menyusun daftar daripada memikirkan faktor-faktor penting aktual dalam mencapai tujuan. Ini juga menyajikan daftar yang dihasilkan tanpa kritik dan tanpa prioritas yang jelas sehingga, misalnya, peluang yang lemah dapat menyeimbangkan ancaman yang kuat.

Adalah bijaksana untuk tidak menghilangkan kandidat SWOT yang masuk terlalu cepat. Pentingnya SWOT individu akan terungkap oleh nilai strategi yang mereka hasilkan. Item SWOT yang menghasilkan strategi berharga adalah penting. Item SWOT yang tidak menghasilkan strategi tidak penting.

Penggunaan

Kegunaan analisis SWOT tidak terbatas pada organisasi pencari keuntungan. Analisis SWOT dapat digunakan dalam situasi pengambilan keputusan ketika keadaan akhir yang diinginkan (objektif) didefinisikan. Contohnya termasuk organisasi nirlaba, unit pemerintah, dan individu. Analisis SWOT juga dapat digunakan dalam perencanaan pra-krisis dan manajemen krisis preventif. Analisis SWOT juga dapat digunakan dalam membuat rekomendasi selama studi kelayakan / survei.

Pembangunan Strategi

Analisis SWOT dapat digunakan secara efektif untuk membangun strategi organisasi atau pribadi. Langkah-langkah yang diperlukan untuk melaksanakan analisis berorientasi strategi melibatkan identifikasi faktor internal dan eksternal (menggunakan matriks 2x2 populer), pemilihan dan evaluasi faktor yang paling penting, dan identifikasi hubungan yang ada antara fitur internal dan eksternal.

Misalnya, hubungan yang kuat antara kekuatan dan peluang dapat menyarankan kondisi yang baik di perusahaan dan memungkinkan menggunakan strategi yang agresif. Di sisi lain, interaksi yang kuat antara kelemahan dan ancaman dapat dianalisis sebagai peringatan potensial dan saran untuk menggunakan strategi defensif. Analisis hubungan ini untuk menentukan strategi mana yang harus diimplementasikan sering dilakukan dalam fase perencanaan pertumbuhan untuk bisnis.

Pencocokan dan Konversi

Salah satu cara memanfaatkan SWOT adalah mencocokkan dan mengonversi. Pencocokan digunakan untuk menemukan keunggulan kompetitif dengan mencocokkan kekuatan dengan peluang. Taktik lain adalah mengubah kelemahan atau ancaman menjadi kekuatan atau peluang. Contoh strategi konversi adalah menemukan pasar baru. Jika ancaman atau kelemahan tidak dapat dikonversi, perusahaan harus berusaha meminimalkan atau menghindarinya.

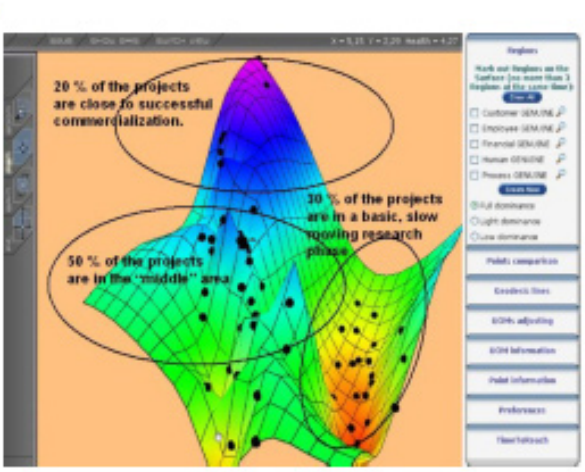
5.4.3 Varian SWOT

Berbagai analisis pelengkap untuk SWOT telah diusulkan, seperti matriks Pertumbuhan-berbagi dan analisis lima kekuatan Porter.

Matrik Tows

Heinz Weihrich mengatakan bahwa beberapa pengguna merasa sulit untuk menerjemahkan hasil analisis SWOT ke dalam tindakan yang berarti yang dapat diadopsi dalam strategi perusahaan yang lebih luas. Dia memperkenalkan TOWS Matrix, kerangka kerja konseptual yang membantu dalam menemukan tindakan yang paling efisien.

Analisis Lansekap SWOT



Lanskap SWOT *secara sistematis menyebarkan hubungan antara tujuan keseluruhan dan faktor-faktor SWOT yang mendasarinya dan menyediakan lanskap 3D interaktif dan dapat melakukan kueri*

Grafik lanskap SWOT membedakan situasi manajerial dengan memvisualisasikan dan meramalkan kinerja dinamis dari objek yang sebanding menurut temuan oleh Brendan Kitts, Leif Edvinsson, dan Tord Beding (2000).

Perubahan dalam kinerja relatif terus diidentifikasi. Proyek (atau unit pengukuran lainnya) yang bisa menjadi objek risiko atau peluang potensial disorot.

Lansekap SWOT juga menunjukkan faktor kekuatan dan kelemahan yang mendasarinya yang memiliki pengaruh atau kemungkinan akan memiliki pengaruh tertinggi dalam konteks nilai yang digunakan (mis. Fluktuasi nilai modal).

Perencanaan Perusahaan

Sebagai bagian dari pengembangan strategi dan rencana untuk memungkinkan organisasi mencapai tujuannya, organisasi itu akan menggunakan proses sistematis / keras yang dikenal sebagai perencanaan perusahaan. SWOT bersama PEST / PESTLE dapat digunakan sebagai dasar untuk analisis faktor bisnis dan lingkungan.

- Tetapkan tujuan - menentukan apa yang akan dilakukan organisasi
- Pemindaian lingkungan
- Penilaian internal SWOT organisasi, ini perlu mencakup penilaian situasi saat ini serta portofolio produk / layanan dan analisis siklus hidup produk / layanan
- Analisis strategi yang ada, ini harus menentukan relevansi dari hasil penilaian internal / eksternal. Ini mungkin termasuk analisis kesenjangan faktor lingkungan
- Masalah Strategis didefinisikan - faktor kunci dalam pengembangan rencana perusahaan yang harus ditangani organisasi
- Mengembangkan strategi baru / revisi - analisis revisi masalah strategis dapat berarti tujuan perlu diubah
- Menetapkan faktor-faktor keberhasilan kritis – pencapaian pencapaian tujuan dan implementasi strategi
- Persiapan rencana operasional, sumber daya, proyek untuk implementasi strategi
- Memantau hasil - pemetaan terhadap rencana, mengambil tindakan korektif, yang dapat berarti mengubah tujuan / strategi

Pemasaran

Dalam banyak analisis pesaing, pemasar membuat profil terperinci masing-masing pesaing di pasar, dengan fokus terutama pada kekuatan dan kelemahan kompetitif relatif mereka menggunakan analisis SWOT. Manajer pemasaran akan memeriksa struktur biaya masing-masing pesaing, sumber laba, sumber daya dan kompetensi, penentuan posisi kompetitif dan diferensiasi produk, tingkat integrasi vertikal, respons historis terhadap perkembangan industri, dan faktor lainnya.

Manajemen pemasaran sering merasa perlu untuk berinvestasi dalam penelitian untuk mengumpulkan data yang diperlukan untuk melakukan analisis pemasaran yang akurat. Karena itu, manajemen sering melakukan riset pasar (riset pemasaran bergantian) untuk mendapatkan informasi ini. Pemasar menggunakan berbagai teknik untuk melakukan riset pasar, tetapi beberapa yang lebih umum termasuk:

- Riset pemasaran kualitatif seperti kelompok fokus
- Penelitian pemasaran kuantitatif seperti survei statistik
- Teknik eksperimental seperti pasar uji
- Teknik observasi seperti observasi etnografi (di tempat)
- Manajer pemasaran juga dapat merancang dan mengawasi berbagai pemindaian lingkungan dan proses intelijen kompetitif untuk membantu mengidentifikasi tren dan menginformasikan analisis pemasaran perusahaan.

Di bawah ini adalah contoh analisis SWOT dari posisi pasar dari konsultan manajemen kecil dengan spesialisasi dalam HRM.

Strengths	Weakness	Opportunities	Threaths
Reputasi di pasar	Kekurangan konsultan di tingkat operasi daripada tingkat mitra	Posisi mapan dengan ceruk pasar yang terdefinisi dengan baik	Konsultan besar beroperasi di tingkat kecil
Keahlian di tingkat mitra dalam konsultasi HRM	Tidak dapat menangani tugas multidisiplin karena ukuran atau kurangnya kemampuan	Pasar yang teridentifikasi untuk konsultasi di bidang selain HRM	Konsultan kecil lainnya yang ingin menyerbu pasar

Penerapan SWOT Dalam Organisasi Masyarakat

Analisis SWOT telah digunakan dalam pekerjaan masyarakat sebagai alat untuk mengidentifikasi faktor-faktor positif dan negatif dalam organisasi, masyarakat, dan masyarakat yang lebih luas yang mempromosikan atau menghambat keberhasilan pelaksanaan layanan sosial dan upaya perubahan sosial. Ini digunakan sebagai sumber daya awal, menilai kekuatan, kelemahan, peluang, dan ancaman dalam komunitas yang dilayani oleh organisasi nirlaba atau komunitas. Alat pengorganisasian ini paling baik digunakan dalam kolaborasi dengan pekerja masyarakat dan / atau anggota masyarakat sebelum mengembangkan tujuan dan sasaran untuk desain program atau menerapkan strategi pengorganisasian. Analisis SWOT adalah bagian dari perencanaan untuk proses perubahan sosial dan tidak akan memberikan strategi rencanakan jika digunakan dengan sendirinya. Setelah analisis SWOT selesai, organisasi perubahan sosial dapat mengubah daftar SWOT menjadi serangkaian rekomendasi untuk dipertimbangkan sebelum mengembangkan rencana strategis.

Analisis SWOT

	Strenghts 1. 2. 3. 4.	Weakness 1. 2. 3. 4.
Opportunities 1. 2. 3. 4.	Strategi Kekuatan-Peluang Menggunakan kekuatan untuk memanfaatkan peluang	Strategi Peluang-Kelemahan mengatasi kelemahan dengan memanfaatkan peluang
Threaths 1. 2. 3. 4.	Strategi Ancaman-Kekuatan Menggunakan kekuatan untuk menghindari ancaman	Strategi Ancaman-Kelemahan Meminimalkan kelemahan untuk menghindari ancaman

one example *of a SWOT Analysis used in community organizing*

Analisis SWOT

Internal		**Eksternal**	
Strenghts	**Weakness**	**Opportunities**	**Threaths**

Analisis SWOT *sederhana yang digunakan dalam Pengorganisasian Masyarakat*

Kekuatan dan Kelemahan: Ini adalah faktor internal dalam suatu organisasi.

- Sumber daya manusia - staf, sukarelawan, anggota dewan, populasi target
- Sumber daya fisik - lokasi, bangunan, peralatan Anda
- Hibah keuangan, agen pendanaan, sumber pendapatan lain
- Kegiatan dan proses - program yang Anda jalankan, sistem yang Anda gunakan
- Pengalaman masa lalu - membangun blok untuk belajar dan sukses, reputasi Anda di masyarakat

Peluang dan Ancaman: Ini adalah faktor eksternal yang berasal dari kekuatan masyarakat atau masyarakat.

- Tren masa depan di bidang Anda atau budaya
- Ekonomi - lokal, nasional, atau internasional
- Sumber pendanaan - yayasan, donor, legislatif
- Demografi - perubahan usia, ras, jenis kelamin, budaya mereka yang Anda layani atau di daerah Anda
- Lingkungan fisik (Apakah bangunan Anda berada di bagian kota yang sedang tumbuh? Apakah perusahaan bus memotong rute?)
- Legislasi (Apakah persyaratan federal yang baru membuat pekerjaan Anda lebih sulit ... atau lebih mudah?)
- Acara lokal, nasional, atau internasional

Meskipun analisis SWOT pada awalnya dirancang sebagai metode organisasi untuk bisnis dan industri, analisis SWOT telah direplikasi dalam berbagai kerja komunitas sebagai alat untuk mengidentifikasi dukungan eksternal dan internal untuk memerangi oposisi internal dan eksternal. Analisis SWOT diperlukan untuk memberikan arahan ke tahap selanjutnya dari proses perubahan. Ini telah digunakan oleh pengorganisir masyarakat dan anggota masyarakat untuk memajukan keadilan sosial dalam konteks praktik Pekerjaan Sosial.

Aplikasi dalam Kelompok Organisasi

Elemen untuk Dipertimbangkan

Elemen yang perlu dipertimbangkan dalam analisis SWOT termasuk memahami komunitas tempat suatu organisasi bekerja. Ini dapat dilakukan melalui forum publik, kampanye mendengarkan, dan wawancara informasi. Pengumpulan data akan membantu menginformasikan anggota masyarakat dan pekerja ketika mengembangkan analisis SWOT. Penilaian kebutuhan dan aset adalah alat yang dapat digunakan untuk mengidentifikasi kebutuhan dan sumber daya masyarakat yang ada. Ketika penilaian ini dilakukan dan data telah dikumpulkan, analisis masyarakat dapat dibuat yang menginformasikan analisis SWOT.

Langkah-langkah untuk Implementasi

Analisis SWOT paling baik dikembangkan dalam pengaturan kelompok seperti pertemuan kerja atau komunitas. Seorang fasilitator dapat melakukan pertemuan dengan terlebih

dahulu menjelaskan apa analisis SWOT serta mengidentifikasi makna dari setiap istilah.

Salah satu cara untuk memfasilitasi pengembangan analisis SWOT termasuk mengembangkan contoh SWOT dengan kelompok yang lebih besar kemudian memisahkan masing-masing kelompok menjadi tim yang lebih kecil untuk dipresentasikan kepada kelompok yang lebih besar setelah menetapkan jumlah waktu. Hal ini memungkinkan individu, yang dapat dibungkam dalam pengaturan kelompok yang lebih besar, untuk berkontribusi. Setelah waktu yang diberikan habis, fasilitator dapat mencatat semua faktor dari masing-masing kelompok ke dalam dokumen besar seperti papan poster, dan kemudian kelompok besar, sebagai kolektif, dapat bekerja melalui masing-masing ancaman dan kelemahan untuk mengeksplorasi opsi yang dapat digunakan untuk memerangi kekuatan negatif dengan kekuatan dan peluang yang ada dalam organisasi dan komunitas. Pertemuan SWOT memungkinkan peserta untuk melakukan brainstorming secara kreatif, mengidentifikasi hambatan, dan mungkin membuat strategi solusi / jalan ke depan untuk keterbatasan ini.

Kapan menggunakan Analisis SWOT

Penggunaan analisis SWOT oleh organisasi masyarakat adalah sebagai berikut: untuk mengatur informasi, memberikan wawasan tentang hambatan yang mungkin ada saat terlibat dalam proses perubahan sosial, dan mengidentifikasi kekuatan yang tersedia yang dapat diaktifkan untuk mengatasi hambatan ini.

- Analisis SWOT dapat digunakan untuk:
- Jelajahi solusi baru untuk masalah
- Identifikasi hambatan yang akan membatasi sasaran / sasaran
- Tentukan arah yang akan paling efektif
- Mengungkapkan kemungkinan dan batasan untuk perubahan
- Untuk merevisi rencana untuk sistem navigasi terbaik, komunitas, dan organisasi
- Sebagai alat tukar pendapat dan perekaman sebagai alat komunikasi
- Untuk meningkatkan “kredibilitas interpretasi” untuk digunakan dalam presentasi kepada para pemimpin atau pendukung utama.

Manfaat SWOT

Analisis SWOT dalam kerangka kerja kerja sosial bermanfaat karena membantu organisasi memutuskan apakah suatu tujuan dapat diperoleh atau tidak dan karenanya memungkinkan organisasi untuk menetapkan tujuan, sasaran, dan langkah yang dapat dicapai untuk memajukan perubahan sosial atau upaya pengembangan masyarakat. Ini memungkinkan penyelenggara untuk mengambil visi dan menghasilkan hasil yang praktis dan efisien yang mempengaruhi perubahan jangka panjang, dan membantu organisasi mengumpulkan informasi yang bermakna untuk memaksimalkan potensi mereka. Menyelesaikan analisis SWOT adalah proses yang berguna mengenai pertimbangan prioritas organisasi utama, seperti jender dan keanekaragaman budaya dan tujuan penggalangan dana.

Keterbatasan

Beberapa temuan dari Menon et al. (1999) dan Hill dan Westbrook (1997) telah menyarankan bahwa SWOT dapat merusak kinerja dan bahwa "tidak ada yang kemudian menggunakan output dalam tahap selanjutnya dari strategi".

Kritik lain termasuk penyalahgunaan analisis SWOT sebagai teknik yang dapat dengan cepat dirancang tanpa pemikiran kritis yang mengarah pada penggambaran yang keliru tentang kekuatan, kelemahan, peluang, dan ancaman dalam lingkungan internal dan eksternal organisasi.

Keterbatasan lain termasuk pengembangan analisis SWOT hanya untuk mempertahankan tujuan dan sasaran yang telah diputuskan sebelumnya. Penyalahgunaan ini mengarah pada pembatasan kemungkinan brainstorming dan identifikasi hambatan yang "nyata". Penyalahgunaan ini juga menempatkan kepentingan organisasi di atas kesejahteraan komunitas. Selanjutnya, analisis SWOT harus dikembangkan sebagai kolaborasi dengan berbagai kontribusi yang dibuat oleh peserta termasuk anggota masyarakat. Rancangan analisis SWOT oleh satu atau dua pekerja komunitas terbatas pada realitas kekuatan, khususnya faktor eksternal, dan merendahkan kontribusi yang mungkin diberikan anggota masyarakat.

5.5 Riset Pemasaran

Riset pemasaran adalah "proses atau serangkaian proses yang menghubungkan produsen, pelanggan, dan pengguna akhir dengan pemasar melalui informasi - informasi yang digunakan untuk mengidentifikasi dan menentukan peluang dan masalah pemasaran;

menghasilkan, memperbaiki, dan mengevaluasi tindakan pemasaran; memantau kinerja pemasaran; dan meningkatkan pemahaman pemasaran sebagai suatu proses. Riset pemasaran menetapkan informasi yang diperlukan untuk mengatasi masalah ini, merancang metode untuk mengumpulkan informasi, mengelola dan mengimplementasikan proses pengumpulan data, menganalisis hasil, dan mengomunikasikan temuan dan implikasinya. ”

Ini adalah pengumpulan, pencatatan, dan analisis sistematis data kualitatif dan kuantitatif tentang masalah yang berkaitan dengan produk dan layanan pemasaran. Tujuan riset pemasaran adalah untuk mengidentifikasi dan menilai bagaimana elemen-elemen perubahan bauran pemasaran memengaruhi perilaku pelanggan. Istilah ini umumnya dipertukarkan dengan riset pasar; namun, para praktisi ahli mungkin ingin membuat perbedaan, karena riset pasar berkaitan secara khusus dengan pasar, sementara riset pemasaran memperhatikan secara khusus tentang proses pemasaran.

Riset pemasaran sering dipartisi menjadi dua set pasangan kategori, baik berdasarkan target pasar:

- Riset pemasaran konsumen, dan
- Riset pemasaran Business-to-business (B2B)

Atau, sebagai alternatif, dengan pendekatan metodologis:

- Riset pemasaran kualitatif, dan
- Penelitian pemasaran kuantitatif

Riset pemasaran konsumen adalah bentuk sosiologi terapan yang berkonsentrasi pada pemahaman preferensi, sikap, dan perilaku konsumen dalam ekonomi berbasis pasar, dan bertujuan untuk memahami efek dan keberhasilan komparatif kampanye pemasaran. Bidang riset pemasaran konsumen sebagai ilmu statistik dipelopori oleh Arthur Nielsen dengan berdirinya Perusahaan ACNielsen pada tahun 1923.

Dengan demikian, riset pemasaran juga dapat digambarkan sebagai identifikasi sistematis, objektif, pengumpulan, analisis, dan penyebaran informasi untuk tujuan membantu manajemen dalam pengambilan keputusan terkait dengan identifikasi dan solusi masalah dan peluang dalam pemasaran.

Alur Wewenang dan Tanggung Jawab

Tugas riset pemasaran (MR) adalah memberikan informasi pasar yang relevan, akurat, andal, valid, dan terkini kepada manajemen. Lingkungan pemasaran yang kompetitif dan biaya yang terus meningkat yang disebabkan oleh pengambilan keputusan yang buruk mengharuskan penelitian pemasaran memberikan informasi yang baik. Keputusan yang tepat tidak didasarkan pada perasaan, intuisi, atau bahkan penilaian murni.

Manajer membuat banyak keputusan strategis dan taktis dalam proses mengidentifikasi dan memuaskan kebutuhan pelanggan. Mereka membuat keputusan tentang peluang potensial, pemilihan target pasar, segmentasi pasar, perencanaan dan implementasi program pemasaran, kinerja pemasaran, dan kontrol. Keputusan-keputusan ini diperumit oleh interaksi antara variabel pemasaran yang dapat dikendalikan dari produk, harga, promosi, dan distribusi. Komplikasi lebih lanjut ditambahkan oleh faktor lingkungan yang tidak terkendali seperti kondisi ekonomi secara umum, teknologi, kebijakan publik dan hukum, lingkungan politik, persaingan, dan perubahan sosial dan budaya. Faktor lain dalam campuran ini adalah kompleksitas konsumen. Riset pemasaran membantu manajer pemasaran menghubungkan variabel pemasaran dengan lingkungan dan konsumen. Ini membantu menghilangkan beberapa ketidakpastian dengan memberikan informasi yang relevan tentang variabel pemasaran, lingkungan, dan konsumen. Dengan tidak adanya informasi yang relevan, respons konsumen terhadap program pemasaran tidak dapat diprediksi secara andal atau akurat. Program riset pemasaran yang sedang berlangsung memberikan informasi tentang faktor dan konsumen yang dapat dikendalikan dan tidak dapat dikendalikan; informasi ini meningkatkan efektivitas keputusan yang dibuat oleh manajer pemasaran.

Secara tradisional, peneliti pemasaran bertanggung jawab untuk menyediakan informasi yang relevan dan keputusan pemasaran dibuat oleh para manajer. Namun, perannya berubah dan peneliti pemasaran menjadi lebih terlibat dalam pengambilan keputusan, sedangkan manajer pemasaran menjadi lebih terlibat dengan penelitian. Peran riset pemasaran dalam pengambilan keputusan manajerial dijelaskan lebih lanjut dengan menggunakan kerangka kerja model “DECIDE”.

Sejarah

Riset pemasaran telah berkembang dalam beberapa dekade sejak Arthur Nielsen menetapkannya sebagai industri yang aktif, yang akan tumbuh seiring dengan ekonomi

B2B dan B2C. Pasar secara alami berevolusi, dan sejak lahirnya ACNielsen, ketika penelitian terutama dilakukan oleh kelompok fokus secara langsung dan survei pena-dan-kertas, kebangkitan Internet dan proliferasi situs web perusahaan telah mengubah cara penelitian dieksekusi.

Analisis web lahir dari kebutuhan untuk melacak perilaku pengunjung situs dan, seiring dengan semakin populernya e-commerce dan iklan web, bisnis menuntut rincian tentang informasi yang diciptakan oleh praktik baru dalam pengumpulan data web, seperti rasio klik-tayang dan keluar. Ketika Internet berkembang pesat, situs web menjadi lebih besar dan lebih kompleks dan kemungkinan komunikasi dua arah antara bisnis dan konsumen mereka menjadi kenyataan. Disediakan dengan kapasitas untuk berinteraksi dengan pelanggan online, Peneliti dapat mengumpulkan sejumlah besar data yang sebelumnya tidak tersedia, lebih lanjut mendorong Industri Riset Pemasaran.

Pada milenium baru, ketika Internet terus berkembang dan situs web menjadi lebih interaktif, pengumpulan dan analisis data menjadi lebih umum bagi Firma Riset Pemasaran yang kliennya memiliki keberadaan web. Dengan pertumbuhan eksplosif dari pasar online, muncul kompetisi baru untuk perusahaan; bisnis tidak lagi hanya bersaing dengan toko di ujung jalan - persaingan sekarang diwakili oleh kekuatan global. Gerai ritel muncul secara online dan kebutuhan sebelumnya untuk toko batu bata dan mortir berkurang pada kecepatan yang lebih besar daripada kompetisi online yang berkembang. Dengan begitu banyak saluran online bagi konsumen untuk melakukan pembelian, perusahaan membutuhkan metode yang lebih baru dan lebih menarik, di kombinasi dengan pesan yang beresonansi lebih efektif, untuk menarik perhatian konsumen rata-rata.

Memiliki akses ke data web tidak secara otomatis memberikan alasan kepada perusahaan di balik perilaku pengguna yang mengunjungi situs mereka, yang memicu industri riset pemasaran untuk mengembangkan cara-cara baru dan lebih baik untuk melacak, mengumpulkan, dan menafsirkan informasi. Ini mengarah pada pengembangan berbagai alat seperti kelompok fokus online dan survei pop-up atau intersep situs web. Jenis layanan ini memungkinkan perusahaan untuk menggali lebih dalam motivasi konsumen, menambah wawasan mereka dan memanfaatkan data ini untuk mendorong pangsa pasar.

Ketika informasi di seluruh dunia menjadi lebih mudah diakses, meningkatnya persaingan mendorong perusahaan untuk menuntut lebih banyak Peneliti Pasar. Itu tidak lagi cukup untuk mengikuti tren dalam perilaku web atau melacak data penjualan; perusahaan

sekarang membutuhkan akses ke perilaku konsumen di seluruh proses pembelian. Ini berarti Industri Riset Pemasaran, sekali lagi, perlu beradaptasi dengan kebutuhan pasar yang berubah dengan cepat, dan terhadap tuntutan perusahaan yang mencari keunggulan kompetitif.

Saat ini, Riset Pemasaran telah beradaptasi dengan inovasi dalam teknologi dan kemudahan terkait dengan informasi yang tersedia. Perusahaan B2B dan B2C bekerja keras untuk tetap kompetitif dan mereka sekarang menuntut riset pemasaran kuantitatif ("Apa") dan kualitatif ("Mengapa?") Untuk lebih memahami target audiens mereka dan motivasi di balik perilaku pelanggan.

Permintaan ini mendorong Peneliti Pemasaran untuk mengembangkan platform baru untuk komunikasi dua arah yang interaktif antara perusahaan dan konsumen mereka. Perangkat seluler seperti SmartPhones adalah contoh terbaik dari platform yang muncul yang memungkinkan perusahaan untuk terhubung dengan pelanggan mereka di seluruh proses pembelian. Perusahaan-perusahaan riset yang inovatif, seperti OnResearch dengan aplikasi OnMobile mereka, sekarang menyediakan bisnis dengan sarana untuk menjangkau konsumen dari titik penyelidikan awal hingga keputusan dan, pada akhirnya, pembelian.

Ketika perangkat seluler pribadi menjadi lebih mampu dan tersebar luas, Industri Riset Pemasaran akan berupaya memanfaatkan tren ini lebih jauh. Perangkat seluler menghadirkan saluran yang sempurna untuk Firma Riset untuk mendapatkan tayangan langsung dari pembeli dan memberikan pandangan menyeluruh tentang konsumen kepada konsumen dalam pasar target mereka, dan selanjutnya. Sekarang, lebih dari sebelumnya, inovasi adalah kunci kesuksesan bagi Peneliti Pemasaran. Klien Riset Pemasaran mulai menuntut produk yang sangat personal dan berfokus khusus dari perusahaan MR; big data sangat bagus untuk mengidentifikasi segmen pasar umum, tetapi kurang mampu mengidentifikasi faktor-faktor kunci dari ceruk pasar, yang sekarang mendefinisikan keunggulan kompetitif yang dicari perusahaan di era mobile-digital ini.

Karakteristik

Pertama, riset pemasaran sistematis. Dengan demikian perencanaan sistematis diperlukan pada semua tahap proses riset pemasaran. Prosedur yang diikuti pada setiap tahap secara metodologi baik, didokumentasikan dengan baik, dan, sebanyak mungkin, direncanakan sebelumnya. Riset pemasaran menggunakan metode ilmiah dalam hal data dikumpulkan

dan dianalisis untuk menguji gagasan atau hipotesis sebelumnya. Para ahli dalam riset pemasaran telah menunjukkan bahwa studi yang menampilkan banyak hipotesis dan sering bersaing menghasilkan hasil yang lebih bermakna daripada yang hanya menampilkan satu hipotesis dominan.

Riset pemasaran adalah objektif. Ia berusaha memberikan informasi akurat yang mencerminkan keadaan sebenarnya. Itu harus dilakukan tanpa memihak. Sementara penelitian selalu dipengaruhi oleh filosofi penelitian peneliti, itu harus bebas dari bias pribadi atau politik peneliti atau manajemen. Penelitian yang dimotivasi oleh keuntungan pribadi atau politik melibatkan pelanggaran standar profesional. Penelitian semacam itu sengaja dibuat bias sehingga menghasilkan temuan yang telah ditentukan. Sifat obyektif dari riset pemasaran menggarisbawahi pentingnya pertimbangan etis. Selain itu, peneliti harus selalu objektif sehubungan dengan pemilihan informasi yang akan ditampilkan dalam teks referensi karena literatur tersebut harus menawarkan pandangan yang komprehensif tentang pemasaran. Namun, penelitian telah menunjukkan bahwa banyak buku teks pemasaran tidak menampilkan prinsip-prinsip penting dalam riset pemasaran.

Penelitian Bisnis Terkait

Bentuk lain dari riset bisnis meliputi:

- Riset pasar lebih luas cakupannya dan memeriksa semua aspek lingkungan bisnis. Ini mengajukan pertanyaan tentang pesaing, struktur pasar, peraturan pemerintah, tren ekonomi, kemajuan teknologi, dan berbagai faktor lain yang membentuk lingkungan bisnis. Terkadang istilah ini lebih mengacu pada analisis keuangan perusahaan, industri, atau sektor. Dalam hal ini, analis keuangan biasanya melakukan penelitian dan memberikan hasilnya kepada penasihat investasi dan investor potensial.
- Penelitian produk - Ini melihat produk apa yang dapat diproduksi dengan teknologi yang tersedia, dan apa yang dapat dikembangkan oleh inovasi produk baru di masa mendatang.
- Penelitian periklanan - adalah bentuk khusus riset pemasaran yang dilakukan untuk meningkatkan kemanjuran iklan. Pengujian salin, juga dikenal sebagai “pra-pengujian,” adalah bentuk penelitian khusus yang memprediksi kinerja iklan di pasar sebelum ditayangkan, dengan menganalisis tingkat perhatian pemirsa, keterkaitan merek, motivasi, hiburan, dan komunikasi, juga sebagai penguraian aliran perhatian iklan dan aliran emisi. Pra-pengujian juga digunakan pada iklan yang masih dalam bentuk kasar (ripomatic atau animatic). (Young, p. 213)

Klasifikasi

Organisasi terlibat dalam riset pemasaran karena dua alasan: (1) untuk mengidentifikasi dan (2) memecahkan masalah pemasaran. Perbedaan ini berfungsi sebagai dasar untuk mengklasifikasikan riset pemasaran menjadi riset identifikasi masalah dan riset penyelesaian masalah.

Penelitian identifikasi masalah dilakukan untuk membantu mengidentifikasi masalah yang, mungkin, tidak terlihat di permukaan dan belum ada atau kemungkinan akan muncul di masa depan seperti citra perusahaan, karakteristik pasar, analisis penjualan, perkiraan jangka pendek, perkiraan jarak jauh, dan penelitian tren bisnis. Penelitian jenis ini memberikan informasi tentang lingkungan pemasaran dan membantu mendiagnosis masalah. Sebagai contoh, Temuan-temuan penelitian penyelesaian masalah digunakan dalam membuat keputusan yang akan memecahkan masalah pemasaran tertentu.

Stanford Research Institute, di sisi lain, melakukan survei tahunan terhadap konsumen yang digunakan untuk mengklasifikasikan orang ke dalam kelompok-kelompok homogen untuk tujuan segmentasi. Panel Diary Pembelian Nasional (NPD) mempertahankan panel buku harian terbesar di Amerika Serikat.

Layanan standar adalah studi penelitian yang dilakukan untuk perusahaan klien yang berbeda tetapi dengan cara standar. Misalnya, prosedur untuk mengukur efektivitas periklanan telah distandarisasi sehingga hasilnya dapat dibandingkan lintas studi dan norma-norma evaluatif dapat ditetapkan. The Starch Readership Survey adalah layanan yang paling banyak digunakan untuk mengevaluasi iklan cetak; layanan lain yang terkenal adalah Studi Dampak Majalah Gallup dan Robinson. Layanan ini juga dijual secara sindikasi.

- Layanan khusus menawarkan berbagai layanan riset pemasaran yang disesuaikan dengan kebutuhan spesifik klien. Setiap proyek riset pemasaran diperlakukan secara unik.
- Pemasok layanan terbatas berspesialisasi dalam satu atau beberapa fase proyek riset pemasaran. Layanan yang ditawarkan oleh pemasok tersebut diklasifikasikan sebagai layanan lapangan, koding dan entri data, analisis data, layanan analitis, dan produk bermerek. Layanan lapangan mengumpulkan data melalui internet, surat tradisional, wawancara langsung, atau wawancara telepon, dan perusahaan yang berspesialisasi dalam wawancara disebut organisasi layanan lapangan. Organisasi-organisasi ini dapat berkisar dari organisasi berpemilik kecil yang beroperasi secara

lokal hingga organisasi multinasional besar dengan fasilitas wawancara line WATS. Beberapa organisasi memiliki fasilitas wawancara yang luas di seluruh negeri untuk mewawancarai pembeli di mal.

- Layanan pengodean dan entri data meliputi pengeditan kuesioner yang diisi lengkap, pengembangan skema pengkodean, dan menyalin data ke disket atau pita magnetik untuk input ke komputer. Sistem Data NRC menyediakan layanan tersebut.
- Layanan analisis meliputi perancangan dan pretest kuesioner, penentuan cara terbaik untuk mengumpulkan data, merancang rencana pengambilan sampel, dan aspek lain dari desain penelitian. Beberapa proyek riset pemasaran yang kompleks membutuhkan pengetahuan tentang prosedur yang canggih, termasuk desain eksperimental khusus, dan teknik analisis seperti analisis bersama dan penskalaan multidimensi. Keahlian semacam ini dapat diperoleh dari perusahaan dan konsultan yang berspesialisasi dalam layanan analitis.
- Layanan analisis data ditawarkan oleh perusahaan, juga dikenal sebagai rumah tab, yang berspesialisasi dalam analisis komputer data kuantitatif seperti yang diperoleh dalam survei besar. Awalnya sebagian besar perusahaan analisis data hanya menyediakan tabulasi (jumlah frekuensi) dan tabulasi silang (jumlah frekuensi yang menggambarkan dua atau lebih variabel secara bersamaan). Dengan proliferasi perangkat lunak, banyak perusahaan sekarang memiliki kemampuan untuk menganalisis data mereka sendiri, tetapi, perusahaan analisis data masih diminati.
- Produk dan layanan riset pemasaran bermerek adalah prosedur pengumpulan dan analisis data khusus yang dikembangkan untuk mengatasi jenis masalah riset pemasaran tertentu. Prosedur ini dipatenkan, diberi nama merek, dan dipasarkan seperti produk bermerek lainnya.

Jenis

Teknik riset pemasaran datang dalam berbagai bentuk, termasuk:

- Pelacakan Iklan - penelitian di pasar secara berkala atau berkelanjutan untuk memantau kinerja suatu merek menggunakan langkah-langkah seperti kesadaran merek, preferensi merek, dan penggunaan produk. (Young, 2005)
- Penelitian Periklanan - digunakan untuk memprediksi pengujian salinan atau melacak kemanjuran iklan untuk media apa pun, diukur dengan kemampuan iklan untuk mendapatkan perhatian (diukur dengan Attention-Tracking), mengkomunikasikan pesan, membangun citra merek, dan memotivasi konsumen untuk membeli produk atau layanan. (Young, 2005)

- Penelitian kesadaran merek - sejauh mana konsumen dapat mengingat atau mengenali nama merek atau nama produk
- Penelitian asosiasi merek - apa yang diasosiasikan konsumen dengan merek?
- Penelitian atribut merek - sifat-sifat utama apa yang menggambarkan janji merek?
- Pengujian nama merek - apa yang dirasakan konsumen tentang nama-nama produk?
- Proses pengambilan keputusan pembeli — untuk menentukan apa yang memotivasi orang untuk membeli dan proses pengambilan keputusan apa yang mereka gunakan; selama dekade terakhir, Neuromarketing muncul dari konvergensi neuroscience dan pemasaran, yang bertujuan untuk memahami proses pengambilan keputusan konsumen
- Penelitian pelacakan mata komersial - memeriksa iklan, desain paket, situs web, dll. Dengan menganalisis perilaku visual konsumen
- Pengujian konsep - untuk menguji penerimaan konsep oleh target konsumen
- Coolhunting (juga dikenal sebagai trendspotting) - untuk melakukan pengamatan dan prediksi dalam perubahan tren budaya baru atau yang sudah ada di berbagai bidang seperti mode, musik, film, televisi, budaya dan gaya hidup kaum muda
- Pengujian salin - memprediksi kinerja iklan di pasar sebelum mengudara dengan menganalisis tingkat perhatian pemirsa, keterkaitan merek, motivasi, hiburan, dan komunikasi, serta memecah aliran perhatian dan aliran emosi iklan. (Young, p 213)
- Penelitian kepuasan pelanggan - studi kuantitatif atau kualitatif yang menghasilkan pemahaman tentang kepuasan pelanggan dengan transaksi
- Estimasi permintaan - untuk menentukan perkiraan tingkat permintaan produk
- Audit saluran distribusi - untuk menilai sikap distributor dan pengecer terhadap produk, merek, atau perusahaan
- Intelijen strategis Internet - mencari pendapat pelanggan di Internet: obrolan, forum, halaman web, blog ... di mana orang mengekspresikan secara bebas tentang pengalaman mereka dengan produk, menjadi pembentuk opini yang kuat.
- Keefektifan dan analisis pemasaran - Membangun model dan mengukur hasil untuk menentukan efektivitas aktivitas pemasaran individu.
- Mystery consumer atau Mystery shopping - Seorang karyawan atau perwakilan dari firma riset pasar secara anonim menghubungi seorang tenaga penjualan dan menunjukkan dia sedang berbelanja untuk suatu produk. Pembeli kemudian mencatat seluruh pengalaman. Metode ini sering digunakan untuk kontrol kualitas atau untuk meneliti produk pesaing.
- Penelitian positioning - bagaimana pasar sasaran melihat merek relatif terhadap pesaing? Untuk apa merek ini berdiri?

- Uji elastisitas harga - untuk menentukan seberapa sensitif pelanggan terhadap perubahan harga
- Perkiraan penjualan - untuk menentukan tingkat penjualan yang diharapkan berdasarkan tingkat permintaan. Sehubungan dengan faktor-faktor lain seperti pengeluaran iklan, promosi penjualan dll.
- Penelitian segmentasi - untuk menentukan karakteristik demografis, psikografis, budaya, dan perilaku pembeli potensial
- Panel online - sekelompok individu yang menerima untuk menanggapi riset pemasaran online
- Audit toko - untuk mengukur penjualan suatu produk atau lini produk pada sampel toko yang dipilih secara statistik untuk menentukan pangsa pasar, atau untuk menentukan apakah toko ritel menyediakan layanan yang memadai.
- Tes pemasaran - peluncuran produk skala kecil yang digunakan untuk menentukan kemungkinan penerimaan produk ketika diperkenalkan ke pasar yang lebih luas
- Riset Pemasaran Viral - merujuk pada riset pemasaran yang dirancang untuk memperkirakan probabilitas bahwa komunikasi spesifik akan ditransmisikan ke seluruh Jejaring Sosial individu. Estimasi Potensi Jejaring Sosial (SNP) dikombinasikan dengan perkiraan efektivitas penjualan untuk memperkirakan ROI pada kombinasi pesan dan media tertentu.

Semua bentuk riset pemasaran ini dapat digolongkan sebagai riset identifikasi masalah atau riset pemecahan masalah.

Ada dua sumber utama data - primer dan sekunder. Penelitian utama dilakukan dari awal. Ini asli dan dikumpulkan untuk menyelesaikan masalah yang ada. Penelitian sekunder sudah ada sejak dikumpulkan untuk tujuan lain. Ini dilakukan pada data yang diterbitkan sebelumnya dan biasanya oleh orang lain. Biaya penelitian sekunder jauh lebih murah daripada penelitian primer, tetapi jarang muncul dalam bentuk yang benar-benar memenuhi kebutuhan peneliti.

Perbedaan serupa ada antara penelitian eksplorasi dan penelitian konklusif. Penelitian eksplorasi memberikan wawasan dan pemahaman tentang suatu masalah atau situasi. Itu harus menarik kesimpulan pasti hanya dengan sangat hati-hati. Penelitian konklusif menarik kesimpulan: hasil penelitian dapat digeneralisasi untuk seluruh populasi.

Penelitian eksplorasi dilakukan untuk mengeksplorasi masalah untuk mendapatkan beberapa ide dasar tentang solusi pada tahap awal penelitian. Ini dapat berfungsi

sebagai input untuk penelitian konklusif. Informasi penelitian eksplorasi dikumpulkan dengan wawancara kelompok fokus, meninjau literatur atau buku, berdiskusi dengan para pakar, dll. Ini sifatnya tidak terstruktur dan kualitatif. Jika sumber data sekunder tidak dapat melayani tujuan, sampel kenyamanan ukuran kecil dapat dikumpulkan. Penelitian konklusif dilakukan untuk menarik beberapa kesimpulan tentang masalah tersebut. Ini pada dasarnya, penelitian terstruktur dan kuantitatif, dan output dari penelitian ini adalah input untuk sistem informasi manajemen (SIM).

Penelitian eksplorasi juga dilakukan untuk menyederhanakan temuan penelitian konklusif atau deskriptif, jika temuan tersebut sangat sulit untuk ditafsirkan bagi para manajer pemasaran.

Metode

Secara metodologis, riset pemasaran menggunakan tipe-tipe desain penelitian berikut:

Berdasarkan pertanyaan

- Penelitian pemasaran kualitatif - umumnya digunakan untuk tujuan eksplorasi - sejumlah kecil responden - tidak dapat digeneralisasikan untuk seluruh populasi - signifikansi statistik dan kepercayaan tidak dihitung - contoh termasuk kelompok fokus, wawancara mendalam, dan teknik proyektif
- Penelitian pemasaran kuantitatif - umumnya digunakan untuk menarik kesimpulan - menguji hipotesis tertentu - menggunakan teknik pengambilan sampel acak sehingga dapat menyimpulkan dari sampel ke populasi - melibatkan sejumlah besar responden - contoh termasuk survei dan kuesioner. Teknik meliputi pemodelan pilihan, penskalaan preferensi perbedaan maksimum, dan analisis kovarian.

Berdasarkan pengamatan

- Studi etnografi - pada dasarnya kualitatif, peneliti mengamati fenomena sosial dalam pengaturan alaminya - pengamatan dapat terjadi secara melintang (pengamatan dilakukan pada satu waktu) atau memanjang (pengamatan terjadi selama beberapa periode waktu) - contohnya mencakup analisis penggunaan produk dan jejak cookie komputer.
- Teknik eksperimental - secara alami kuantitatif, peneliti menciptakan lingkungan buatan semu untuk mencoba mengendalikan faktor palsu, kemudian memanipulasi setidaknya satu dari variabel - contohnya termasuk laboratorium pembelian dan pasar pengujian

Para peneliti sering menggunakan lebih dari satu desain penelitian. Mereka dapat mulai dengan penelitian sekunder untuk mendapatkan informasi latar belakang, kemudian melakukan kelompok fokus (desain penelitian kualitatif) untuk mengeksplorasi masalah tersebut. Akhirnya mereka dapat melakukan survei nasional penuh (desain penelitian kuantitatif) untuk menyusun rekomendasi spesifik untuk klien.

Bisnis to Bisnis

Riset bisnis ke bisnis (B2B) pasti lebih rumit daripada riset konsumen. Para peneliti perlu mengetahui apa jenis pendekatan multi-faceted yang akan menjawab tujuan, karena jarang mungkin untuk menemukan jawaban menggunakan hanya satu metode. Menemukan responden yang tepat sangat penting dalam penelitian B2B karena mereka sering sibuk, dan mungkin tidak ingin berpartisipasi. Mendorong mereka untuk "membuka diri" adalah keterampilan lain yang dibutuhkan oleh peneliti B2B. Terakhir, namun tidak kalah pentingnya, sebagian besar penelitian bisnis mengarah pada keputusan strategis dan ini berarti bahwa peneliti bisnis harus memiliki keahlian dalam mengembangkan strategi yang berakar kuat dalam temuan penelitian dan dapat diterima oleh klien.

Ada empat faktor utama yang membuat riset pasar B2B istimewa dan berbeda dari pasar konsumen:

- Unit pengambilan keputusan jauh lebih kompleks di pasar B2B daripada di pasar konsumen
- Produk B2B dan aplikasinya lebih kompleks daripada produk konsumen
- Pemasar B2B menangani jumlah pelanggan yang jauh lebih kecil yang sangat jauh lebih besar dalam konsumsi produk mereka daripada yang terjadi di pasar konsumen
- Hubungan pribadi sangat penting dalam pasar B2B.

Bisnis Kecil dan Organisasi Nirlaba

Riset pemasaran tidak hanya terjadi di perusahaan besar dengan banyak karyawan dan anggaran besar. Informasi pemasaran dapat diperoleh dengan mengamati lingkungan lokasi mereka dan lokasi kompetisi. Survei skala kecil dan kelompok fokus adalah cara berbiaya rendah untuk mengumpulkan informasi dari pelanggan potensial dan yang sudah ada. Sebagian besar data sekunder (statistik, demografi, dll.) Tersedia untuk umum di perpustakaan atau di internet dan dapat dengan mudah diakses oleh pemilik usaha kecil.

Berikut adalah beberapa langkah yang bisa dilakukan oleh UKM (Small Medium Entreprise) untuk menganalisis pasar:

1. Memberikan data sekunder dan atau primer (jika perlu);
2. Menganalisis data Ekonomi Makro & Mikro (mis. Pasokan & Permintaan, PDB, Perubahan harga, pertumbuhan ekonomi, Penjualan menurut sektor / industri, tingkat bunga, jumlah investasi / divestasi, I / O, CPI, Analisis sosial, dll.);
3. Menerapkan konsep bauran pemasaran, yang terdiri dari: Tempat, Harga, Produk, Promosi, Orang, Proses, Bukti Fisik dan juga situasi politik & sosial untuk menganalisis situasi pasar global);
4. Menganalisis tren pasar, pertumbuhan, ukuran pasar, pangsa pasar, persaingan pasar (misalnya analisis SWOT, Analisis B / C, identitas pemetaan saluran saluran utama, pendorong loyalitas dan kepuasan pelanggan, persepsi merek, tingkat kepuasan, saluran pesaing saat ini analisis hubungan, dll.), dll .;
5. Menentukan segmen pasar, target pasar, perkiraan pasar dan posisi pasar;
6. Merumuskan strategi pasar & juga menyelidiki kemungkinan kemitraan / kolaborasi (mis. Profiling & analisis SWOT dari mitra potensial, mengevaluasi kemitraan bisnis.)
7. Menggabungkan analisis tersebut dengan rencana bisnis / analisis model bisnis UKM (misalnya, deskripsi bisnis, proses bisnis, strategi bisnis, model pendapatan, ekspansi bisnis, pengembalian investasi, analisis keuangan (sejarah perusahaan, asumsi keuangan, analisis biaya / manfaat, proyeksi) untung & Rugi, Arus Kas, Neraca & Rasio Bisnis, dll.).

Catatan penting: Analisis keseluruhan harus didasarkan pada pertanyaan 6W + 1H (Apa, Kapan, Di mana, Siapa, Mengapa, dan Bagaimana)

Pemasaran Internasional

Riset Pemasaran Internasional mengikuti jalur yang sama dengan penelitian dalam negeri, tetapi ada beberapa masalah lagi yang mungkin muncul. Pelanggan di pasar internasional mungkin memiliki kebiasaan, budaya, dan harapan yang sangat berbeda dari perusahaan yang sama. Dalam hal ini, Riset Pemasaran lebih mengandalkan data primer daripada informasi sekunder. Mengumpulkan data primer dapat dihambat oleh bahasa, literasi dan akses ke teknologi. Informasi dasar Budaya dan Pasar akan diperlukan untuk memaksimalkan efektivitas penelitian. Beberapa langkah yang akan membantu mengatasi hambatan termasuk; 1. Mengumpulkan informasi sekunder tentang negara yang sedang dipelajari dari sumber internasional yang andal, mis. WHO dan IMF 2. Kumpulkan informasi sekunder tentang produk / layanan yang sedang dipelajari dari sumber yang tersedia 3.

Kumpulkan informasi sekunder tentang produsen produk dan penyedia layanan yang diteliti di negara terkait 4. Kumpulkan informasi sekunder tentang budaya dan praktik bisnis umum 5. Tanyakan pertanyaan untuk mendapatkan pemahaman yang lebih baik tentang alasan di balik rekomendasi untuk metodologi tertentu

Ketentuan Umum

Teknik riset pasar menyerupai yang digunakan dalam jajak pendapat politik dan penelitian ilmu sosial. Meta-analisis (juga disebut teknik Schmidt-Hunter) mengacu pada metode statistik yang menggabungkan data dari beberapa studi atau dari beberapa jenis studi. Konseptualisasi berarti proses mengubah citra mental yang samar menjadi konsep yang dapat didefinisikan. Operasionalisasi adalah proses mengubah konsep menjadi perilaku spesifik yang dapat diamati yang dapat diukur oleh peneliti. Presisi mengacu pada ketepatan ukuran yang diberikan. Reliabilitas mengacu pada kemungkinan bahwa konstruk operasional yang diberikan akan menghasilkan hasil yang sama jika diukur kembali. Validitas mengacu pada sejauh mana suatu ukuran menyediakan data yang menangkap makna konstruk yang dioperasionalkan sebagaimana didefinisikan dalam penelitian. Itu bertanya, “Apakah kita mengukur apa yang ingin kita ukur?”

- Penelitian terapan ditetapkan untuk membuktikan hipotesis nilai tertentu untuk klien yang membayar untuk penelitian. Misalnya, perusahaan rokok dapat melakukan penelitian yang berupaya menunjukkan bahwa rokok baik untuk kesehatan seseorang. Banyak peneliti memiliki keraguan etis tentang melakukan penelitian terapan.
- Sugging (dari SUG, untuk “penjualan dengan kedok” dari riset pasar) membentuk teknik penjualan di mana orang-orang penjualan berpura-pura melakukan riset pemasaran, tetapi dengan tujuan sebenarnya untuk memperoleh motivasi pembeli dan informasi pengambilan keputusan pembeli menjadi digunakan dalam panggilan penjualan berikutnya.
- Frugging terdiri dari praktik meminta dana dengan dalih menjadi organisasi penelitian.

Jenjang Karir

Beberapa posisi yang tersedia dalam riset pemasaran termasuk wakil presiden riset pemasaran, direktur penelitian, asisten direktur penelitian, manajer proyek, direktur kerja

lapangan, spesialis statistik / pemrosesan data, analis senior, analis, analis junior, dan penyelia operasional.

Posisi entry-level yang paling umum dalam riset pemasaran untuk orang dengan gelar sarjana (mis., BBA) adalah sebagai pengawas operasional. Orang-orang ini bertanggung jawab untuk mengawasi serangkaian operasi yang terdefinisi dengan baik, termasuk kerja lapangan, pengeditan data, dan pengkodean, dan mungkin terlibat dalam pemrograman dan analisis data. Posisi entry-level lain untuk BBA adalah asisten manajer proyek. Asisten manajer proyek akan belajar dan membantu dalam desain kuesioner, meninjau instruksi lapangan, dan memonitor waktu dan biaya studi. Namun, dalam industri riset pemasaran, ada preferensi yang semakin besar untuk orang-orang dengan gelar master. Mereka yang memiliki gelar MBA atau setara cenderung dipekerjakan sebagai manajer proyek.

Sejumlah kecil sekolah bisnis juga menawarkan gelar Master of Marketing Research (MMR) yang lebih terspesialisasi. MMR biasanya mempersiapkan siswa untuk berbagai metodologi penelitian dan berfokus pada pembelajaran baik di ruang kelas maupun di lapangan.

Posisi entry-level khas dalam perusahaan bisnis akan menjadi analis penelitian junior (untuk BBA) atau analis penelitian (untuk MBA atau MMR). Analis yunior dan analis riset belajar tentang industri tertentu dan menerima pelatihan dari anggota staf senior, biasanya manajer riset pemasaran. Posisi analis junior mencakup program pelatihan untuk mempersiapkan individu untuk tanggung jawab analis riset, termasuk berkoordinasi dengan departemen pemasaran dan tenaga penjualan untuk mengembangkan tujuan paparan produk. Tanggung jawab analis riset mencakup memeriksa semua data untuk akurasi, membandingkan dan membedakan penelitian baru dengan norma yang berlaku, dan menganalisis data primer dan sekunder untuk tujuan perkiraan pasar.

Seperti yang ditunjukkan oleh jabatan ini, orang-orang dengan berbagai latar belakang dan keterampilan diperlukan dalam riset pemasaran. Spesialis teknis seperti ahli statistik jelas membutuhkan latar belakang yang kuat dalam statistik dan analisis data. Posisi lain, seperti direktur penelitian, meminta untuk mengelola pekerjaan orang lain dan membutuhkan keterampilan yang lebih umum. Untuk mempersiapkan karir dalam riset pemasaran, siswa biasanya:

- Mengikuti semua kursus pemasaran.
- Mengikuti kursus statistik dan metode kuantitatif.
- Memperoleh keterampilan komputer.

- Mengkuti kursus psikologi dan perilaku konsumen.
- Memperoleh keterampilan komunikasi tertulis dan verbal yang efektif.
- Berpikir kreatif.

Hirarki Perusahaan

- Wakil Presiden Riset Pemasaran: Ini adalah posisi senior dalam riset pemasaran. VP bertanggung jawab atas keseluruhan operasi riset pemasaran perusahaan dan melayani di tim manajemen puncak. Tetapkan tujuan dan sasaran departemen riset pemasaran.
- Direktur Penelitian: Juga posisi senior, direktur memiliki tanggung jawab keseluruhan untuk pengembangan dan pelaksanaan semua proyek riset pemasaran.
- Asisten Direktur Penelitian: Berfungsi sebagai asisten administrasi untuk direktur dan mengawasi beberapa anggota staf riset pemasaran lainnya.
- Manajer Proyek (Senior): Memiliki tanggung jawab keseluruhan untuk desain, implementasi, dan manajemen proyek penelitian.
- Spesialis Statistik / Pemrosesan Data: Berfungsi sebagai pakar teori dan penerapan teknik statistik. Tanggung jawab meliputi desain eksperimental, pemrosesan data, dan analisis.
- Analis Senior: Berpartisipasi dalam pengembangan proyek dan mengarahkan pelaksanaan operasional proyek yang ditugaskan. Bekerja sama dengan analis, analis junior, dan personel lain dalam mengembangkan desain penelitian dan pengumpulan data. Mempersiapkan laporan akhir. Tanggung jawab utama untuk memenuhi waktu dan kendala biaya ada pada analis senior.
- Analis: Menangani detail yang terlibat dalam melaksanakan proyek. Desain dan pretest kuesioner dan melakukan analisis awal data.
- Junior Analyst: Menangani penugasan rutin seperti analisis data sekunder, pengeditan dan pengkodean kuesioner, dan analisis statistik sederhana.
- Direktur Pekerjaan Lapangan: Bertanggung jawab atas pemilihan, pelatihan, pengawasan, dan evaluasi pewawancara dan pekerja lapangan lainnya.

Bab 6 : Aspek Esensi Intelejensi Bisnis

Aspek penting dari intelijen bisnis adalah *Context Analysis*, manajemen kinerja bisnis, penemuan proses bisnis, sistem informasi, intelijen organisasi dan proses penambangan. Metode untuk menganalisis lingkungan bisnis apa pun dikenal sebagai *Context Analysis*. Topik yang dibahas dalam bagian ini sangat penting untuk memperluas pengetahuan yang ada tentang intelijen bisnis.

6.1 Context Analysis

Context Analysis (Analisis Konteks) adalah metode untuk menganalisis lingkungan di mana bisnis beroperasi. Pemindaian lingkungan terutama berfokus pada lingkungan makro bisnis. Tetapi analisis konteks mempertimbangkan seluruh lingkungan bisnis, lingkungan internal dan eksternal. Ini adalah aspek penting dari perencanaan bisnis. Satu jenis analisis konteks, yang disebut analisis SWOT, memungkinkan bisnis untuk mendapatkan wawasan tentang kekuatan dan kelemahan mereka dan juga peluang dan ancaman yang ditimbulkan oleh pasar tempat mereka beroperasi. Tujuan utama dari analisis konteks, SWOT atau yang lain, adalah untuk menganalisis lingkungan untuk mengembangkan rencana aksi strategis untuk bisnis.

Analisis konteks juga merujuk pada metode analisis sosiologis yang terkait dengan Scheflen (1963) yang meyakini bahwa 'tindakan tertentu, baik sekilas pada orang lain, perubahan postur, atau komentar tentang cuaca, tidak memiliki makna intrinsik . Tindakan semacam itu hanya dapat dipahami ketika dilakukan dalam kaitannya dengan satu sama lain. '(Kendon, 1990: 16). Ini tidak dibahas di sini; hanya Analisis Konteks dalam pengertian bisnis.

Menentukan Pasar atau Subjek

Langkah pertama dari metode ini adalah mendefinisikan pasar (atau subjek) tertentu yang ingin dianalisis dan memfokuskan semua teknik analisis pada apa yang telah didefinisikan. Subjek, misalnya, dapat berupa ide produk yang baru diusulkan.

Analisis Tren

Langkah selanjutnya dari metode ini adalah melakukan analisis tren. Analisis tren adalah analisis faktor lingkungan makro di lingkungan eksternal bisnis, juga disebut analisis PEST. Ini terdiri dari menganalisis tren politik, ekonomi, sosial, teknologi dan demografi. Ini dapat dilakukan dengan terlebih dahulu menentukan faktor mana, pada setiap level, yang relevan untuk subjek yang dipilih dan untuk menilai setiap item untuk menentukan kepentingannya. Ini memungkinkan bisnis untuk mengidentifikasi faktor-faktor yang dapat memengaruhi mereka. Mereka tidak dapat mengendalikan faktor-faktor ini tetapi mereka dapat mencoba mengatasinya dengan menyesuaikan diri. Tren (faktor) yang dibahas dalam analisis PEST adalah Politik, Ekonomi, Sosial dan Teknologi; tetapi untuk analisis konteks, tren demografis juga penting. Tren demografis adalah faktor-faktor yang berkaitan dengan populasi, seperti misalnya usia rata-rata, agama, pendidikan, dll. Informasi demografis sangat penting jika, misalnya selama riset pasar, suatu bisnis ingin menentukan segmen pasar tertentu untuk ditargetkan. Tren lain dijelaskan dalam pemindaian lingkungan dan analisis PEST. Analisis tren hanya mencakup sebagian dari lingkungan eksternal. Aspek penting lain dari lingkungan eksternal yang harus dipertimbangkan oleh bisnis adalah pesaingnya. Ini adalah langkah selanjutnya dari metode ini, analisis pesaing.

Analisis Pesaing

Seperti yang dapat dibayangkan, penting bagi bisnis untuk mengetahui siapa pesaingnya, bagaimana mereka melakukan bisnis mereka dan seberapa kuat mereka sehingga mereka dapat bertahan dan tersinggung. Dalam analisis Pesaing beberapa teknik diperkenalkan bagaimana melakukan analisis tersebut. Di sini saya akan memperkenalkan teknik lain yang melibatkan melakukan empat sub analisis, yaitu: penentuan tingkat kompetisi, kekuatan kompetitif, perilaku pesaing dan strategi pesaing.

Tingkat Persaingan

Bisnis bersaing di beberapa level dan penting bagi mereka untuk menganalisis level ini sehingga mereka dapat memahami permintaan. Persaingan diidentifikasi pada empat tingkatan:

- Kebutuhan konsumen: tingkat persaingan yang mengacu pada kebutuhan dan keinginan konsumen. Sebuah bisnis harus bertanya: Apa keinginan konsumen?
- Persaingan umum: Jenis permintaan konsumen. Misalnya: apakah konsumen lebih suka bercukur dengan pisau cukur listrik atau pisau cukur?
- Merek: Tingkat ini mengacu pada persaingan merek. Merek mana yang lebih disukai konsumen?

- Produk: Tingkat ini mengacu pada jenis permintaan. Jadi, jenis produk apa yang disukai konsumen?

Aspek penting lain dari analisis persaingan adalah meningkatkan wawasan konsumen. Sebagai contoh: [Ducati] telah, dengan mewawancarai banyak pelanggan mereka, menyimpulkan bahwa pesaing utama mereka bukanlah sepeda lain, tetapi mobil sport seperti [Porsche] atau [GM]. Ini tentu saja akan mempengaruhi tingkat persaingan dalam bisnis ini.

Kekuatan Kompetitif

Ini adalah kekuatan yang menentukan tingkat persaingan dalam pasar tertentu. Ada enam kekuatan yang harus dipertimbangkan, kekuatan persaingan, ancaman pendatang baru, daya tawar pembeli dan pemasok, ancaman produk pengganti dan pentingnya produk pelengkap. Analisis ini dijelaskan dalam analisis pasukan Porter 5.

Perilaku Pesaing

Perilaku pesaing adalah tindakan defensif dan ofensif kompetisi.

Strategi Pesaing

Strategi-strategi ini merujuk pada bagaimana suatu organisasi bersaing dengan organisasi lain. Dan ini adalah: strategi harga rendah dan strategi diferensiasi produk.

Peluang dan Ancaman

Langkah selanjutnya, setelah analisis tren dan analisis pesaing dilakukan, adalah menentukan ancaman dan peluang yang ditimbulkan oleh pasar. Analisis tren mengungkapkan serangkaian tren yang dapat memengaruhi bisnis baik secara positif maupun negatif. Dengan demikian, ini dapat diklasifikasikan sebagai peluang atau ancaman. Demikian juga, analisis pesaing mengungkapkan masalah persaingan positif dan negatif yang dapat diklasifikasikan sebagai peluang atau ancaman.

Analisis Organisasi

Fase terakhir dari metode ini adalah analisis lingkungan internal organisasi, dengan demikian organisasi itu sendiri. Tujuannya adalah untuk menentukan keterampilan, pengetahuan, dan teknologi yang dimiliki bisnis. Ini memerlukan melakukan analisis internal dan analisis kompetensi.

Analisis Internal

Analisis internal, juga disebut analisis SWOT, melibatkan pengidentifikasian kekuatan dan kelemahan organisasi. Kekuatan mengacu pada faktor-faktor yang dapat menghasilkan keunggulan pasar dan kelemahan pada faktor-faktor yang memberikan kerugian karena bisnis tidak dapat memenuhi kebutuhan pasar.

Analisis Kompetensi

Kompetensi adalah kombinasi dari pengetahuan, keterampilan, dan teknologi bisnis yang dapat memberi mereka keunggulan dibandingkan pesaing. Melakukan analisis semacam itu melibatkan mengidentifikasi kompetensi terkait pasar, kompetensi terkait integritas, dan kompetensi terkait fungsional.

Matriks SWOT-i

Bagian sebelumnya menjelaskan langkah-langkah utama yang terlibat dalam analisis konteks. Semua langkah ini menghasilkan data yang dapat digunakan untuk mengembangkan strategi. Ini dirangkum dalam matriks SWOT-i. Analisis tren dan pesaing mengungkapkan peluang dan ancaman yang ditimbulkan oleh pasar. Analisis organisasi mengungkapkan kompetensi organisasi dan juga kekuatan dan kelemahannya. Kekuatan, kelemahan, peluang dan ancaman ini merangkum seluruh analisis konteks. Matriks SWOT-i, digambarkan dalam tabel di bawah, digunakan untuk menggambarkan ini dan untuk membantu memvisualisasikan strategi yang akan disusun. SWOT-i adalah singkatan dari Strengths, Weaknesses, Opportunities, Threats and Issues. Masalah mengacu pada masalah strategis yang akan digunakan untuk menyusun rencana strategis.

	Opportunities (O_1, O_2, ..., O_n)	**Threats (T_1, T_2, ..., T_n)**
Strengths (S_1, S_2, ..., S_n)	$S_1O_1 \dots S_nO_1$ ... $S_1O_n \dots S_nO_n$	$S_1T_1 \dots S_nT_1$ ... $S_1T_n \dots S_nT_n$
Weaknesses (W_1, W_2, ..., W_n)	$W_1O_1 \dots W_nO_1$ ... $W_1O_n \dots W_nO_n$	$W_1T_1 \dots W_nT_1$ ... $W_1T_n \dots W_nT_n$

Matriks ini menggabungkan kekuatan dengan peluang dan ancaman, dan kelemahan dengan peluang dan ancaman yang diidentifikasi selama analisis. Dengan demikian matriks mengungkapkan empat kelompok:

- Kekuatan dan peluang klaster: gunakan kekuatan untuk memanfaatkan peluang.

- Kekuatan dan ancaman cluster: gunakan kekuatan untuk mengatasi ancaman
- Kelemahan dan peluang klaster: kelemahan tertentu menghambat organisasi untuk mengambil keuntungan dari peluang, oleh karena itu mereka harus mencari cara untuk mengubah kelemahan itu.
- Kelemahan dan ancaman kelompok: tidak ada cara bagi organisasi untuk mengatasi ancaman tanpa harus membuat perubahan besar.

Rencana Strategis

Tujuan akhir dari analisis konteks adalah untuk mengembangkan rencana strategis. Bagian sebelumnya menjelaskan semua langkah yang membentuk batu loncatan untuk mengembangkan rencana aksi strategis untuk organisasi. Tren dan analisis pesaing memberikan wawasan tentang peluang dan ancaman di pasar dan analisis internal memberikan wawasan tentang kompetensi organisasi. . Dan ini digabungkan dalam matriks SWOT-i. Matriks SWOT-i membantu mengidentifikasi masalah yang perlu ditangani. Masalah-masalah ini perlu diselesaikan dengan merumuskan tujuan dan rencana untuk mencapai tujuan itu, sebuah strategi.

Studi Kasus

Anton Sanjaya sedang dalam proses menulis rencana bisnis untuk ide bisnisnya, Jaya Systems. Jaya Systems akan menjadi bisnis perangkat lunak yang berfokus pada pengembangan perangkat lunak untuk bisnis kecil. Anton menyadari bahwa ini adalah pasar yang sulit karena ada banyak perusahaan perangkat lunak yang mengembangkan perangkat lunak bisnis. Oleh karena itu, ia melakukan analisis konteks untuk mendapatkan wawasan tentang lingkungan bisnis untuk mengembangkan rencana aksi strategis untuk mencapai keunggulan kompetitif dalam pasar.

Tentukan Pasar

Langkah pertama adalah mendefinisikan pasar untuk analisis. Anton memutuskan bahwa ia ingin fokus pada bisnis kecil yang terdiri dari paling banyak 20 karyawan.

Analisis Tren

Langkah selanjutnya adalah melakukan analisis tren. Faktor-faktor lingkungan makro yang harus dipertimbangkan Anton adalah sebagai berikut:

- Tren politik: Hak kekayaan intelektual
- Tren ekonomi: Pertumbuhan ekonomi

- Tren sosial: Mengurangi biaya operasional; Kemudahan untuk melakukan administrasi bisnis
- Tren teknologi: Perangkat lunak suite; Aplikasi web
- Tren demografis: Peningkatan lulusan studi terkait TI

Analisis pesaing

Analisis tren berikut adalah analisis pesaing. Anton menganalisis persaingan pada empat level untuk mendapatkan wawasan tentang bagaimana mereka beroperasi dan di mana letak keunggulannya.

- Tingkat kompetisi:
- Kebutuhan konsumen: Jaya Systems akan bersaing pada kenyataan bahwa konsumen menginginkan penyelenggaraan bisnis yang efisien dan efektif
- Merek: Ada bisnis perangkat lunak yang telah membuat perangkat lunak bisnis untuk sementara waktu dan dengan demikian telah menjadi sangat populer di pasar. Bersaing berdasarkan merek akan sulit.
- Produk: Mereka akan dikemas perangkat lunak seperti kompetisi utama.
- Kekuatan kompetitif: Pasukan yang dapat memengaruhi Sistem Jaya khususnya:
- Daya tawar pembeli: sejauh mana mereka dapat beralih dari satu produk ke produk lainnya.
- Ancaman pendatang baru: sangat mudah bagi seseorang untuk mengembangkan produk perangkat lunak baru yang bisa lebih baik daripada Jaya.
- Kekuatan persaingan: para pemimpin pasar memiliki sebagian besar uang tunai dan pelanggan; mereka harus berkuasa untuk membentuk pasar.
- Perilaku pesaing: Fokus kompetisi adalah untuk mengambil alih posisi pemimpin pasar.
- Strategi pesaing: Antonbermaksud bersaing berdasarkan diferensiasi produk.

Peluang dan Ancaman

Sekarang Anton telah menganalisis persaingan dan tren di pasar, ia dapat mendefinisikan peluang dan ancaman.

Peluang:

- Karena pesaing berfokus untuk mengambil alih posisi kepemimpinan, Jaya dapat fokus pada segmen pasar yang diabaikan oleh pemimpin pasar. Ini memungkinkan mereka mengambil alih posisi pemimpin pasar yang menunjukkan kelemahan.

- Fakta bahwa ada lulusan TI baru, Jaya dapat mempekerjakan atau bermitra dengan seseorang yang mungkin memiliki ide cemerlang.

Ancaman:

- Lulusan IT dengan ide segar dapat memulai bisnis perangkat lunak mereka sendiri dan membentuk persaingan besar untuk Jaya Systems.

Analisis Organisasi

Setelah Antonmengidentifikasi peluang dan ancaman pasar, ia dapat mencoba mencari tahu apa kekuatan dan kelemahan Jaya System dengan melakukan analisis organisasi.

- Analisis internal:
- Kekuatan: Diferensiasi produk
- Kelemahan: Tidak memiliki orang yang inovatif dalam organisasi
- Analisis kompetensi:
- Kompetensi terkait fungsional: Jaya Systems menyediakan fungsionalitas sistem yang sesuai dengan bisnis kecil.
- Kompetensi terkait pasar: Jaya Systems memiliki kesempatan untuk fokus pada bagian pasar yang diabaikan.

Matriks SWOT-i

Setelah analisis sebelumnya, Antondapat membuat matriks SWOT-i untuk melakukan analisis SWOT.

Rencana Strategis

Setelah membuat matriks SWOT-i, Anton sekarang dapat menyusun rencana strategis.

- Fokuskan semua upaya pengembangan perangkat lunak ke bagian pasar yang diabaikan oleh para pemimpin pasar, usaha kecil.
- Mempekerjakan lulusan It inovatif baru-baru ini untuk merangsang inovasi dalam Jaya Systems.

6.2 Manajemen Kinerja Bisnis

Manajemen kinerja bisnis (*Business Performance Management*-BPM) adalah serangkaian manajemen kinerja dan proses analisis yang memungkinkan manajemen kinerja organisasi untuk mencapai satu atau lebih tujuan yang telah dipilih sebelumnya. Sinonim untuk “manajemen kinerja bisnis” termasuk “manajemen kinerja perusahaan (CPM)” dan “manajemen kinerja perusahaan”.

Manajemen kinerja bisnis terdapat dalam pendekatan manajemen proses bisnis.

Manajemen kinerja bisnis memiliki tiga kegiatan utama:

1. pemilihan sasaran,
2. konsolidasi informasi pengukuran yang relevan dengan kemajuan organisasi terhadap sasaran-sasaran ini, dan
3. intervensi yang dilakukan oleh manajer sehubungan dengan informasi ini dengan maksud untuk meningkatkan kinerja masa depan terhadap tujuan-tujuan ini.

Meskipun disajikan di sini secara berurutan, biasanya ketiga kegiatan akan berjalan secara bersamaan, dengan intervensi oleh manajer yang mempengaruhi pilihan tujuan, informasi pengukuran dipantau, dan kegiatan yang dilakukan oleh organisasi.

Karena kegiatan manajemen kinerja bisnis dalam organisasi besar sering melibatkan pengumpulan dan pelaporan volume data yang besar, banyak vendor perangkat lunak, terutama yang menawarkan alat intelijen bisnis, memasarkan produk yang dimaksudkan untuk membantu proses ini. Sebagai hasil dari upaya pemasaran ini, manajemen kinerja bisnis sering dipahami secara keliru sebagai kegiatan yang selalu bergantung pada sistem perangkat lunak untuk bekerja, dan banyak definisi manajemen kinerja bisnis secara eksplisit menyarankan perangkat lunak sebagai komponen yang pasti dari pendekatan tersebut.

Minat dalam manajemen kinerja bisnis dari komunitas perangkat lunak ini didorong oleh penjualan - “Area pertumbuhan terbesar dalam analisis BI operasional adalah di bidang manajemen kinerja bisnis.”

Sejak tahun 1992, manajemen kinerja bisnis telah sangat dipengaruhi oleh munculnya kerangka kerja balanced scorecard. Adalah umum bagi manajer untuk menggunakan kerangka kerja balanced scorecard untuk memperjelas tujuan organisasi, untuk mengidentifikasi cara melacaknya, dan untuk menyusun mekanisme dengan mana intervensi akan dipicu. Langkah-langkah ini sama dengan yang ditemukan di BPM, dan sebagai hasilnya, balanced scorecard sering digunakan sebagai dasar untuk aktivitas manajemen kinerja bisnis dengan organisasi.

Di masa lalu, pemilik telah berusaha untuk mendorong strategi turun dan melintasi organisasi mereka, mengubah strategi ini menjadi metrik yang dapat ditindaklanjuti dan

menggunakan analisis untuk mengekspos hubungan sebab dan akibat yang, jika dipahami, dapat memberikan wawasan tentang pengambilan keputusan.

Sejarah

Referensi ke manajemen kinerja non-bisnis muncul di The Art of War milik Sun Tzu. Sun Tzu mengklaim bahwa untuk berhasil dalam perang, seseorang harus memiliki pengetahuan penuh tentang kekuatan dan kelemahannya sendiri serta musuh-musuhnya. Kurangnya seperangkat pengetahuan mungkin menghasilkan kekalahan. Paralel antara tantangan dalam bisnis dan tantangan perang meliputi:

- mengumpulkan data - baik internal maupun eksternal
- pola dan makna yang tajam dalam data (menganalisis)
- menanggapi informasi yang dihasilkan

Sebelum dimulainya Era Informasi di akhir abad ke-20, perusahaan terkadang mengambil kesulitan untuk mengumpulkan data dari sumber-sumber non-otomatis dengan susah payah. Karena mereka tidak memiliki sumber daya komputasi untuk menganalisis data dengan baik, mereka sering membuat keputusan komersial terutama berdasarkan intuisi.

Ketika bisnis mulai mengotomatisasi semakin banyak sistem, semakin banyak data yang tersedia. Namun, pengumpulan sering tetap menjadi tantangan karena kurangnya infrastruktur untuk pertukaran data atau karena ketidaksesuaian antar sistem. Laporan tentang data yang dikumpulkan terkadang membutuhkan waktu berbulan-bulan untuk menghasilkan. Laporan tersebut memungkinkan pengambilan keputusan strategis jangka panjang yang terinformasi. Namun, pengambilan keputusan taktis jangka pendek sering terus mengandalkan intuisi.

Pada tahun 1989 Howard Dresner, seorang analis penelitian di Gartner, mempopulerkan “intelijen bisnis” (BI) sebagai istilah umum untuk menggambarkan serangkaian konsep dan metode untuk meningkatkan pengambilan keputusan bisnis dengan menggunakan sistem pendukung berbasis fakta. Manajemen kinerja dibangun di atas dasar BI, tetapi menggabungkannya dengan siklus perencanaan dan pengendalian perusahaan - dengan kemampuan perencanaan, konsolidasi, dan pemodelan perusahaan.

Peningkatan standar, otomatisasi, dan teknologi telah menyebabkan sejumlah besar data tersedia. Teknologi data warehouse telah memungkinkan pembangunan repositori untuk menyimpan data ini. Peningkatan ETL dan alat integrasi aplikasi perusahaan telah meningkatkan pengumpulan data yang tepat waktu. Teknologi pelaporan OLAP memungkinkan pembuatan laporan baru yang lebih cepat yang menganalisis data. Pada 2010, intelijen bisnis telah menjadi seni penyaringan melalui sejumlah besar data, mengekstraksi informasi yang berguna dan mengubah informasi itu menjadi pengetahuan yang dapat ditindaklanjuti.

Definisi dan Ruang Lingkup

Manajemen kinerja bisnis terdiri dari serangkaian proses manajemen dan analisis, didukung oleh teknologi, yang memungkinkan bisnis untuk menentukan tujuan strategis dan kemudian mengukur dan mengelola kinerja terhadap tujuan tersebut. Proses manajemen kinerja bisnis inti meliputi perencanaan keuangan, perencanaan operasional, pemodelan bisnis, konsolidasi dan pelaporan, analisis, dan pemantauan indikator kinerja utama yang terkait dengan strategi. Manajemen kinerja bisnis melibatkan konsolidasi data dari berbagai sumber, permintaan, dan analisis data, dan mempraktikkan hasilnya.

Kerangka kerja

Ada berbagai kerangka kerja untuk menerapkan manajemen kinerja bisnis. Disiplin memberi perusahaan kerangka kerja top-down yang digunakan untuk menyelaraskan perencanaan dan pelaksanaan, strategi dan taktik, dan tujuan unit bisnis dan perusahaan. Reaksi dapat mencakup strategi Six Sigma, balanced scorecard, penetapan biaya berbasis aktivitas (ABC), Tujuan dan Hasil Utama (OKR), Manajemen Kualitas Total, nilai tambah ekonomis, pengukuran strategis terintegrasi dan Teori Kendala.

Balanced scorecard adalah kerangka kerja manajemen kinerja yang paling banyak diadopsi.

Metrik dan Indikator Kinerja Utama

Beberapa area di mana manajemen bank dapat memperoleh pengetahuan dengan menggunakan manajemen kinerja bisnis meliputi:

- nomor yang berhubungan dengan pelanggan:
- pelanggan baru diperoleh

- status pelanggan yang ada
- gesekan pelanggan (termasuk putus karena alasan gesekan)
- omset yang dihasilkan oleh segmen pelanggan - mungkin menggunakan filter demografis
- saldo terutang yang dimiliki oleh segmen pelanggan dan ketentuan pembayaran - mungkin menggunakan filter demografis
- pengumpulan kredit macet dalam hubungan pelanggan
- analisis demografis individu (calon pelanggan) yang melamar menjadi pelanggan, dan tingkat persetujuan, penolakan, dan angka yang tertunda
- analisis kenakalan pelanggan di balik pembayaran
- profitabilitas pelanggan berdasarkan segmen demografis dan segmentasi pelanggan berdasarkan profitabilitas
- manajemen kampanye
- dasbor real-time pada metrik operasional utama
- efektivitas peralatan secara keseluruhan
- analisis clickstream di situs web
- pelacak portofolio produk utama
- analisis saluran pemasaran
- analisis data penjualan berdasarkan segmen produk
- metrik pusat panggilan

Meskipun daftar di atas menggambarkan apa yang bank dapat pantau, itu bisa merujuk ke perusahaan telepon atau ke perusahaan sektor jasa serupa.

Item-item yang penting secara umum meliputi:

1. data terkait Key Performance Indicator (KPI) yang konsisten dan benar memberikan wawasan tentang aspek operasional perusahaan
2. ketersediaan data terkait KPI tepat waktu
3. KPI yang dirancang untuk secara langsung mencerminkan efisiensi dan efektivitas bisnis
4. informasi yang disajikan dalam format yang membantu pengambilan keputusan untuk manajemen dan pengambil keputusan
5. kemampuan untuk membedakan pola atau tren dari informasi yang terorganisir

Manajemen kinerja bisnis mengintegrasikan proses perusahaan dengan CRM atau ERP. Perusahaan harus menjadi lebih mampu mengukur kepuasan pelanggan, mengendalikan tren pelanggan dan mempengaruhi nilai pemegang saham.

Jenis Perangkat Lunak Aplikasi

Orang yang bekerja dalam intelijen bisnis telah mengembangkan alat yang memudahkan pekerjaan manajemen kinerja bisnis, terutama ketika tugas intelijen bisnis melibatkan pengumpulan dan analisis sejumlah besar data yang tidak terstruktur.

Kategori alat yang biasa digunakan untuk manajemen kinerja bisnis meliputi:

- MOLAP — Multidimensional online analytical processing, terkadang hanya disebut "analitik" (berdasarkan dimensi analisis dan yang disebut "hypercube" atau "cube") scorecarding, dashboarding and data visualization
- data warehouses
- document warehouses
- text mining
- DM — data mining
- BPO — business performance optimization
- EPM — enterprise performance management
- EIS — executive information systems
- DSS — decision support systems
- MIS — management information systems
- SEMS — strategic enterprise management software
- EOI — Operational intelligence Enterprise Operational Intelligence Software

Desain dan Implementasi

Pertanyaan yang diajukan ketika menerapkan program manajemen kinerja bisnis meliputi:

Query-alignment queries

Tentukan tujuan program jangka pendek dan menengah. Apa tujuan strategis organisasi yang akan dituju oleh program? Misi / visi organisasi apa yang terkait dengannya? Hipotesis perlu dibuat yang merinci bagaimana inisiatif ini pada akhirnya akan meningkatkan hasil / kinerja (yaitu peta strategi).

Baseline queries

Menilai kompetensi pengumpulan-informasi saat ini. Apakah organisasi memiliki kemampuan untuk memonitor sumber informasi penting? Data apa yang sedang dikumpulkan dan bagaimana cara menyimpannya? Apa parameter statistik dari data

ini, mis., Berapa banyak variasi acak yang dikandungnya? Apakah ini sedang diukur?

Kueri biaya dan risiko

Perkirakan konsekuensi keuangan dari inisiatif BI baru. Menilai biaya operasi saat ini dan peningkatan biaya yang terkait dengan inisiatif BPM. Apa risiko bahwa inisiatif akan gagal? Penilaian risiko ini harus dikonversi menjadi metrik keuangan dan dimasukkan dalam perencanaan.

Permintaan pelanggan dan pemangku kepentingan

Tentukan siapa yang akan mendapat manfaat dari inisiatif dan siapa yang akan membayar. Siapa yang memiliki kepentingan dalam prosedur saat ini? Pelanggan / pemangku kepentingan seperti apa yang akan mendapat manfaat langsung dari inisiatif ini? Siapa yang akan mendapat manfaat secara tidak langsung? Apa manfaat kuantitatif / kualitatif yang diikuti? Apakah inisiatif yang ditentukan merupakan cara terbaik atau satu-satunya untuk meningkatkan kepuasan bagi semua jenis pelanggan? Bagaimana manfaat pelanggan akan dipantau? Bagaimana dengan karyawan, pemegang saham, dan anggota saluran distribusi?

Kueri terkait metrik

Persyaratan informasi memerlukan operasionalisasi ke dalam metrik yang jelas. Putuskan metrik mana yang akan digunakan untuk setiap informasi yang dikumpulkan. Apakah ini metrik terbaik dan mengapa? Berapa banyak metrik yang perlu dilacak? Jika ini jumlah yang besar (biasanya), sistem seperti apa yang dapat melacaknya? Apakah metrik terstandarisasi, sehingga dapat dibandingkan dengan kinerja di organisasi lain? Apa metrik standar industri yang tersedia?

Pertanyaan terkait metodologi pengukuran

Tetapkan metodologi atau prosedur untuk menentukan cara terbaik (atau dapat diterima) untuk mengukur metrik yang diperlukan. Seberapa sering data akan dikumpulkan? Apakah ada standar industri untuk ini? Apakah ini cara terbaik untuk melakukan pengukuran? Bagaimana kita tahu itu?

Kueri terkait hasil

Pantau program BPM untuk memastikan bahwa ia memenuhi tujuan. Program itu sendiri mungkin memerlukan penyesuaian. Program harus diuji untuk keakuratan, keandalan, dan validitas. Bagaimana bisa ditunjukkan bahwa inisiatif BI, dan bukan sesuatu yang lain, berkontribusi pada perubahan hasil? Berapa banyak perubahan itu mungkin acak?

6.2 Business Process Discovery

Business Process Discovery (BPD) terkait dengan proses penambangan adalah seperangkat teknik yang secara otomatis membangun representasi dari proses bisnis organisasi saat ini dan variasi proses utamanya. Teknik-teknik ini menggunakan bukti yang ditemukan dalam sistem teknologi yang ada yang menjalankan proses bisnis dalam suatu organisasi.

Teknik *Business Process Discovery*

Teknik *Business Process Discovery* mewujudkan properti berikut:

- Paradigma yang muncul - Metode saat ini didasarkan pada wawancara manual terstruktur top-down yang mengandalkan representasi dari proses bisnis / perilaku sistem. Proses penemuan otomatis bergantung pada pengumpulan data dari sistem informasi selama periode waktu tertentu. Data ini kemudian dapat dianalisis untuk membentuk model proses.
- Penemuan proses otomatis - Dengan mengotomatisasi analisis data, subjektivitas teknik analisis proses manual saat ini dihapus. Sistem otomatis memiliki metodologi yang tertanam bahwa - melalui uji coba berulang - telah terbukti secara akurat menemukan proses dan memproses variasi tanpa bias.
- Informasi yang akurat- Karena informasi dikumpulkan dari sumber yang sebenarnya, maka informasi itu tidak akurat, dan bukan dikumpulkan dari perwakilan pihak kedua.
- Informasi lengkap - Suatu proses otomatis menangkap semua informasi yang terjadi di dalam sistem dan merepresentasikannya berdasarkan waktu, tanggal, pengguna, dll. Karena informasi tersebut dikumpulkan dari interaksi real-time, itu tidak dapat hilang atau hilang. masalah memori selektif. Ini termasuk kelengkapan tentang pengecualian dalam proses. Seringkali, pengecualian diperlakukan sebagai “kebisingan” statistik, yang dapat mengecualikan inefisiensi penting dalam proses bisnis.
- Proses Standar - Pengumpulan otomatis informasi menghasilkan data proses yang dapat dikelompokkan, dikuantifikasi dan diklasifikasikan. Ini memasok dasar untuk pengembangan dan pemantauan proses saat ini dan baru, di mana tolok ukur dapat ditetapkan. Tolok ukur ini adalah akar dari desain proses baru dan penentuan akar masalah. Selain itu, data proses terstandarisasi dapat mengatur tahapan untuk upaya peningkatan proses yang berkelanjutan.

Aplikasi / Teknik

Business Process Discovery melengkapi dan dibangun di atas pekerjaan di banyak bidang lainnya.

- *Process Discovery* adalah salah satu dari tiga jenis utama penambangan proses. Dua jenis proses penambangan lainnya adalah pengecekan kesesuaian dan perluasan / peningkatan model. Semua teknik ini bertujuan mengekstraksi pengetahuan terkait proses dari log peristiwa. Dalam hal penemuan proses, tidak ada model proses sebelumnya; model ditemukan berdasarkan log peristiwa. Pemeriksaan kesesuaian bertujuan untuk menemukan perbedaan antara model proses yang diberikan dan log peristiwa. Dengan cara ini dimungkinkan untuk mengukur kepatuhan dan menganalisis perbedaan. Peningkatan menggunakan model apriori dan meningkatkan atau memperluasnya menggunakan informasi dari log peristiwa, mis., Menunjukkan kemacetan.
- *Business Process Discovery*adalah tingkat pemahaman berikutnya dalam bidang analisis bisnis yang muncul, yang memungkinkan organisasi untuk melihat, menganalisis, dan menyesuaikan struktur dan proses yang mendasarinya yang masuk ke dalam operasi sehari-hari. Penemuan ini mencakup pengumpulan informasi semua komponen proses bisnis, termasuk teknologi, orang, prosedur departemen, dan protokol.
- *Business Process Discovery* menciptakan master proses yang melengkapi analisis proses bisnis (BPA). Alat dan metodologi BPA sangat cocok untuk dekomposisi proses hierarki top-down, dan analisis proses yang akan dilakukan. BPD menyediakan analisis bottoms-up yang menikah dengan top-down untuk menyediakan proses bisnis yang lengkap, yang dikelola secara hierarkis oleh BPA.
- Intelejensi Bisnis memberi organisasi pelaporan dan analisis data di organisasi mereka. Namun, BI tidak memiliki model proses, kesadaran atau analisis. BPD melengkapi BI dengan memberikan pandangan proses eksplisit untuk operasi saat ini, dan memberikan analisis pada model proses untuk membantu organisasi mengidentifikasi dan bertindak atas ketidakefisienan proses bisnis, atau anomali.
- Analisis web adalah contoh terbatas BPD dalam analisis web yang merekonstruksi proses pengguna web saat mereka berinteraksi dengan situs web. Namun, analisis ini terbatas pada proses seperti yang terkandung dalam sesi, dari perspektif pengguna dan sehubungan dengan hanya sistem dan proses berbasis web.
- Triase bisnis menyediakan kerangka kerja untuk mengkategorikan proses yang diidentifikasi oleh analisis proses bisnis (BPA) berdasarkan kepentingan relatifnya untuk mencapai tujuan atau hasil yang dinyatakan dan diukur. Memanfaatkan kategori

yang sama yang digunakan oleh layanan medis medis dan bencana medis, proses bisnis dikategorikan sebagai:

- Esensial / kritis (proses merah) - Proses penting untuk mencapai hasil / sasaran
- Penting / mendesak (proses kuning) - Proses yang mempercepat pencapaian hasil / sasaran
- Opsional / suportif (proses hijau) - Proses tidak diperlukan untuk mencapai hasil / sasaran

Sumber daya dialokasikan berdasarkan kategori proses dengan sumber daya pertama didedikasikan untuk proses merah, kemudian proses kuning dan akhirnya proses hijau. Dalam hal sumber daya menjadi terbatas, sumber daya pertama-tama ditahan dari Proses Hijau, kemudian Proses Kuning. Sumber daya hanya ditahan dari Proses Merah jika kegagalan untuk mencapai hasil / tujuan dapat diterima.

Tujuan / Contoh

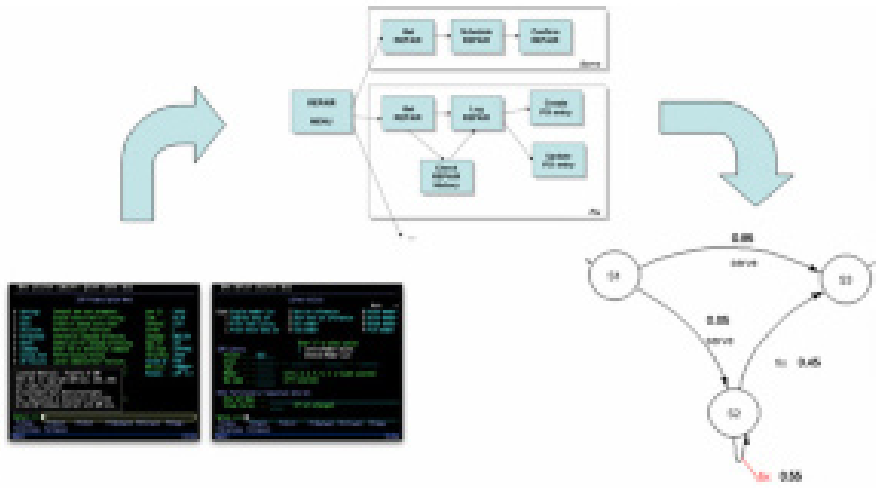
Contoh kecil dapat menggambarkan teknologi *Business Process Discovery* yang diperlukan saat ini. Alat *Business Process Discovery* Otomatis menangkap data yang diperlukan, dan mengubahnya menjadi dataset terstruktur untuk diagnosis aktual; Tantangan utama adalah pengelompokan tindakan berulang dari pengguna ke dalam peristiwa yang bermakna. Selanjutnya, alat penemuan proses bisnis ini mengusulkan model proses probabilistik. Perilaku probabilistik sangat penting untuk analisis dan diagnosis proses. Berikut ini menunjukkan contoh di mana proses perbaikan probabilistik pulih dari tindakan pengguna. Model proses “apa adanya” menunjukkan dengan tepat di mana rasa sakitnya dalam bisnis ini. Lima persen perbaikan yang salah adalah pertanda buruk, tetapi yang lebih buruk, perbaikan berulang yang diperlukan untuk menyelesaikan perbaikan itu rumit.

Sejarah

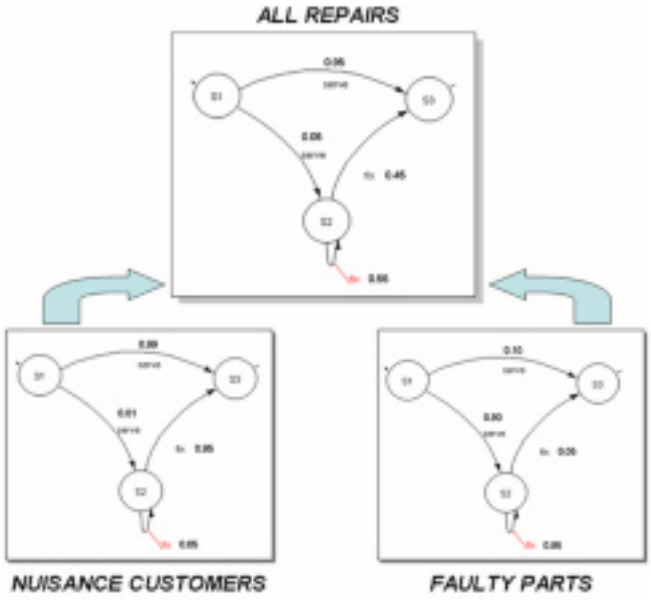
- Intelejensi Bisnis (BI) muncul lebih dari 20 tahun yang lalu dan sangat penting untuk melaporkan apa yang terjadi dalam sistem organisasi. Namun aplikasi BI saat ini dan teknologi penambangan data tidak selalu cocok untuk mengevaluasi tingkat detail yang diperlukan untuk menganalisis data yang tidak terstruktur dan dinamika manusia dalam proses bisnis.
- Six-Sigma dan pendekatan kuantitatif lainnya untuk perbaikan proses bisnis telah digunakan selama lebih dari satu dekade dengan berbagai tingkat keberhasilan.

Keterbatasan utama untuk keberhasilan pendekatan ini adalah ketersediaan data yang akurat untuk menjadi dasar analisis. Dengan BPD, banyak organisasi enam-sigma menemukan kemampuan untuk memperluas analisis mereka ke dalam proses bisnis utama secara efektif.

- Proses penambangan Menurut para peneliti di Universitas Teknologi Eindhoven, (PM) muncul sebagai disiplin ilmu sekitar tahun 1990 ketika teknik seperti algoritma Alpha memungkinkan untuk mengekstraksi model proses (biasanya direpresentasikan sebagai Petri nets) dari log peristiwa. Namun, pengakuan akan disiplin ilmiah calon ini sangat terbatas di beberapa negara. Ketika hype Proses Penambangan yang dilakukan oleh Universitas Teknologi Eindhoven semakin berkembang, semakin banyak kritik yang muncul yang menunjukkan bahwa Proses Penambangan tidak lebih dari sekumpulan algoritma yang memecahkan masalah bisnis yang spesifik dan sederhana: penemuan proses bisnis dan metode evaluasi tambahan. Saat ini, ada lebih dari 100 algoritma penambangan proses yang dapat menemukan model proses yang juga mencakup concurrency, mis., Teknik penemuan proses genetik, algoritma penambangan heuristik, algoritma penambangan berbasis wilayah, dan algoritma penambangan fuzzy.



Analisis yang *lebih dalam dari data proses "apa adanya" dapat mengungkapkan bagian mana* yang salah *yang bertanggung jawab atas perilaku keseluruhan dalam contoh ini. Ini mungkin* mengarah pada *penemuan subkelompok perbaikan yang sebenarnya membutuhkan fokus* manajemen untuk *perbaikan.*



Dalam hal *ini, akan menjadi jelas bahwa bagian yang salah juga bertanggung jawab atas* perbaikan berulang. *Aplikasi serupa telah didokumentasikan, seperti kasus Penyedia Asuransi* Kesehatan di *mana dalam 4 bulan ROI Analisis Proses Bisnis diperoleh dari pemahaman yang* tepat tentang *proses penanganan klaim dan menemukan bagian yang salah.*

6.3 Sistem Informasi

Sistem informasi (SI) adalah sistem terorganisir untuk pengumpulan, pengorganisasian, penyimpanan, dan komunikasi informasi. Lebih khusus lagi, ini adalah studi tentang jaringan pelengkap yang digunakan orang dan organisasi untuk mengumpulkan, memfilter, memproses, membuat, dan mendistribusikan data.

"Information System (IS) adalah sekelompok komponen yang berinteraksi untuk menghasilkan informasi"

Sistem informasi komputer adalah sistem yang terdiri dari orang dan komputer yang memproses atau menginterpretasikan informasi. Istilah ini juga terkadang digunakan dalam pengertian yang lebih terbatas untuk merujuk hanya pada perangkat lunak yang digunakan untuk menjalankan basis data yang terkomputerisasi atau untuk merujuk hanya pada sistem komputer.

Sistem informasi adalah studi akademis tentang sistem dengan referensi khusus untuk informasi dan jaringan pelengkap perangkat keras dan perangkat lunak yang digunakan orang dan organisasi untuk mengumpulkan, menyaring, memproses, membuat, dan juga

mendistribusikan data. Penekanan ditempatkan pada Sistem Informasi yang memiliki Batas definitif, Pengguna, Pemroses, Toko, Input, Output dan jaringan komunikasi yang disebutkan di atas.

Setiap sistem informasi spesifik bertujuan untuk mendukung operasi, manajemen dan pengambilan keputusan. Sistem informasi adalah teknologi informasi dan komunikasi (TIK) yang digunakan organisasi, dan juga cara orang berinteraksi dengan teknologi ini dalam mendukung proses bisnis.

Beberapa penulis membuat perbedaan yang jelas antara sistem informasi, sistem komputer, dan proses bisnis. Sistem informasi biasanya memasukkan komponen TIK tetapi tidak sepenuhnya berkaitan dengan TIK, sebaliknya berfokus pada penggunaan akhir teknologi informasi. Sistem informasi juga berbeda dari proses bisnis. Sistem informasi membantu mengendalikan kinerja proses bisnis.

Alter berpendapat keuntungan melihat sistem informasi sebagai jenis khusus sistem kerja. Sistem kerja adalah sistem di mana manusia atau mesin melakukan proses dan aktivitas menggunakan sumber daya untuk menghasilkan produk atau layanan tertentu untuk pelanggan. Sistem informasi adalah sistem kerja yang kegiatannya dikhususkan untuk menangkap, mentransmisikan, menyimpan, mengambil, memanipulasi dan menampilkan informasi.

Dengan demikian, sistem informasi saling berhubungan dengan sistem data di satu sisi dan sistem aktivitas di sisi lain. Sistem informasi adalah bentuk sistem komunikasi di mana data mewakili dan diproses sebagai bentuk memori sosial. Sistem informasi juga dapat dianggap sebagai bahasa semi formal yang mendukung pengambilan keputusan dan tindakan manusia.

Sistem informasi adalah fokus utama studi untuk informatika organisasi.

Gambaran

Silver dkk (1995) memberikan dua pandangan tentang IS yang meliputi perangkat lunak, perangkat keras, data, orang, dan prosedur. Zheng memberikan pandangan sistem lain tentang sistem informasi yang juga menambahkan proses dan elemen sistem penting seperti lingkungan, batas, tujuan, dan interaksi. Asosiasi Mesin Komputasi mendefinisikan “Spesialis sistem informasi [sebagai] fokus pada mengintegrasikan solusi teknologi

informasi dan proses bisnis untuk memenuhi kebutuhan informasi bisnis dan perusahaan lain."

Ada berbagai jenis sistem informasi, misalnya: sistem pemrosesan transaksi, sistem pendukung keputusan, sistem manajemen pengetahuan, sistem manajemen pembelajaran, sistem manajemen basis data, dan sistem informasi kantor. Yang paling penting bagi sebagian besar sistem informasi adalah teknologi informasi, yang biasanya dirancang untuk memungkinkan manusia melakukan tugas yang tidak cocok dengan otak manusia, seperti: menangani informasi dalam jumlah besar, melakukan perhitungan yang rumit, dan mengendalikan banyak proses simultan.

Teknologi informasi adalah sumber daya yang sangat penting dan dapat ditempa yang tersedia untuk para eksekutif. Banyak perusahaan telah menciptakan posisi chief information officer (CIO) yang duduk di dewan eksekutif dengan chief executive officer (CEO), chief financial officer (CFO), chief operating officer (COO), dan chief technical officer (CTO). CTO juga dapat berfungsi sebagai CIO, dan sebaliknya. Kepala petugas keamanan informasi (CISO) berfokus pada manajemen keamanan informasi.

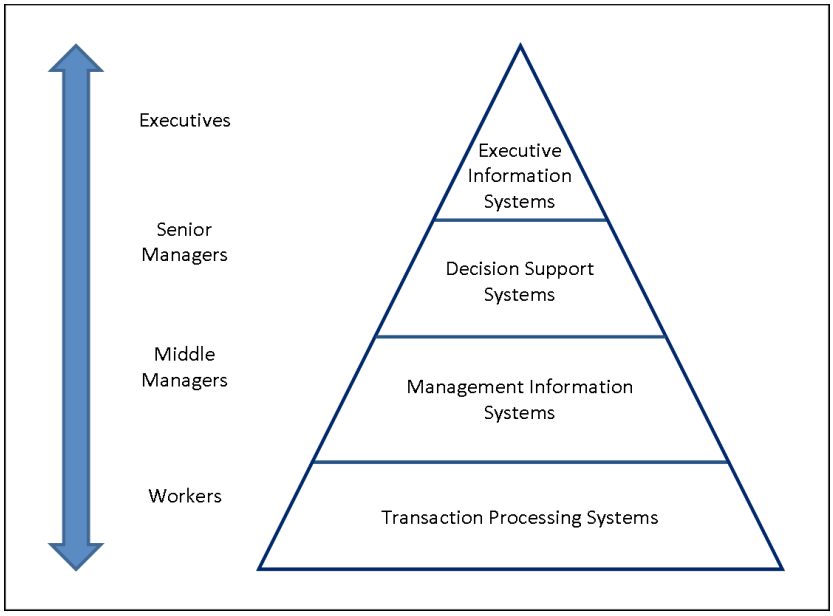
Enam komponen yang harus disatukan untuk menghasilkan sistem informasi adalah:

1. Perangkat Keras: Istilah perangkat keras mengacu pada mesin. Kategori ini mencakup komputer itu sendiri, yang sering disebut sebagai central processing unit (CPU), dan semua peralatan pendukungnya. Di antara peralatan pendukung adalah perangkat input dan output, perangkat penyimpanan dan perangkat komunikasi.
2. Perangkat Lunak: Istilah perangkat lunak mengacu pada program komputer dan manual (jika ada) yang mendukungnya. Program komputer adalah instruksi yang dapat dibaca mesin yang mengarahkan sirkuit di dalam bagian perangkat keras sistem untuk berfungsi dengan cara yang menghasilkan informasi yang berguna dari data. Program umumnya disimpan pada beberapa media input / output, seringkali disk atau tape.
3. Data: Data adalah fakta yang digunakan oleh program untuk menghasilkan informasi yang bermanfaat. Seperti halnya program, data umumnya disimpan dalam bentuk yang dapat dibaca mesin pada disk atau kaset sampai komputer membutuhkannya.
4. Prosedur: Prosedur adalah kebijakan yang mengatur pengoperasian sistem komputer. "Prosedur adalah untuk orang-orang apa perangkat lunak untuk perangkat keras" adalah analogi umum yang digunakan untuk menggambarkan peran prosedur dalam suatu sistem.

5. People: Setiap sistem membutuhkan orang jika itu berguna. Seringkali elemen yang paling kelihatan dari sistem adalah orang-orang, mungkin komponen yang paling mempengaruhi keberhasilan atau kegagalan sistem informasi. Ini termasuk “tidak hanya pengguna, tetapi mereka yang mengoperasikan dan melayani komputer, mereka yang memelihara data, dan mereka yang mendukung jaringan komputer.” <Kroenke, D. M. (2015). Essential MIS. Pendidikan Pearson>
6. Umpan balik: itu adalah komponen lain dari IS, yang mendefinisikan bahwa IS dapat diberikan dengan umpan balik (Meskipun komponen ini tidak diperlukan untuk berfungsi).

Data adalah jembatan antara perangkat keras dan manusia. Ini berarti bahwa data yang kami kumpulkan hanya data, sampai kami melibatkan orang. Pada titik itu, data sekarang menjadi informasi.

Jenis-jenis Sistem Informasi



Tingkat empat

Pandangan “klasik” dari sistem Informasi yang ditemukan dalam buku teks pada 1980-an adalah piramida sistem yang mencerminkan hierarki organisasi, biasanya sistem pemrosesan transaksi di bagian bawah piramida, diikuti oleh sistem informasi manajemen, sistem pendukung keputusan, dan diakhiri dengan sistem informasi eksekutif di bagian atas. Meskipun model piramida tetap bermanfaat, sejak pertama kali dirumuskan sejumlah

teknologi baru telah dikembangkan dan kategori baru sistem informasi telah muncul, beberapa di antaranya tidak lagi cocok dengan mudah ke dalam model piramida asli.

Beberapa contoh sistem tersebut adalah:

- gudang data
- Perencanaan Sumberdaya Perusahaan
- sistem perusahaan
- sistem pakar
- mesin pencari
- sistem Informasi Geografis
- sistem informasi global
- otomatisasi kantor

Sistem informasi berbasis komputer pada dasarnya adalah SI yang menggunakan teknologi komputer untuk melaksanakan beberapa atau semua tugas yang direncanakan. Komponen dasar dari sistem informasi berbasis komputer adalah:

- Perangkat Keras - ini adalah perangkat seperti monitor, prosesor, printer dan keyboard, yang semuanya bekerja bersama untuk menerima, memproses, menampilkan data dan informasi.
- Perangkat Lunak - adalah program yang memungkinkan perangkat keras untuk memproses data.
- Database- adalah pengumpulan file atau tabel terkait yang berisi data terkait.
- Jaringan - adalah sistem penghubung yang memungkinkan beragam komputer untuk mendistribusikan sumber daya.
- Prosedur - adalah perintah untuk menggabungkan komponen di atas untuk memproses informasi dan menghasilkan output yang disukai.

Empat komponen pertama (perangkat keras, perangkat lunak, database, dan jaringan) membentuk apa yang dikenal sebagai platform teknologi informasi. Pekerja teknologi informasi kemudian dapat menggunakan komponen-komponen ini untuk membuat sistem informasi yang mengawasi langkah-langkah keselamatan, risiko dan pengelolaan data. Tindakan ini dikenal sebagai layanan teknologi informasi.

Sistem informasi tertentu mendukung bagian-bagian organisasi, yang lain mendukung seluruh organisasi, dan yang lain lagi, mendukung kelompok-kelompok organisasi. Ingatlah bahwa setiap departemen atau area fungsional dalam suatu organisasi memiliki

koleksi program aplikasi, atau sistem informasi sendiri. Sistem informasi area fungsional (*functional area information systems*-FAIS) ini merupakan pilar pendukung untuk IS yang lebih umum yaitu, sistem intelijen bisnis dan dasbor. Seperti namanya, setiap FAIS mendukung fungsi tertentu dalam organisasi, mis .: IS akuntansi, IS keuangan, IS produksi / operasi, IS pemasaran, IS sumber daya manusia. Dalam bidang keuangan dan akuntansi, manajer menggunakan sistem TI untuk meramalkan pendapatan dan aktivitas bisnis, untuk menentukan sumber dan penggunaan dana terbaik, dan untuk melakukan audit untuk memastikan bahwa organisasi secara fundamental sehat dan bahwa semua laporan keuangan dan dokumen akurat. Jenis lain dari sistem informasi organisasi adalah FAIS, sistem pemrosesan transaksi, perencanaan sumber daya perusahaan, sistem otomasi kantor, sistem informasi manajemen, sistem pendukung keputusan, sistem pakar, dasbor eksekutif, sistem manajemen rantai pasokan, dan sistem perdagangan elektronik. Dasbor adalah bentuk khusus IS yang mendukung semua manajer organisasi. Mereka menyediakan akses cepat ke informasi tepat waktu dan akses langsung ke informasi terstruktur dalam bentuk laporan. Sistem pakar berupaya menduplikasi pekerjaan pakar manusia dengan menerapkan kemampuan penalaran, pengetahuan, dan keahlian dalam domain tertentu.

Pengembangan Sistem Informasi

Departemen teknologi informasi dalam organisasi yang lebih besar cenderung sangat mempengaruhi pengembangan, penggunaan, dan penerapan teknologi informasi dalam organisasi. Serangkaian metodologi dan proses dapat digunakan untuk mengembangkan dan menggunakan sistem informasi. Banyak pengembang sekarang menggunakan pendekatan teknik seperti siklus hidup pengembangan sistem (SDLC), yang merupakan prosedur sistematis untuk mengembangkan sistem informasi melalui tahapan yang terjadi secara berurutan. Penelitian terbaru bertujuan untuk memungkinkan dan mengukur pengembangan kolektif berkelanjutan dari sistem tersebut dalam suatu organisasi oleh keseluruhan aktor manusia itu sendiri. Sistem informasi dapat dikembangkan secara internal (dalam organisasi) atau outsourcing. Ini dapat dicapai dengan melakukan outsourcing komponen tertentu atau seluruh sistem. Kasus spesifik adalah distribusi geografis dari tim pengembangan (offshoring, sistem informasi global).

Sistem informasi berbasis komputer, mengikuti definisi Langefors, adalah media yang diimplementasikan secara teknologi untuk:

- merekam, menyimpan, dan menyebarluaskan ekspresi linguistik,
- serta untuk menarik kesimpulan dari ekspresi seperti itu.

Sistem informasi geografis, sistem informasi pertanahan, dan sistem informasi bencana adalah contoh dari sistem informasi yang muncul, tetapi mereka dapat secara luas dianggap sebagai sistem informasi spasial. Pengembangan sistem dilakukan secara bertahap yang meliputi:

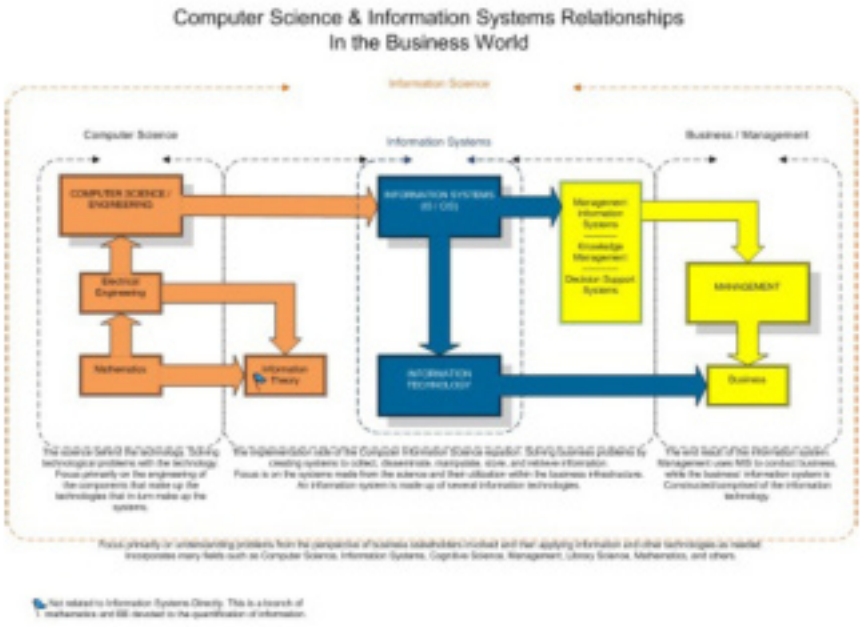
- Pengenalan masalah dan spesifikasi
- Pengumpulan informasi
- Spesifikasi persyaratan untuk sistem baru
- Desain sistem
- Konstruksi sistem
- Implementasi sistem
- Tinjauan dan pemeliharaan.

Sebagai Disiplin Akademik

Bidang studi yang disebut sistem informasi mencakup berbagai topik termasuk analisis dan desain sistem, jaringan komputer, keamanan informasi, manajemen basis data, dan sistem pendukung keputusan. Manajemen informasi berkaitan dengan masalah praktis dan teoritis dalam mengumpulkan dan menganalisis informasi dalam area fungsi bisnis termasuk alat produktivitas bisnis, pemrograman aplikasi dan implementasi, perdagangan elektronik, produksi media digital, penambangan data, dan dukungan keputusan. Komunikasi dan jaringan berhubungan dengan teknologi telekomunikasi. Sistem informasi menjembatani bisnis dan ilmu komputer menggunakan landasan teori informasi dan komputasi untuk mempelajari berbagai model bisnis dan proses algoritmik terkait dalam membangun sistem TI dalam disiplin ilmu komputer. Sistem informasi komputer (CIS) adalah bidang yang mempelajari komputer dan proses algoritmik, termasuk prinsip-prinsip mereka, desain perangkat lunak dan perangkat keras, aplikasi mereka, dan dampaknya terhadap masyarakat, sedangkan IS menekankan fungsi daripada desain.

Beberapa sarjana IS telah memperdebatkan sifat dan dasar Sistem Informasi yang berakar pada disiplin referensi lain seperti Ilmu Komputer, Teknik, Matematika, Ilmu Manajemen, Sibernetika, dan lainnya. Sistem informasi juga dapat didefinisikan sebagai kumpulan perangkat keras, perangkat lunak, data, orang dan prosedur yang bekerja bersama untuk menghasilkan informasi yang berkualitas.

Membedakan IS dari Disiplin Terkait



Hubungan Sistem Informasi dengan Teknologi Informasi, Ilmu Komputer, Ilmu Informasi, dan Bisnis.

Mirip dengan ilmu komputer, disiplin ilmu lain dapat dilihat sebagai disiplin ilmu terkait dan landasan SI. Wilayah studi IS melibatkan studi teori dan praktik yang berkaitan dengan fenomena sosial dan teknologi, yang menentukan pengembangan, penggunaan, dan efek sistem informasi dalam organisasi dan masyarakat. Tetapi, sementara mungkin ada tumpang tindih yang cukup dari disiplin pada batas-batas, disiplin masih dibedakan oleh fokus, tujuan, dan orientasi kegiatan mereka.

Dalam ruang lingkup yang luas, istilah Sistem Informasi adalah bidang studi ilmiah yang membahas berbagai kegiatan strategis, manajerial, dan operasional yang terlibat dalam pengumpulan, pemrosesan, penyimpanan, distribusi, dan penggunaan informasi dan teknologi yang terkait dalam masyarakat dan organisasi. . Istilah sistem informasi juga digunakan untuk menggambarkan fungsi organisasi yang menerapkan pengetahuan IS di industri, lembaga pemerintah, dan organisasi nirlaba. Sistem Informasi sering merujuk pada interaksi antara proses algoritmik dan teknologi. Interaksi ini dapat terjadi di dalam atau melintasi batas-batas organisasi. Sistem informasi adalah teknologi yang digunakan organisasi dan juga cara di mana organisasi berinteraksi dengan teknologi dan cara

teknologi bekerja dengan proses bisnis organisasi. Sistem informasi berbeda dari teknologi informasi (TI) di mana sistem informasi memiliki komponen teknologi informasi yang berinteraksi dengan komponen proses.

Satu masalah dengan pendekatan itu adalah bahwa hal itu mencegah bidang IS dari tertarik pada penggunaan non-organisasi TIK, seperti dalam jejaring sosial, permainan komputer, penggunaan pribadi seluler, dll. Cara berbeda untuk membedakan bidang IS dari bidangnya. tetangga bertanya, “Aspek realitas manakah yang paling berarti di bidang IS dan bidang lainnya?” Pendekatan ini, berdasarkan pada filosofi, membantu untuk menentukan tidak hanya fokus, tujuan dan orientasi, tetapi juga martabat, nasib dan tanggung jawab bidang di antara bidang-bidang lainnya. Jurnal Internasional Manajemen Informasi, 30, 13-20.

Jenjang Karir

Sistem Informasi memiliki sejumlah bidang pekerjaan yang berbeda:

- Strategi SI
- Manajemen IS
- Pengembangan IS
- IS iterasi
- Organisasi IS

Ada berbagai jalur karier dalam disiplin sistem informasi. “Pekerja dengan pengetahuan teknis khusus dan keterampilan komunikasi yang kuat akan memiliki prospek terbaik. Pekerja dengan keterampilan manajemen dan pemahaman tentang praktik dan prinsip bisnis akan memiliki peluang bagus, karena perusahaan semakin mencari teknologi untuk mendorong pendapatan mereka. “

Teknologi informasi penting untuk operasi bisnis kontemporer, ia menawarkan banyak peluang kerja. Bidang sistem informasi mencakup orang-orang dalam organisasi yang merancang dan membangun sistem informasi, orang-orang yang menggunakan sistem itu, dan orang-orang yang bertanggung jawab untuk mengelola sistem-sistem itu. Permintaan akan staf TI tradisional seperti programmer, analis bisnis, analis sistem, dan desainer sangat penting. Banyak pekerjaan bergaji baik ada di bidang teknologi informasi. Di bagian atas daftar adalah chief information officer (CIO).

CIO adalah eksekutif yang bertanggung jawab atas fungsi IS. Di sebagian besar organisasi, CIO bekerja dengan chief executive officer (CEO), chief financial officer (CFO), dan eksekutif senior lainnya. Oleh karena itu, ia secara aktif berpartisipasi dalam proses perencanaan strategis organisasi.

Penelitian

Penelitian sistem informasi pada umumnya interdisipliner berkaitan dengan studi tentang efek sistem informasi pada perilaku individu, kelompok, dan organisasi. Hevner et al. (2004) mengategorikan penelitian dalam SI ke dalam dua paradigma ilmiah termasuk ilmu perilaku yang mengembangkan dan memverifikasi teori yang menjelaskan atau memprediksi perilaku manusia atau organisasi dan ilmu desain yang memperluas batas kemampuan manusia dan organisasi dengan menciptakan artefak baru dan inovatif.

Salvatore March dan Gerald Smith mengusulkan kerangka kerja untuk meneliti berbagai aspek Teknologi Informasi termasuk hasil penelitian (hasil penelitian) dan kegiatan untuk melakukan penelitian ini (kegiatan penelitian). Mereka mengidentifikasi hasil penelitian sebagai berikut:

1. Construct : yang merupakan konsep yang membentuk kosa kata suatu domain. Mereka merupakan konsep yang digunakan untuk menggambarkan masalah dalam domain dan untuk menentukan solusi mereka.
2. Model : yang merupakan sekumpulan proposisi atau pernyataan yang mengungkapkan hubungan antar konstruk.
3. Metode : yang merupakan serangkaian langkah (algoritma atau pedoman) yang digunakan untuk melakukan tugas. Metode didasarkan pada seperangkat konstruksi yang mendasari dan representasi (model) dari ruang solusi.
4. Instansiasi : adalah realisasi artefak di lingkungannya. Juga kegiatan penelitian termasuk:
    1. Membangun artefak untuk melakukan tugas tertentu.
    2. Mengevaluasi artefak untuk menentukan apakah ada kemajuan yang telah dicapai.
    3. Mengingat artefak yang kinerjanya telah dievaluasi, penting untuk menentukan mengapa dan bagaimana artefak bekerja atau tidak bekerja dalam lingkungannya. Oleh karena itu, berteori dan membenarkan teori tentang artefak TI.

Meskipun Sistem Informasi sebagai suatu disiplin ilmu telah berevolusi selama lebih dari 30 tahun sekarang, fokus inti atau identitas penelitian IS masih menjadi bahan perdebatan di antara para sarjana. Ada dua pandangan utama seputar debat ini: pandangan sempit yang berfokus pada artefak TI sebagai materi pokok penelitian IS, dan pandangan luas yang berfokus pada interaksi antara aspek sosial dan teknis TI yang tertanam dalam dinamika konteks yang berkembang. Pandangan ketiga menyerukan para sarjana IS untuk memberikan perhatian yang seimbang pada artefak TI dan konteksnya.

Karena studi tentang sistem informasi adalah bidang yang diterapkan, praktisi industri mengharapkan penelitian sistem informasi untuk menghasilkan temuan yang segera dapat diterapkan dalam praktik. Namun, ini tidak selalu terjadi, karena para peneliti sistem informasi sering mengeksplorasi masalah perilaku secara lebih mendalam daripada yang diharapkan oleh para praktisi. Ini mungkin membuat hasil penelitian sistem informasi sulit dipahami, dan telah menimbulkan kritik.

Dalam sepuluh tahun terakhir tren bisnis diwakili oleh meningkatnya peran Fungsi Sistem Informasi (*Information Systems Function*-ISF), terutama yang berkaitan dengan strategi perusahaan dan operasi pendukungnya. Itu menjadi faktor kunci untuk meningkatkan produktivitas dan mendukung penciptaan nilai baru.

Badan internasional para peneliti Sistem Informasi, Asosiasi Sistem Informasi (*Association for Information Systems*-AIS), dan *Senior Scholars Forum Subcommittee* untuk Jurnal (23 April 2007), mengusulkan 'keranjang' jurnal yang dianggap AIS sebagai 'sangat baik', dan dinominasikan: Management Information Systems Quarterly (MISQ), Information Systems Research (ISR), Journal of the Association for Information Systems (JAIS), Journal of Management Information Systems (JMIS), European Journal of Information Systems (EJIS), dan Information Systems Journal (ISJ).

Sejumlah konferensi sistem informasi tahunan diselenggarakan di berbagai belahan dunia, yang sebagian besar di antaranya ditinjau oleh rekan sejawat. AIS secara langsung menjalankan Konferensi Internasional tentang Sistem Informasi (*International Conference on Information Systems*-ICIS) dan Konferensi Amerika tentang Sistem Informasi ((ICIS) dan *Americas Conference on Information System*-AMCIS), sementara konferensi yang berafiliasi dengan AIS meliputi Konferensi Asia Pasifik mengenai Sistem Informasi (*Pacific Asia Conference on Information Systems*-PACIS), Konferensi Eropa tentang Sistem Informasi (*European Conference on Information Systems*-ECIS), dan Konferensi Mediterania tentang Sistem Informasi (*Mediterranean Conference on Information Systems*-

MCIS), Konferensi Internasional tentang Manajemen Sumber Daya Informasi (*International Conference on Information Resources Management*-IRM) dan Konferensi Internasional Wuhan tentang E-Bisnis (*Wuhan International Conference on E-Business*-WHICEB). Konferensi bab AIS meliputi Konferensi Australasia tentang Sistem Informasi (ACIS), Konferensi Penelitian Sistem Informasi di Skandinavia (IRIS), Konferensi Internasional Sistem Informasi (ISICO), Konferensi Bab Italia AIS (C*onference of the Italian Chapter of AIS*-itAIS), Konferensi AIS Tahunan Mid-Barat (*Annual Mid-Western AIS Conference*-MWAIS)) dan Konferensi Tahunan AIS Selatan (*Annual Conference of the Southern*-AIS SAIS).

6.4 Kecerdasan Organisasi (OI)

Organizational Intelligence (OI) adalah kemampuan organisasi untuk memahami dan menyimpulkan pengetahuan yang relevan dengan tujuan bisnisnya. Dengan kata lain, ini adalah kapasitas intelektual seluruh organisasi. Dengan kecerdasan organisasi yang relevan muncul nilai potensial yang besar bagi perusahaan dan oleh karena itu organisasi menemukan studi di mana kekuatan dan kelemahannya terletak dalam merespons perubahan dan kompleksitas. Kecerdasan Organisasi mencakup baik manajemen pengetahuan dan pembelajaran organisasi, karena merupakan penerapan konsep manajemen pengetahuan ke lingkungan bisnis, selain itu termasuk mekanisme pembelajaran, model pemahaman dan model jaringan nilai bisnis, seperti konsep balanced scorecard. Inteligensi Organisasi terdiri dari kemampuan untuk memahami situasi yang kompleks dan bertindak secara efektif, untuk menafsirkan dan bertindak berdasarkan peristiwa dan sinyal yang relevan di lingkungan. Ini juga mencakup kemampuan untuk mengembangkan, berbagi, dan menggunakan pengetahuan yang relevan dengan tujuan bisnisnya serta kemampuan untuk mencerminkan dan belajar dari pengalaman.

Sementara organisasi di masa lalu telah dipandang sebagai kompilasi tugas, produk, karyawan, pusat laba dan proses, hari ini mereka dipandang sebagai sistem cerdas yang dirancang untuk mengelola pengetahuan. Para ahli telah menunjukkan bahwa organisasi terlibat dalam proses pembelajaran menggunakan bentuk diam-diam dari pengetahuan intuitif, data keras yang disimpan dalam jaringan komputer dan informasi yang diperoleh dari lingkungan, yang semuanya digunakan untuk membuat keputusan yang masuk akal. Karena proses kompleks ini melibatkan sejumlah besar orang yang berinteraksi dengan sistem informasi yang beragam, kecerdasan organisasi lebih dari sekadar kecerdasan agregat anggota organisasi; itu adalah kecerdasan organisasi itu sendiri sebagai sistem yang lebih besar.

Kecerdasan Organisasi vs Kecerdasan Operasional

Kecerdasan Organisasi dan kecerdasan operasional biasanya dipandang sebagai himpunan bagian dari analisis bisnis, karena keduanya adalah jenis pengetahuan yang memiliki tujuan untuk meningkatkan kinerja bisnis di seluruh perusahaan. Kecerdasan Operasional sering dikaitkan dengan atau dibandingkan dengan intelijen bisnis real-time (BI) karena keduanya memberikan visibilitas dan wawasan ke dalam operasi bisnis. Kecerdasan Operasional berbeda dari BI terutama menjadi aktivitas-sentris, sedangkan BI terutama data-sentris dan bergantung pada database (atau cluster Hadoop) serta pendekatan setelah-fakta dan berbasis laporan untuk mengidentifikasi pola dalam data. Menurut definisi, Operational Intelligence bekerja secara real-time dan mengubah aliran data yang tidak terstruktur — dari file log, sensor, data jaringan dan layanan — menjadi intelijen real-time dan dapat ditindaklanjuti.

Sementara Kecerdasan Operasional berfokus pada aktivitas dan BI berfokus pada data, Kecerdasan Organisasi berbeda dari pendekatan-pendekatan lain dalam hal fokus tenaga kerja atau organisasi. Kecerdasan Organisasi membantu perusahaan memahami hubungan yang menggerakkan bisnis mereka — dengan mengidentifikasi komunitas serta alur kerja karyawan dan pola komunikasi kolaboratif di seluruh geografi, divisi, dan organisasi internal dan eksternal.

Proses Informasi

Ada banyak aspek yang harus dipertimbangkan organisasi dalam tiga langkah yang mereka ambil untuk mendapatkan informasi. Tanpa pertimbangan ini, organisasi dapat mengalami tantangan strategis.

Memperoleh Informasi

Pertama-tama, organisasi harus memperoleh informasi yang berlaku untuk membuat prediksi yang bermanfaat. Suatu organisasi harus bertanya apa yang sudah mereka ketahui dan perlu ketahui. Mereka juga harus mengetahui kerangka waktu di mana informasi dibutuhkan dan di mana serta untuk menemukannya. Untuk membuat penilaian terbaik, mereka juga harus mengevaluasi nilai informasi. Tampaknya informasi berharga yang harganya lebih mahal untuk ditemukan daripada memperolehnya dapat merugikan perusahaan. Jika dinilai bernilai, organisasi harus menemukan cara paling efisien untuk mendapatkannya.

Memproses informasi

Setelah memperoleh informasi yang benar, organisasi harus tahu cara memprosesnya dengan benar. Mereka perlu tahu bagaimana mereka dapat membuat informasi baru lebih mudah diakses dan bagaimana mereka dapat memastikan bahwa informasi tersebut disebarluaskan kepada orang yang tepat. Organisasi harus mencari cara bagaimana mengamankannya dan berapa lama dan jika lama, bagaimana mereka harus melestarikannya.

Pemanfaatan Informasi

Langkah terakhir termasuk pemanfaatan informasi. Suatu organisasi harus bertanya pada diri sendiri apakah mereka melihat informasi yang benar dan jika demikian, apakah mereka menempatkan mereka dalam konteks yang benar. Mereka harus mempertimbangkan kemungkinan perubahan lingkungan mengubah nilai informasi dan menentukan semua koneksi dan pola yang relevan. Tidak lupa untuk mengetahui apakah mereka termasuk orang yang tepat dalam proses pengambilan keputusan dan apakah ada teknologi yang dapat meningkatkan pengambilan keputusan.

Permasalan SI dalam Organisasi

Ada empat dimensi masalah yang dihadapi banyak organisasi ketika berhadapan dengan informasi. Ini juga disebut sebagai ketidaktahuan organisasi.

Ketidakpastian

Suatu organisasi mungkin tidak pasti ketika tidak memiliki cukup atau informasi yang benar. Sebagai contoh, sebuah perusahaan mungkin tidak pasti dalam lanskap kompetitif karena tidak memiliki informasi yang cukup untuk melihat bagaimana para pesaing akan bertindak. Ini tidak menyiratkan bahwa konteks situasinya kompleks atau tidak jelas. Ketidakpastian bahkan bisa ada ketika berbagai kemungkinan kecil dan sederhana. Ada berbagai tingkat ketidakpastian. Pertama-tama suatu organisasi dapat sepenuhnya ditentukan (kepastian lengkap), memiliki beberapa probabilitas (risiko), probabilitas diperkirakan dengan kepercayaan yang lebih rendah (ketidakpastian subyektif), probabilitas tidak diketahui (ketidakpastian tradisional) atau tidak pasti (ketidakpastian lengkap). Namun bahkan dengan kurangnya kejelasan, ketidakpastian mengasumsikan bahwa konteks masalahnya jelas dan dipahami dengan baik.

Kompleksitas

Suatu organisasi mungkin memproses lebih banyak informasi daripada yang dapat mereka kelola. Kompleksitas tidak selalu berkorelasi dengan ketidakjelasan atau ketidakpastian. Sebaliknya, itu terjadi ketika ada terlalu banyak atau ketika ruang lingkup terlalu besar untuk diproses. Organisasi dengan masalah kompleksitas memiliki variabel, solusi, dan metode yang saling terkait. Mengelola masalah-masalah ini tergantung pada individu dan organisasi. Misalnya, orang yang belum mendapat informasi dan novis harus berurusan dengan masing-masing elemen dan hubungan satu per satu tetapi para ahli dapat memahami situasi dengan lebih baik dan menemukan pola yang akrab dengan lebih mudah. Organisasi yang menghadapi kompleksitas harus memiliki kapasitas untuk menemukan, memetakan, mengumpulkan, berbagi, mengeksploitasi apa yang perlu diketahui organisasi.

Makna Ganda

Suatu organisasi mungkin tidak memiliki kerangka kerja konseptual untuk menafsirkan informasi. Jika ketidakpastian mewakili tidak memiliki jawaban, dan kompleksitas merepresentasikan kesulitan dalam menemukannya, ambiguitas berarti tidak mampu merumuskan pertanyaan yang tepat. Ambiguitas tidak dapat diatasi dengan meningkatkan jumlah informasi. Suatu organisasi harus dapat menafsirkan dan menjelaskan informasi dalam perjanjian bersama. Hipotesis harus terus-menerus dibuat dan dibahas dan kegiatan komunikasi utama seperti percakapan tatap muka harus dilakukan. Menyelesaikan ambiguitas pada tahap-tahap awal daripada pesaing memberi banyak keuntungan bagi organisasi karena membantu organisasi untuk membuat keputusan yang lebih tepat dan strategis dan memiliki kesadaran yang lebih baik.

Equivocality

Suatu organisasi mungkin memiliki kerangka kerja yang bersaing untuk menafsirkan suatu pekerjaan. Equivocality mengacu pada beberapa interpretasi lapangan. Setiap penafsiran tidak ambigu tetapi berbeda satu sama lain dan mungkin saling eksklusif atau bertentangan. Equivocality hasil tidak hanya karena pengalaman dan nilai-nilai setiap orang adalah unik tetapi juga dari preferensi dan tujuan yang tidak dapat diandalkan atau bertentangan, minat yang berbeda atau peran dan tanggung jawab yang tidak jelas.

Organisasi dan Budaya Informasi

Budaya organisasi menggambarkan bagaimana organisasi akan bekerja agar berhasil. Itu hanya dapat digambarkan sebagai atmosfer atau nilai-nilai organisasi. Budaya organisasi adalah penting karena dapat digunakan sebagai alat kepemimpinan yang sukses untuk membentuk dan meningkatkan organisasi. Setelah budaya diselesaikan, itu dapat digunakan oleh pemimpin untuk menyampaikan visinya kepada organisasi. Selain itu, jika pemimpin sangat memahami budaya organisasi, ia juga dapat menggunakannya untuk memprediksi hasil di masa depan dalam situasi tertentu.

Kontrol

Organisasi dengan budaya kontrol berorientasi pada perusahaan dan berorientasi pada kenyataan. Mereka akan berhasil dengan mengendalikan dan menjaga batasan. Organisasi akan menghargai ketepatan waktu informasi, keamanan dan standarisasi hirarkis. Mereka membuat rencana dan memelihara suatu proses. Organisasi ini memiliki stabilitas, kepastian dan otoritas. Misalnya, organisasi dengan budaya kontrol dapat bersifat monarki.

Kompetensi

Organisasi dengan budaya kompetensi berorientasi pada perusahaan dan berorientasi pada kemungkinan. Mereka akan berhasil dengan menjadi yang terbaik dengan eksklusivitas informasi. Organisasi menghargai efisiensi, akurasi, dan pencapaian. Mereka mencari kreativitas dan keahlian dari orang-orang di organisasi. Misalnya, organisasi dengan budaya kompetensi dapat

Penanaman

Sebuah organisasi dengan budaya budidaya berorientasi pada orang dan berorientasi pada kemungkinan. Mereka akan berhasil dengan menumbuhkan orang, yang memenuhi visi bersama. Organisasi menghargai aktualisasi diri dan kecemerlangan. Mereka juga memprioritaskan ide dari orang-orang. Misalnya, organisasi dengan budaya budi daya dapat menjadi utopianisme teknologi.

Kolaborasi

Organisasi dengan budaya kolaborasi berorientasi pada orang dan berorientasi pada kenyataan. Mereka akan berhasil dengan bekerja bersama. Organisasi menghargai afiliasi dan kerja tim. Mereka juga memprioritaskan orang dalam organisasi. Organisasi

ini memiliki aksesibilitas dan inklusifitas informasi. Misalnya, organisasi dengan budaya kolaborasi dapat menjadi anarki.

Kecerdasan Organisasi dan Inovasi

Efektivitas kepemimpinan organisasi terkait erat dengan kecerdasan dan inovasi organisasi. Ada enam faktor kepemimpinan yang menentukan atmosfer organisasi: fleksibilitas (seberapa bebas orang dapat berkomunikasi satu sama lain dan berinovasi), tanggung jawab (rasa loyalitas terhadap organisasi), standar yang ditetapkan oleh orang-orang dalam organisasi, umpan balik dan penghargaan yang sesuai , visi yang jelas dibagikan oleh orang-orang dan jumlah komitmen terhadap tujuan. Kombinasi dari faktor-faktor ini menghasilkan enam gaya kepemimpinan yang berbeda: Koersif / Memerintah, Otoritatif / Visioner, Afiliasi, Demokratis, Pelatihan dan Penentu Kecepatan.

Selanjutnya, kecerdasan organisasi adalah kumpulan kecerdasan individu. Gaya kepemimpinan organisasi dan atmosfernya terkait dengan inovasi organisasi. Inovasi terjadi ketika ada informasi baru yang dibagikan dan diproses secara efisien dalam organisasi.

6.4.6 Teori-Teori

Round Table

Dalam Round Table King Arthur, profesor Harvard David Perkins menggunakan metafora Round Table untuk membahas bagaimana percakapan kolaboratif menciptakan organisasi yang lebih cerdas. Meja Bundar adalah salah satu kisah legenda Arthurian yang paling dikenal karena dimaksudkan untuk memberi sinyal pergeseran kekuasaan dari seorang raja yang biasanya duduk di ujung meja panjang dan membuat pernyataan panjang sementara semua orang mendengarkan. Dengan mengurangi hierarki dan membuat kolaborasi lebih mudah, Arthur menemukan sumber kekuatan penting — kecerdasan organisasi — yang memungkinkannya menyatukan Inggris abad pertengahan.

Paradoks Pemotong Rumput

Paradoks mesin pemotong rumput, metafora lain dari buku Perkins, menjelaskan fakta bahwa, meskipun menyatukan upaya fisik itu mudah, menyatukan upaya mental itu sulit. “Jauh lebih mudah bagi 10 orang untuk berkolaborasi dalam memotong rumput besar daripada bagi 10 orang untuk berkolaborasi dalam mendesain mesin pemotong rumput.” Kecerdasan sebuah organisasi tercermin dari jenis percakapan — tatap muka

dan elektronik, dari ruang surat ke ruang dewan — yang dimiliki anggota satu sama lain. “Di tingkat atas, tingkat atas, kecerdasan organisasi bergantung pada cara-cara berinteraksi satu sama lain yang menunjukkan pemrosesan pengetahuan yang baik dan perilaku simbolis positif.”

Harold Wilensky berpendapat bahwa intelijen organisasi diuntungkan oleh argumen sehat dan persaingan konstruktif.

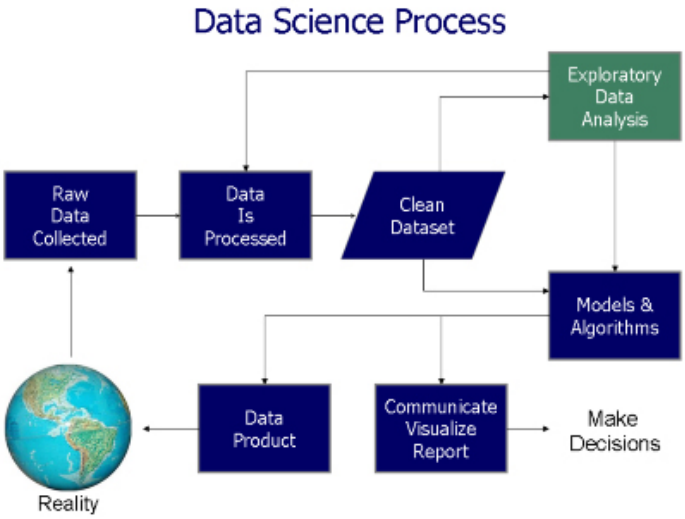
6.5 Visualisasi Data

Visualisasi data atau visualisasi data dipandang oleh banyak disiplin ilmu sebagai padanan modern komunikasi visual. Ini melibatkan penciptaan dan studi representasi visual dari data, yang berarti “informasi yang telah diabstraksikan dalam beberapa bentuk skematis, termasuk atribut atau variabel untuk unit informasi”.

Tujuan utama dari visualisasi data adalah untuk mengkomunikasikan informasi dengan jelas dan efisien melalui grafik statistik, plot dan grafik informasi. Data numerik dapat dikodekan menggunakan titik, garis, atau bilah, untuk mengkomunikasikan pesan kuantitatif secara visual. Visualisasi yang efektif membantu pengguna menganalisis dan alasan tentang data dan bukti. Itu membuat data yang kompleks lebih mudah diakses, dimengerti dan digunakan. Pengguna dapat memiliki tugas analitis tertentu, seperti membuat perbandingan atau memahami hubungan sebab akibat, dan prinsip desain grafik (yaitu, menunjukkan perbandingan atau menunjukkan hubungan sebab akibat) mengikuti tugas tersebut. Tabel umumnya digunakan di mana pengguna akan mencari pengukuran tertentu, sementara grafik dari berbagai jenis digunakan untuk menunjukkan pola atau hubungan dalam data untuk satu variabel atau lebih.

Visualisasi data adalah seni dan sains. Ini dipandang sebagai cabang statistik deskriptif oleh beberapa orang, tetapi juga sebagai alat pengembangan teori yang didasarkan oleh orang lain. Tingkat di mana data dihasilkan telah meningkat. Data yang dibuat oleh aktivitas internet dan sejumlah sensor di lingkungan, seperti satelit, disebut sebagai “Big Data”. Memproses, menganalisis, dan mengomunikasikan data ini menghadirkan berbagai tantangan etis dan analitis untuk visualisasi data. Bidang ilmu data dan praktisi yang disebut ilmuwan data telah muncul untuk membantu mengatasi tantangan ini.

Gambaran



Visualisasi data *adalah salah satu langkah dalam menganalisis data dan menyajikannya kepada* pengguna.

Visualisasi data mengacu pada teknik yang digunakan untuk mengkomunikasikan data atau informasi dengan menyandikannya sebagai objek visual (mis., Titik, garis, atau bilah) yang terkandung dalam grafik. Tujuannya adalah untuk mengkomunikasikan informasi dengan jelas dan efisien kepada pengguna. Ini adalah salah satu langkah dalam analisis data atau ilmu data. Menurut Friedman (2008) “tujuan utama visualisasi data adalah untuk mengkomunikasikan informasi dengan jelas dan efektif melalui cara grafis. Itu tidak berarti bahwa visualisasi data perlu terlihat membosankan agar fungsional atau sangat canggih untuk terlihat cantik. Untuk menyampaikan gagasan secara efektif, baik bentuk estetika maupun fungsionalitas harus berjalan seiring, memberikan wawasan tentang kumpulan data yang agak jarang dan rumit dengan mengomunikasikan aspek-aspek kuncinya dengan cara yang lebih intuitif. Namun desainer sering gagal untuk mencapai keseimbangan antara bentuk dan fungsi, menciptakan visualisasi data yang indah yang gagal untuk melayani tujuan utama mereka - untuk mengkomunikasikan informasi ”.

Memang, Fernanda Viegas dan Martin M. Wattenberg telah menyarankan bahwa visualisasi yang ideal seharusnya tidak hanya berkomunikasi dengan jelas, tetapi merangsang keterlibatan dan perhatian penonton.

Tidak terbatas pada komunikasi informasi, visualisasi data yang disusun dengan baik juga merupakan cara untuk pemahaman yang lebih baik dari data (dalam perspektif

penelitian berbasis data), karena membantu mengungkap tren, mewujudkan wawasan, mengeksplorasi sumber, dan memberi tahu cerita.

Visualisasi data berkaitan erat dengan grafik informasi, visualisasi informasi, visualisasi ilmiah, analisis data eksplorasi dan grafik statistik. Dalam milenium baru, visualisasi data telah menjadi bidang aktif penelitian, pengajaran dan pengembangan. Menurut Post dkk (2002), telah menyatukan visualisasi ilmiah dan informasi.

Karakteristik Tampilan Grafis Efektif

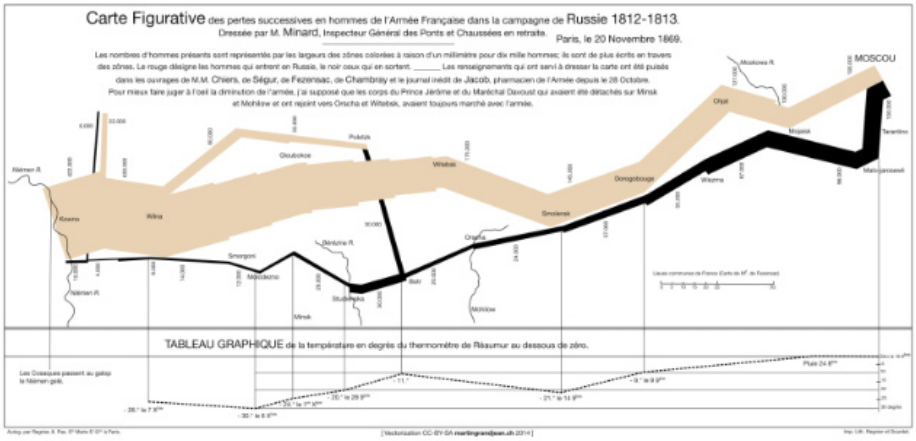


Diagram Napoleon *pada Maret 1869 karya Charles Joseph Minard - contoh awal grafik informasi.*

Profesor Edward Tufte menjelaskan bahwa pengguna tampilan informasi sedang melakukan tugas analitis tertentu seperti membuat perbandingan atau menentukan hubungan sebab akibat. Prinsip desain grafik informasi harus mendukung tugas analitis, menunjukkan perbandingan atau hubungan sebab akibat.

Dalam bukunya tahun 1983 *The Visual Display of Quantitative Information,* Edward Tufte mendefinisikan 'tampilan grafis' dan prinsip-prinsip untuk tampilan grafis yang efektif dalam bagian berikut: "Keunggulan dalam grafik statistik terdiri dari ide-ide kompleks yang dikomunikasikan dengan jelas, presisi dan efisiensi. Tampilan grafis harus:

- menunjukkan data
- mendorong pemirsa untuk berpikir tentang substansi daripada tentang metodologi, desain grafis, teknologi produksi grafis atau sesuatu yang lain
- menghindari mendistorsi apa yang dikatakan data
- menyajikan banyak angka dalam ruang kecil

- membuat set data besar yang koheren
- mendorong mata untuk membandingkan bagian data yang berbeda yang mengungkapkan data pada beberapa tingkat detail, dari tinjauan luas hingga struktur halus
- melayani tujuan yang cukup jelas: deskripsi, eksplorasi, tabulasi, atau dekorasi
- diintegrasikan erat dengan deskripsi statistik dan verbal dari kumpulan data.

Grafik mengungkapkan data. Memang grafik bisa lebih tepat dan mengungkapkan daripada perhitungan statistik konvensional. “

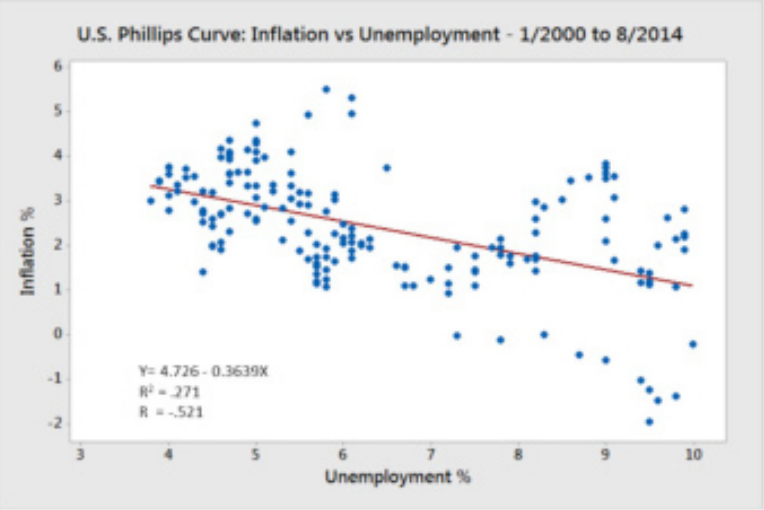
Misalnya, diagram Minard menunjukkan kerugian yang diderita tentara Napoleon pada periode 1812-1813. Enam variabel diplot: ukuran pasukan, lokasinya pada permukaan dua dimensi (x dan y), waktu, arah gerakan, dan suhu. Lebar garis menggambarkan perbandingan (ukuran pasukan pada titik waktu) sedangkan sumbu suhu menunjukkan penyebab perubahan ukuran pasukan. Tampilan multivarian pada permukaan dua dimensi ini menceritakan sebuah kisah yang dapat dipahami segera saat mengidentifikasi sumber data untuk membangun kredibilitas. Tufte menulis pada tahun 1983 bahwa: “Ini mungkin grafik statistik terbaik yang pernah dibuat.”

Tidak menerapkan prinsip-prinsip ini dapat mengakibatkan grafik yang menyesatkan, yang mendistorsi pesan atau mendukung kesimpulan yang salah. Menurut Tufte, chartjunk mengacu pada dekorasi interior luar dari grafik yang tidak meningkatkan pesan, atau efek tiga dimensi atau perspektif yang serampangan. Tidak perlu memisahkan kunci penjelas dari gambar itu sendiri, yang mengharuskan mata untuk melakukan perjalanan bolak-balik dari gambar ke kunci, adalah bentuk “puing administrasi.” Rasio “data terhadap tinta” harus dimaksimalkan, menghapus tinta non-data jika memungkinkan.

Kantor Anggaran Kongres meringkas beberapa praktik terbaik untuk tampilan grafis dalam presentasi Juni 2014. Ini termasuk: a) Mengetahui audiens Anda; b) Merancang grafik yang dapat berdiri sendiri di luar konteks laporan; dan c) Merancang gambar yang mengomunikasikan pesan utama dalam laporan.

Pesan Kuantitatif

Serangkaian waktu *diilustrasikan dengan bagan garis yang menunjukkan tren dalam* pengeluaran dan *pendapatan federal A.S. dari waktu ke waktu.*



Plot sebaran *yang menggambarkan korelasi negatif antara dua variabel (inflasi dan pengangguran) diukur pada titik waktu.*

Penulis Stephen Few menguraikan delapan jenis pesan kuantitatif yang dapat coba dipahami oleh pengguna atau dikomunikasikan dari sekumpulan data dan grafik terkait yang digunakan untuk membantu mengomunikasikan pesan:

1. Rangkaian waktu/*Time-series*: Variabel tunggal ditangkap selama periode waktu tertentu, seperti tingkat pengangguran selama periode 10 tahun. Bagan garis dapat digunakan untuk menunjukkan tren.
2. Pemeringkatan: Subdivisi kategorikal diberi peringkat dalam urutan naik atau turun, seperti peringkat kinerja penjualan (ukuran) oleh tenaga penjualan (kategori, dengan masing-masing tenaga penjualan subdivisi kategorikal) selama satu periode tunggal.

Bagan batang dapat digunakan untuk menunjukkan perbandingan di seluruh tenaga penjualan.

3. Bagian-ke-keseluruhan/ *Part-to-whole*: Subdivisi kategorikal diukur sebagai rasio terhadap keseluruhan (mis., Persentase dari 100%). Pie chart atau diagram batang dapat menunjukkan perbandingan rasio, seperti pangsa pasar yang diwakili oleh pesaing di pasar.
4. Penyimpangan: Subdivisi kategorikal dibandingkan dengan referensi, seperti perbandingan pengeluaran aktual vs anggaran untuk beberapa departemen bisnis untuk periode waktu tertentu. Bagan batang dapat menunjukkan perbandingan jumlah aktual versus jumlah referensi.
5. Distribusi frekuensi: Menunjukkan jumlah pengamatan dari variabel tertentu untuk interval yang diberikan, seperti jumlah tahun di mana pengembalian pasar saham antara interval seperti 0-10%, 11-20%, dll. Histogram, jenis diagram batang, dapat digunakan untuk analisis ini. Boxplot membantu memvisualisasikan statistik utama tentang distribusi, seperti median, kuartil, outlier, dll.
6. Korelasi: Perbandingan antara pengamatan yang diwakili oleh dua variabel (X, Y) untuk menentukan apakah mereka cenderung bergerak ke arah yang sama atau berlawanan. Misalnya, merencanakan pengangguran (X) dan inflasi (Y) untuk sampel bulan. Plot sebar biasanya digunakan untuk pesan ini.
7. Perbandingan nominal: Membandingkan subdivisi kategori tanpa urutan tertentu, seperti volume penjualan menurut kode produk. Bagan batang dapat digunakan untuk perbandingan ini.
8. Geografis atau geospasial: Perbandingan variabel di seluruh peta atau tata letak, seperti tingkat pengangguran menurut negara atau jumlah orang di berbagai lantai bangunan. Kartogram adalah grafik tipikal yang digunakan.

Analis yang meninjau sekumpulan data dapat mempertimbangkan apakah beberapa atau semua pesan dan tipe grafik di atas berlaku untuk tugas dan audiens mereka. Proses coba-coba untuk mengidentifikasi hubungan dan pesan yang bermakna dalam data adalah bagian dari analisis data eksplorasi.

Persepsi Visual dan Visualisasi Data

Manusia dapat membedakan perbedaan panjang garis, orientasi bentuk, dan warna (rona) tanpa upaya pemrosesan yang signifikan; ini disebut sebagai “atribut pra-perhatian.” Misalnya, mungkin memerlukan waktu dan upaya yang signifikan (“pemrosesan yang penuh perhatian”) untuk mengidentifikasi berapa kali angka “5” muncul dalam serangkaian

angka; tetapi jika digit itu berbeda dalam ukuran, orientasi, atau warna, contoh digit dapat dicatat dengan cepat melalui pemrosesan pra-perhatian.

Grafik yang efektif memanfaatkan pemrosesan dan atribut yang penuh perhatian dan kekuatan relatif dari atribut ini. Misalnya, karena manusia dapat lebih mudah memproses perbedaan panjang garis daripada luas permukaan, mungkin lebih efektif untuk menggunakan bagan batang (yang memanfaatkan panjang garis untuk menunjukkan perbandingan) daripada bagan pie (yang menggunakan area permukaan untuk menunjukkan perbandingan).

Persepsi Manusia / Kognisi dan Visualisasi Data

Ada sisi manusiawi untuk visualisasi data. Dengan “mempelajari persepsi dan kognisi manusia ...” kita lebih mampu memahami target data yang kita tampilkan. Kognisi mengacu pada proses dalam manusia seperti persepsi, perhatian, pembelajaran, memori, pemikiran, pembentukan konsep, membaca, dan pemecahan masalah. Dasar visualisasi data berkembang karena ketika sebuah gambar bernilai ribuan kata, data yang ditampilkan secara grafis memungkinkan untuk pemahaman informasi yang lebih mudah. Visualisasi yang tepat memberikan pendekatan berbeda untuk menunjukkan koneksi potensial, hubungan, dll. Yang tidak sejelas data kuantitatif non-visual. Visualisasi menjadi sarana eksplorasi data. Neuron otak manusia melibatkan banyak fungsi tetapi 2/3 dari neuron otak didedikasikan untuk penglihatan. Dengan indera penglihatan yang berkembang dengan baik, analisis data dapat dilakukan pada data, apakah data itu kuantitatif atau kualitatif. Visualisasi yang efektif mengikuti pemahaman proses persepsi manusia dan mampu menerapkannya pada visualisasi intuitif adalah penting. Memahami bagaimana manusia melihat dan mengatur dunia sangat penting untuk mengkomunikasikan data secara efektif kepada pembaca. Ini mengarah pada desain yang lebih intuitif.

Sejarah Visualisasi Data

Ada sejarah visualisasi data: dimulai pada abad ke-2 SM dengan pengaturan data menjadi kolom dan baris dan berkembang ke representasi kuantitatif awal pada abad ke-17. Menurut Yayasan Desain Interaksi, filsuf dan matematikawan Prancis René Descartes meletakkan dasar bagi Scotsman William Playfair. Descartes mengembangkan sistem koordinat dua dimensi untuk menampilkan nilai, yang pada akhir abad ke 18 Playfair melihat potensi komunikasi grafis data kuantitatif. Pada paruh kedua abad ke-20, Jacques Bertin menggunakan grafik kuantitatif untuk merepresentasikan informasi “secara intuitif,

jelas, akurat, dan efisien". John Tukey dan terutama Edward Tufte mendorong batas-batas visualisasi data. Tukey dengan pendekatan statistik barunya: analisis data eksplorasi dan Tufte dengan bukunya "Tampilan Visual Informasi Kuantitatif", jalan itu diaspal untuk menyempurnakan teknik visualisasi data untuk lebih dari ahli statistik. Dengan kemajuan teknologi muncul perkembangan visualisasi data; dimulai dengan visualisasi yang digambar tangan dan berkembang menjadi aplikasi yang lebih teknis - termasuk desain interaktif yang mengarah ke visualisasi perangkat lunak. Program seperti SAS, SOFA, R, Minitab, dan lainnya memungkinkan untuk visualisasi data di bidang statistik. Aplikasi visualisasi data lainnya, lebih fokus dan unik untuk individu, bahasa pemrograman seperti D3, Python dan JavaScript membantu membuat visualisasi data kuantitatif menjadi suatu kemungkinan.

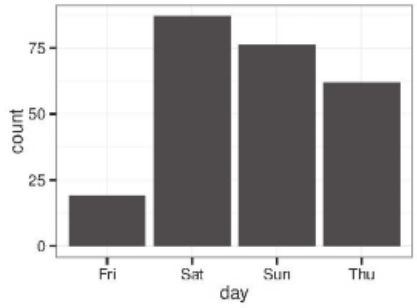
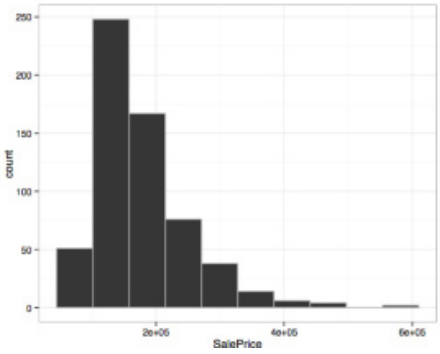
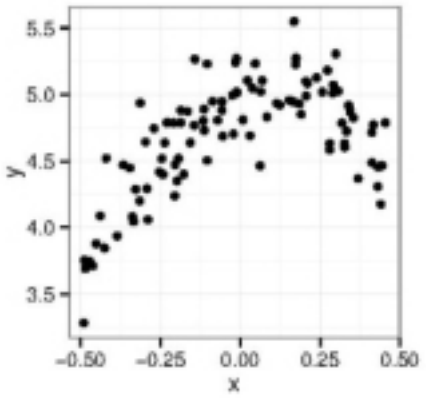
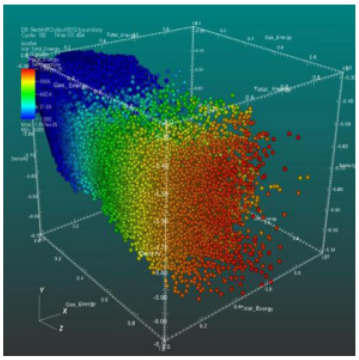
Terminologi

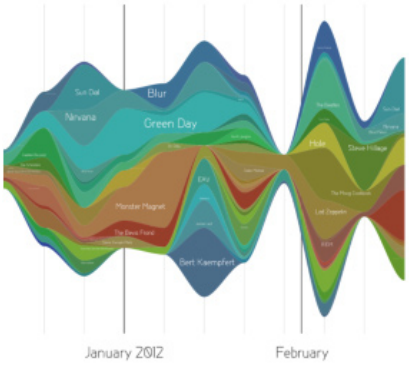
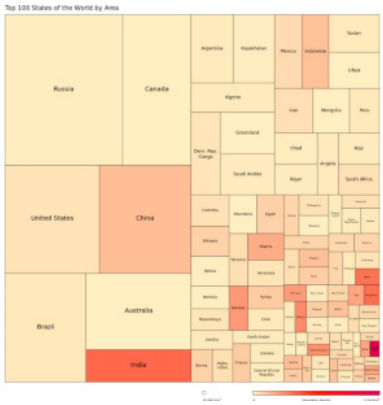
Visualisasi data melibatkan terminologi tertentu, beberapa di antaranya berasal dari statistik. Sebagai contoh, penulis Stephen Few mendefinisikan dua jenis data, yang digunakan dalam kombinasi untuk mendukung analisis atau visualisasi yang bermakna:

- Categorical: Label teks yang menggambarkan sifat data, seperti "Nama" atau "Usia". Istilah ini juga mencakup data kualitatif (non-numerik).
- Kuantitatif: Ukuran numerik, seperti "25" untuk mewakili usia dalam tahun. Dua tipe utama tampilan informasi adalah tabel dan grafik.
- Tabel berisi data kuantitatif yang disusun dalam baris dan kolom dengan label kategorikal. Ini terutama digunakan untuk mencari nilai-nilai spesifik. Dalam contoh di atas, tabel mungkin memiliki label kolom kategoris yang mewakili nama (variabel kualitatif) dan usia (variabel kuantitatif), dengan setiap baris data mewakili satu orang (unit eksperimen sampel atau subdivisi kategori).
- Grafik terutama digunakan untuk menunjukkan hubungan antara data dan menggambarkan nilai yang dikodekan sebagai objek visual (mis., Garis, bilah, atau titik). Nilai numerik ditampilkan dalam area yang digambarkan oleh satu atau lebih sumbu. Sumbu ini memberikan skala (kuantitatif dan kategorikal) yang digunakan untuk memberi label dan menetapkan nilai pada objek visual. Banyak grafik juga disebut sebagai grafik.

Perpustakaan KPI telah mengembangkan "Tabel Periodik Metode Visualisasi," grafik interaktif yang menampilkan berbagai metode visualisasi data. Ini mencakup enam jenis metode visualisasi data: data, informasi, konsep, strategi, metafora dan gabungan.

Contoh Diagram Data Visualization

	Nama	Dimensi Visual	Contoh Penggunaan
	Bar chart	• panjang / hitungan • kategori • warna)	• Perbandingan nilai, seperti kinerja penjualan untuk beberapa orang atau bisnis dalam satu periode waktu. Untuk variabel tunggal yang diukur dari waktu ke waktu (tren), diagram garis lebih disukai.
	Histogram	• bin limit • panjang / hitungan • warna)	• Menentukan frekuensi persentase pengembalian pasar saham tahunan dalam rentang tertentu (bins) seperti 0-10%, 11-20%, dll. Tinggi batang mewakili jumlah pengamatan (tahun) dengan persentase pengembalian dalam kisaran yang diwakili oleh tempat sampah.
	Scatter plot	• posisi x • posisi y • (simbol / mesin terbang) • (warna) (ukuran)	• Menentukan hubungan (misalnya, korelasi) antara pengangguran (x) dan inflasi (y) untuk beberapa periode waktu.
	Scatter plot (3D)	• posisi x • posisi y • posisi z • warna	

	Network	• ukuran node • warna node • ketebalan ikatan • dasi warna • spasialisasi	• Menemukan cluster di jaringan (mis. Mengelompokkan teman Facebook ke dalam cluster berbeda). • Menentukan node yang paling berpengaruh di jaringan (misalnya, sebuah perusahaan ingin menargetkan sekelompok kecil orang di Twitter untuk kampanye pemasaran).
	Streamgraph	• lebar • warna • waktu (aliran)	
		• ukuran • warna	• ruang disk berdasarkan lokasi / jenis file

Data Presentation Architecture

Arsitektur penyajian data (*Data Presentation Architecture*-DPA) adalah seperangkat keterampilan yang berupaya mengidentifikasi, menemukan, memanipulasi, memformat, dan menyajikan data sedemikian rupa untuk mengomunikasikan makna dan pengetahuan yang tepat secara optimal.

Secara historis, istilah arsitektur penyajian data dikaitkan dengan Kelly Lautt: “Arsitektur Presentasi Data (DPA) adalah keterampilan yang jarang diterapkan yang sangat penting bagi keberhasilan dan nilai Intelejensi Bisnis. Arsitektur penyajian data menggabungkan ilmu angka, data, dan statistik dalam menemukan informasi berharga dari data dan menjadikannya bermanfaat, relevan, dan dapat ditindaklanjuti dengan seni visualisasi data, komunikasi, psikologi organisasi, dan manajemen perubahan untuk memberikan solusi intelijen bisnis dengan data ruang lingkup, waktu pengiriman, format dan visualisasi yang akan paling efektif mendukung dan mendorong perilaku operasional, taktis dan strategis menuju tujuan bisnis (atau organisasi) yang dipahami. DPA bukanlah TI atau keahlian bisnis tetapi ada sebagai bidang keahlian yang terpisah. Sering bingung dengan visualisasi data, arsitektur penyajian data adalah keahlian yang jauh lebih luas yang mencakup menentukan data apa pada jadwal apa dan dalam format apa yang harus disajikan, bukan hanya cara terbaik untuk menyajikan data yang telah dipilih (yaitu data visualisasi). Keterampilan visualisasi data adalah salah satu elemen dari DPA. ”

Tujuan

DPA memiliki dua tujuan utama:

- Untuk menggunakan data untuk memberikan pengetahuan dengan cara seefisien mungkin (meminimalkan kebisingan, kompleksitas, dan data atau detail yang tidak perlu mengingat kebutuhan dan peran setiap audiens)
- Untuk menggunakan data untuk memberikan pengetahuan dengan cara yang seefektif mungkin (memberikan data yang relevan, tepat waktu, dan lengkap kepada setiap anggota audiens dengan cara yang jelas dan dapat dimengerti yang menyampaikan makna penting, dapat ditindaklanjuti dan dapat memengaruhi pemahaman, perilaku, dan keputusan)

Cakupan

Dengan tujuan di atas dalam pikiran, karya aktual arsitektur presentasi data terdiri dari:

- Menciptakan mekanisme pengiriman yang efektif untuk setiap anggota audiens tergantung pada peran, tugas, lokasi, dan akses mereka ke teknologi
- Menentukan arti penting (pengetahuan yang relevan) yang dibutuhkan oleh setiap anggota audiens dalam setiap konteks
- Menentukan periode yang diperlukan pembaruan data (mata uang data)
- Menentukan waktu yang tepat untuk presentasi data (kapan dan seberapa sering pengguna perlu melihat data)

- Menemukan data yang tepat (area subjek, jangkauan historis, luasnya, tingkat detail, dll.)
- Memanfaatkan analisis yang sesuai, pengelompokan, visualisasi, dan format presentasi lainnya

Bidang terkait

Pekerjaan DPA berbagi kesamaan dengan beberapa bidang lain, termasuk:

- Analisis bisnis dalam menentukan tujuan bisnis, mengumpulkan persyaratan, proses pemetaan.
- Peningkatan proses bisnis dalam tujuannya adalah untuk meningkatkan dan merampingkan tindakan dan keputusan dalam memajukan tujuan bisnis
- Visualisasi data karena menggunakan teori visualisasi yang sudah mapan untuk menambah atau menyoroti makna atau kepentingan dalam penyajian data.
- Desain grafis atau pengguna: Karena istilah DPA digunakan, istilah ini tidak memenuhi desain karena tidak mempertimbangkan detail seperti selera warna, gaya, branding, dan masalah estetika lainnya, kecuali jika elemen desain ini secara khusus diperlukan atau bermanfaat untuk komunikasi makna, dampak, tingkat keparahan atau informasi lainnya dari nilai bisnis. Sebagai contoh:
  - Memilih lokasi untuk berbagai elemen presentasi data pada halaman presentasi (seperti di portal perusahaan, dalam laporan atau pada halaman web) untuk menyampaikan hierarki, prioritas, kepentingan atau perkembangan rasional bagi pengguna adalah bagian dari keahlian DPA.
  - Memilih untuk memberikan warna tertentu dalam elemen grafis yang mewakili data dengan makna atau perhatian khusus adalah bagian dari rangkaian keterampilan DPA
- Arsitektur informasi, tetapi fokus arsitektur informasi adalah pada data yang tidak terstruktur dan karenanya mengecualikan analisis (dalam arti statistik / data) dan transformasi langsung dari konten aktual (data, untuk DPA) menjadi entitas dan kombinasi baru.
- Arsitektur solusi dalam menentukan solusi terperinci yang optimal, termasuk ruang lingkup data untuk dimasukkan, mengingat tujuan bisnis
- Analisis statistik atau analisis data karena menghasilkan informasi dan pengetahuan dari data

6.6 Profiling Data

Profiling terhadap data adalah proses memeriksa data yang tersedia di sumber data informasi yang ada (mis. Database atau file) dan mengumpulkan statistik atau ringkasan kecil tapi informatif tentang data itu. Tujuan dari statistik ini adalah untuk:

1. Cari tahu apakah data yang ada dapat dengan mudah digunakan untuk tujuan lain
2. Tingkatkan kemampuan untuk mencari data dengan memberi tag dengan kata kunci, deskripsi, atau menugaskannya ke suatu kategori
3. Berikan metrik pada kualitas data termasuk apakah data tersebut sesuai dengan standar atau pola tertentu
4. Menilai risiko yang terlibat dalam mengintegrasikan data untuk aplikasi baru, termasuk tantangan bergabung
5. Temukan metadata dari basis data sumber, termasuk pola nilai dan distribusi, kandidat kunci, kandidat kunci asing, dan dependensi fungsional
6. Menilai apakah metadata yang dikenal secara akurat menggambarkan nilai aktual dalam basis data sumber
7. Memahami tantangan data sejak awal dalam setiap proyek intensif data, sehingga kejutan proyek yang terlambat dihindari. Menemukan masalah data di akhir proyek dapat menyebabkan keterlambatan dan pembengkakan biaya.
8. Memiliki pandangan perusahaan tentang semua data, untuk penggunaan seperti manajemen data master di mana data utama diperlukan, atau tata kelola data untuk meningkatkan kualitas data.

Pendahuluan

Profiling terhadap data mengacu pada analisis sumber data kandidat untuk gudang data untuk memperjelas struktur, konten, hubungan, dan aturan derivasi data. Profiling membantu tidak hanya untuk memahami anomali dan menilai kualitas data, tetapi juga untuk menemukan, mendaftar, dan menilai metadata perusahaan. Dengan demikian, tujuan dari profil data adalah untuk memvalidasi metadata ketika itu tersedia dan untuk menemukan metadata ketika tidak. Hasil analisis digunakan baik secara strategis, untuk menentukan kesesuaian sistem sumber kandidat dan memberikan dasar untuk keputusan awal pergi / tidak-pergi, dan secara taktis, untuk mengidentifikasi masalah untuk desain solusi kemudian, dan untuk tingkat harapan sponsor.

Cara Profiling Terhadap Data

Profiling terhadap data menggunakan berbagai jenis statistik deskriptif seperti minimum, maksimum, rata-rata, mode, persentil, standar deviasi, frekuensi, dan variasi serta agregat lain seperti jumlah dan jumlah. Informasi metadata tambahan yang diperoleh selama profil data dapat berupa tipe data, panjang, nilai diskrit, keunikan, kemunculan nilai nol, pola string yang khas, dan pengenalan tipe abstrak. Metadata kemudian dapat digunakan untuk menemukan masalah seperti nilai ilegal, kesalahan ejaan, nilai yang hilang, berbagai representasi nilai, dan duplikat.

Analisis yang berbeda dilakukan untuk tingkat struktural yang berbeda. Misalnya. kolom tunggal dapat diprofilkan secara individual untuk mendapatkan pemahaman tentang distribusi frekuensi dari nilai, tipe, dan penggunaan masing-masing kolom yang berbeda. Ketergantungan nilai tertanam dapat diekspos dalam analisis lintas-kolom. Akhirnya, set nilai yang tumpang tindih yang mungkin mewakili hubungan kunci asing antara entitas dapat dieksplorasi dalam analisis antar-tabel.

Biasanya, alat yang dibuat khusus digunakan untuk profil data untuk memudahkan proses. Kompleksitas perhitungan meningkat ketika pergi dari kolom tunggal, ke tabel tunggal, ke profil struktural tabel silang. Oleh karena itu, kinerja adalah kriteria evaluasi untuk alat profil.

Kapan Melakukan Profiling Terhadap Data

Menurut Kimball, profil data dilakukan beberapa kali dan dengan intensitas yang bervariasi di seluruh proses pengembangan data warehouse. Penilaian profil cahaya harus dilakukan segera setelah sistem sumber kandidat telah diidentifikasi segera setelah akuisisi persyaratan bisnis DW / BI. Tujuannya adalah untuk mengklarifikasi pada tahap awal jika data yang tepat tersedia pada tingkat detail yang sesuai dan bahwa anomali dapat ditangani selanjutnya. Jika ini bukan masalahnya, proyek dapat dihentikan.

Pembuatan profil yang lebih rinci dilakukan sebelum proses pemodelan dimensi untuk melihat apa yang diperlukan untuk mengubah data menjadi model dimensi. Profiling terperinci meluas ke proses desain sistem ETL untuk menentukan data apa yang akan diekstraksi dan filter mana yang diterapkan.

Selain itu, data dapat dilakukan dalam proses pengembangan gudang data setelah data dimasukkan ke dalam pementasan, data mart, dll. Melakukan data pada tahap ini membantu memastikan bahwa pembersihan dan transformasi data telah dilakukan dengan benar sesuai dengan persyaratan.

Manfaat Profiling Terhadap Data

Manfaat dari profil data adalah untuk meningkatkan kualitas data, mempersingkat siklus implementasi proyek-proyek besar, dan meningkatkan pemahaman data untuk pengguna. Menemukan pengetahuan bisnis yang tertanam dalam data itu sendiri adalah salah satu manfaat signifikan yang diperoleh dari profil data. Pembuatan profil data adalah salah satu teknologi paling efektif untuk meningkatkan akurasi data dalam basis data perusahaan.

Meskipun profil data efektif dan berguna untuk setiap sektor dalam kehidupan kita sehari-hari, mungkin sulit untuk tidak masuk ke dalam “kelumpuhan analisis”.

6.7 Data Cleansing

Data Cleansing (Pembersihan data atau *data scrubbing*) adalah proses mendeteksi dan memperbaiki (atau menghapus) catatan yang rusak atau tidak akurat dari kumpulan catatan, tabel, atau basis data dan mengacu pada pengidentifikasian bagian data yang tidak lengkap, salah, tidak akurat atau tidak relevan dari data dan kemudian mengganti, memodifikasi, atau menghapus data kotor atau kasar. Pembersihan data dapat dilakukan secara interaktif dengan alat pertengkaran data, atau sebagai pemrosesan batch melalui skrip.

Setelah pembersihan, set data harus konsisten dengan set data serupa lainnya dalam sistem. Inkonsistensi yang terdeteksi atau dihapus mungkin pada awalnya disebabkan oleh kesalahan entri pengguna, oleh korupsi dalam transmisi atau penyimpanan, atau oleh definisi kamus data yang berbeda dari entitas yang sama di toko yang berbeda. Pembersihan data berbeda dari validasi data dalam validasi yang hampir selalu berarti data ditolak dari sistem pada saat masuk dan dilakukan pada saat masuk, bukan pada kumpulan data.

Proses aktual pembersihan data mungkin melibatkan penghapusan kesalahan ketik atau memvalidasi dan mengoreksi nilai terhadap daftar entitas yang diketahui. Validasi mungkin ketat (seperti menolak alamat apa pun yang tidak memiliki kode pos yang valid)

atau tidak jelas (seperti mengoreksi catatan yang sebagian cocok dengan catatan yang sudah ada dan diketahui). Beberapa solusi pembersihan data akan membersihkan data dengan memeriksa silang dengan set data yang divalidasi. Praktik pembersihan data yang umum adalah peningkatan data, di mana data dibuat lebih lengkap dengan menambahkan informasi terkait. Misalnya, menambahkan alamat dengan nomor telepon apa pun yang terkait dengan alamat itu. Pembersihan data juga dapat melibatkan kegiatan seperti, harmonisasi data, dan standarisasi data. Misalnya, harmonisasi kode pendek (st, rd, dll.) Dengan kata-kata aktual (jalan, jalan, dan sebagainya). Standarisasi data adalah cara mengubah set data referensi ke standar baru, mis, penggunaan kode standar.

Alasan Pengguaan Data Cleansing

Secara administratif, data yang salah atau tidak konsisten dapat mengarah pada kesimpulan yang salah dan investasi yang salah arah pada skala publik dan pribadi. Misalnya, pemerintah mungkin ingin menganalisis angka sensus penduduk untuk memutuskan daerah mana yang memerlukan pengeluaran lebih lanjut dan investasi untuk infrastruktur dan layanan. Dalam hal ini, penting untuk memiliki akses ke data yang dapat diandalkan untuk menghindari keputusan fiskal yang salah.

Di dunia bisnis, data yang salah bisa mahal. Banyak perusahaan menggunakan database informasi pelanggan yang merekam data seperti informasi kontak, alamat, dan preferensi. Misalnya, jika alamat tidak konsisten, perusahaan akan menderita biaya pengiriman ulang surat atau bahkan kehilangan pelanggan.

Profesi akuntansi forensik dan investigasi penipuan menggunakan pembersihan data dalam menyiapkan datanya dan biasanya dilakukan sebelum data dikirim ke gudang data untuk penyelidikan lebih lanjut.

Ada paket yang tersedia sehingga Anda dapat membersihkan / mencuci data alamat saat Anda memasukkannya ke dalam sistem Anda. Ini biasanya dilakukan melalui API dan akan meminta staf saat mereka mengetik alamat.

Kualitas Data

Data berkualitas tinggi harus melewati serangkaian kriteria kualitas. Itu termasuk:

- Validitas: Tingkat kepatuhan terhadap aturan atau batasan bisnis yang ditetapkan. Ketika teknologi database modern digunakan untuk merancang sistem pengumpulan

data, validitas cukup mudah untuk dipastikan: data yang tidak valid muncul terutama dalam konteks lama (di mana kendala tidak diterapkan dalam perangkat lunak) atau di mana teknologi pengambilan data yang tidak tepat digunakan (misalnya, spreadsheet, di mana sangat sulit untuk membatasi apa yang pengguna pilih untuk masuk ke dalam sel). Batasan data termasuk dalam kategori berikut:

- *Data-Type Constraints* - mis., Nilai dalam kolom tertentu harus dari tipe data tertentu, mis., Boolean, numerik (bilangan bulat atau nyata), tanggal, dll.
- *Range Constraints*: biasanya, angka atau tanggal harus berada dalam rentang tertentu. Artinya, mereka memiliki nilai minimum dan / atau maksimum yang diizinkan.
- *Mandatory Constraints*: Kolom tertentu tidak boleh kosong.
- Batasan Unik: Bidang, atau kombinasi bidang, harus unik di seluruh dataset. Misalnya, tidak ada dua orang yang dapat memiliki nomor jaminan sosial yang sama.
- *Set-Membership constraints*: Nilai untuk suatu kolom berasal dari sekumpulan nilai atau kode diskrit. Misalnya, jenis kelamin seseorang mungkin Perempuan, Laki-laki atau Tidak Diketahui (tidak direkam).
- *Foreign-key constraint*: Ini adalah kasus yang lebih umum dari keanggotaan yang ditetapkan. Himpunan nilai dalam kolom didefinisikan dalam kolom tabel lain yang berisi nilai unik. Misalnya, dalam basis data pembayar pajak AS, kolom "negara bagian" diharuskan untuk menjadi milik salah satu negara bagian atau wilayah AS yang ditentukan: set negara bagian / wilayah yang diizinkan dicatat dalam tabel Negara yang terpisah. Istilah kunci asing dipinjam dari terminologi basis data relasional.
- Pola ekspresi reguler: Kadang-kadang, bidang teks harus divalidasi dengan cara ini. Misalnya nomor telepon mungkin diperlukan untuk memiliki pola (999) 999-9999.
- Validasi lintas-bidang: Kondisi tertentu yang menggunakan beberapa bidang harus dimiliki. Misalnya, dalam kedokteran laboratorium, jumlah komponen sel darah putih diferensial harus sama dengan 100 (karena semuanya adalah persentase). Dalam database rumah sakit, tanggal keluarnya pasien dari rumah sakit tidak boleh lebih awal dari tanggal masuk.

- Decleansing mendeteksi kesalahan dan menghapusnya secara sintaksis untuk pemrograman yang lebih baik.
- Akurasi: Tingkat kesesuaian ukuran dengan standar atau nilai sebenarnya. Akurasi sangat sulit dicapai melalui pembersihan data dalam kasus umum, karena memerlukan akses ke sumber data eksternal yang berisi nilai sebenarnya: data "standar emas"

seperti itu seringkali tidak tersedia. Akurasi telah dicapai dalam beberapa konteks pembersihan, terutama data kontak pelanggan, dengan menggunakan database eksternal yang cocok dengan kode pos ke lokasi geografis (kota dan negara bagian), dan juga membantu memverifikasi bahwa alamat jalan dalam kode pos ini benar-benar ada.

- Kelengkapan: Sejauh mana semua tindakan yang diperlukan diketahui. Ketidaklengkapan hampir tidak mungkin untuk diperbaiki dengan metodologi pembersihan data: orang tidak dapat menyimpulkan fakta yang tidak ditangkap ketika data tersebut pada awalnya direkam. (Dalam beberapa konteks, misalnya, data wawancara, dimungkinkan untuk memperbaiki ketidaklengkapan dengan kembali ke sumber data asli, i, e., Mewawancarai kembali subjek, tetapi bahkan ini tidak menjamin kesuksesan karena masalah penarikan kembali misalnya, dalam sebuah wawancara untuk mengumpulkan data tentang konsumsi makanan, tidak ada yang mungkin mengingat dengan tepat apa yang dimakan enam bulan yang lalu. Dalam kasus sistem yang bersikeras kolom tertentu tidak boleh kosong, seseorang dapat mengatasi masalah dengan menetapkan nilai yang mengindikasikan "tidak diketahui" atau "hilang", tetapi pemberian nilai default tidak berarti bahwa data telah dibuat lengkap.
- Konsistensi: Tingkat di mana seperangkat tindakan setara di seluruh sistem. Inkonsistensi terjadi ketika dua item data dalam kumpulan data saling bertentangan: mis., Pelanggan dicatat dalam dua sistem yang berbeda karena memiliki dua alamat saat ini yang berbeda, dan hanya satu di antaranya yang dapat benar. Memperbaiki ketidakkonsistenan tidak selalu mungkin: memerlukan berbagai strategi - misalnya, memutuskan data mana yang direkam baru-baru ini, sumber data mana yang paling dapat diandalkan (pengetahuan terakhir mungkin khusus untuk organisasi tertentu), atau hanya mencoba untuk temukan kebenaran dengan menguji kedua item data (mis., menelpon pelanggan).
- Keseragaman: Tingkat pengukuran data yang ditetapkan dengan menggunakan satuan ukuran yang sama di semua sistem. Dalam kumpulan data yang dikumpulkan dari lokal yang berbeda, berat dapat dicatat dalam pound atau kilo, dan harus dikonversi ke ukuran tunggal menggunakan transformasi aritmatika.

Istilah Integritas mencakup akurasi, konsistensi, dan beberapa aspek validasi tetapi jarang digunakan dengan sendirinya dalam konteks pembersihan data karena tidak cukup spesifik. (Misalnya, "integritas referensial" adalah istilah yang digunakan untuk merujuk pada penegakan batasan kunci asing di atas.)

Proses Pembersihan Data

- Audit data: Data tersebut diaudit dengan menggunakan metode statistik dan basis data untuk mendeteksi anomali dan kontradiksi: ini pada akhirnya memberikan indikasi karakteristik anomali dan lokasi mereka. Beberapa paket perangkat lunak komersial akan memungkinkan Anda menentukan batasan dari berbagai jenis (menggunakan tata bahasa yang sesuai dengan bahasa pemrograman standar, mis., JavaScript atau Visual Basic) dan kemudian menghasilkan kode yang memeriksa data karena melanggar batasan ini. Proses ini dirujuk di bawah ini dalam "spesifikasi alur kerja" dan "eksekusi alur kerja". Untuk pengguna yang tidak memiliki akses ke perangkat lunak pembersihan kelas atas, paket basis data Microcomputer seperti Microsoft Access atau File Maker Pro juga akan membiarkan Anda melakukan pemeriksaan seperti itu, atas dasar kendala-demi-kendala, secara interaktif dengan sedikit atau tanpa pemrograman yang diperlukan dalam banyak kasus .
- Spesifikasi alur kerja: Deteksi dan penghapusan anomali dilakukan dengan urutan operasi pada data yang dikenal sebagai alur kerja. Ini ditentukan setelah proses audit data dan sangat penting dalam mencapai produk akhir dari data berkualitas tinggi. Untuk mencapai alur kerja yang tepat, penyebab anomali dan kesalahan dalam data harus dipertimbangkan secara cermat.
- Eksekusi alur kerja: Pada tahap ini, alur kerja dieksekusi setelah spesifikasinya selesai dan kebenarannya diverifikasi. Implementasi dari alur kerja harus efisien, bahkan pada set data yang besar, yang pasti menimbulkan trade-off karena pelaksanaan operasi pembersihan data bisa mahal secara komputasi.
- • Pasca pemrosesan dan pengendalian: Setelah menjalankan alur kerja pembersihan, hasilnya diperiksa untuk memverifikasi kebenaran. Data yang tidak bisa diperbaiki selama pelaksanaan alur kerja diperbaiki secara manual, jika mungkin. Hasilnya adalah siklus baru dalam proses pembersihan data di mana data diaudit lagi untuk memungkinkan spesifikasi alur kerja tambahan untuk lebih lanjut membersihkan data dengan pemrosesan otomatis.

Sumber data berkualitas baik berkaitan dengan "Budaya Kualitas Data" dan harus dimulai di bagian atas organisasi. Ini bukan hanya masalah menerapkan pemeriksaan validasi yang kuat pada layar input, karena hampir tidak peduli seberapa kuat pemeriksaan ini, mereka sering masih dapat dielakkan oleh pengguna. Ada panduan sembilan langkah untuk organisasi yang ingin meningkatkan kualitas data:

- Menyatakan komitmen tingkat tinggi terhadap budaya kualitas data
- Mendorong rekayasa ulang proses di tingkat eksekutif
- Menghabiskan uang untuk meningkatkan lingkungan entri data
- Menghabiskan uang untuk meningkatkan integrasi aplikasi
- Menghabiskan uang untuk mengubah cara kerja proses
- Mempromosikan kesadaran tim ujung ke ujung
- Mempromosikan kerja sama antar departemen
- Secara umum merayakan keunggulan kualitas data
- Mengukur dan meningkatkan kualitas data secara terus-menerus

Decleanse

Parsing: untuk mendeteksi kesalahan sintaks. Parser memutuskan apakah sebuah string data dapat diterima dalam spesifikasi data yang diizinkan. Ini mirip dengan cara kerja parser dengan tata bahasa dan bahasa.

- Transformasi data: Transformasi data memungkinkan pemetaan data dari format yang diberikan ke format yang diharapkan oleh aplikasi yang sesuai. Ini termasuk konversi nilai atau fungsi terjemahan, serta menormalkan nilai numerik agar sesuai dengan nilai minimum dan maksimum.
- Penghapusan duplikat: Deteksi duplikat membutuhkan algoritma untuk menentukan apakah data mengandung representasi duplikat dari entitas yang sama. Biasanya, data diurutkan berdasarkan kunci yang akan membawa entri duplikat lebih dekat bersama untuk identifikasi yang lebih cepat.
- Metode statistik: Dengan menganalisis data menggunakan nilai rata-rata, standar deviasi, rentang, atau algoritma pengelompokan, adalah mungkin bagi seorang ahli untuk menemukan nilai yang tidak terduga dan dengan demikian salah. Meskipun koreksi data seperti itu sulit karena nilai sebenarnya tidak diketahui, itu bisa diselesaikan dengan menetapkan nilai ke rata-rata atau nilai statistik lainnya. Metode statistik juga dapat digunakan untuk menangani nilai yang hilang yang dapat digantikan oleh satu atau lebih nilai yang masuk akal, yang biasanya diperoleh dengan algoritma augmentasi data yang luas.

Sistem Pembersihan Data

Tugas penting dari sistem ini adalah untuk menemukan keseimbangan yang cocok antara memperbaiki data kotor dan menjaga data sedekat mungkin dengan data asli dari sistem produksi sumber. Ini adalah tantangan untuk Extract, transform, load arsitek.

Sistem harus menawarkan arsitektur yang dapat membersihkan data, merekam peristiwa kualitas dan mengukur / mengontrol kualitas data di gudang data.

Awal yang baik adalah melakukan analisis profil data menyeluruh yang akan membantu menentukan kompleksitas yang diperlukan dari sistem pembersihan data dan juga memberikan gambaran tentang kualitas data saat ini dalam sistem sumber.

Kualitas Layar

Bagian dari sistem pembersihan data adalah seperangkat filter diagnostik yang dikenal sebagai layar berkualitas. Mereka masing-masing menerapkan tes dalam aliran data itu, jika gagal mencatat kesalahan dalam Skema Kejadian Kesalahan. Layar berkualitas dibagi menjadi tiga kategori:

- Layar kolom. Menguji kolom individual, mis. untuk nilai tak terduga seperti nilai NULL; nilai non-numerik yang harus numerik; di luar nilai rentang; dll.
- Struktur layar. Ini digunakan untuk menguji integritas hubungan yang berbeda antara kolom (biasanya kunci asing / primer) dalam tabel yang sama atau berbeda. Mereka juga digunakan untuk menguji bahwa sekelompok kolom valid menurut beberapa definisi struktural yang harus dipatuhi.
- Layar aturan bisnis. Yang paling kompleks dari tiga tes. Mereka menguji apakah data, mungkin di beberapa tabel, mengikuti aturan bisnis tertentu. Contohnya adalah, bahwa jika pelanggan ditandai sebagai jenis pelanggan tertentu, aturan bisnis yang mendefinisikan pelanggan jenis ini harus ditaati.

Ketika layar yang berkualitas merekam kesalahan, itu bisa menghentikan proses aliran data, mengirim data yang salah di tempat lain dari sistem target atau menandai data. Opsi terakhir dianggap solusi terbaik karena opsi pertama mengharuskan, bahwa seseorang harus secara manual menangani masalah setiap kali itu terjadi dan yang kedua menyiratkan bahwa data hilang dari sistem target (integritas) dan seringkali tidak jelas, apa yang harus terjadi pada data ini.

Evaluasi atas Alat dan Proses yang Ada

Alasan utama yang jadi acuan adalah:

- Biaya proyek: biaya biasanya dalam ratusan ribu dolar
- Waktu: kurangnya waktu yang cukup untuk berurusan dengan perangkat lunak pembersihan data skala besar

- Keamanan: kekhawatiran tentang berbagi informasi, memberikan akses aplikasi lintas sistem, dan efek pada sistem lama

Error Event Schema

Skema ini adalah tempat, di mana semua peristiwa kesalahan yang dilemparkan oleh layar berkualitas, direkam. Ini terdiri dari tabel Fakta Kejadian dengan kunci asing ke tabel tiga dimensi yang mewakili tanggal (kapan), pekerjaan batch (di mana) dan layar (yang menghasilkan kesalahan). Itu juga menyimpan informasi tentang kapan tepatnya kesalahan terjadi dan tingkat keparahan kesalahan. Selain itu ada tabel Fakta Kejadian Detail dengan kunci asing ke tabel utama yang berisi informasi terperinci tentang di mana tabel, catatan dan bidang kesalahan terjadi dan kondisi kesalahan.

Tantangan dan Masalah

- Koreksi kesalahan dan kehilangan informasi: Masalah paling sulit dalam pembersihan data adalah koreksi nilai untuk menghapus duplikat dan entri yang tidak valid. Dalam banyak kasus, informasi yang tersedia tentang anomali seperti itu terbatas dan tidak cukup untuk menentukan transformasi atau koreksi yang diperlukan, meninggalkan penghapusan entri tersebut sebagai solusi utama. Penghapusan data, menyebabkan hilangnya informasi; kerugian ini bisa sangat mahal jika ada sejumlah besar data yang dihapus.
- Pemeliharaan data yang dibersihkan: Pembersihan data adalah proses yang mahal dan memakan waktu. Jadi setelah melakukan pembersihan data dan mencapai pengumpulan data yang bebas dari kesalahan, orang ingin menghindari pembersihan ulang data secara keseluruhan setelah beberapa nilai dalam pengumpulan data berubah. Proses hanya harus diulangi pada nilai-nilai yang telah berubah; ini berarti bahwa garis keturunan pembersihan perlu dipertahankan, yang akan membutuhkan pengumpulan data dan teknik manajemen yang efisien.
- Pembersihan data dalam lingkungan yang hampir terintegrasi: Dalam sumber yang hampir terintegrasi seperti IBM DiscoveryLink, pembersihan data harus dilakukan setiap kali data diakses, yang secara signifikan meningkatkan waktu respons dan menurunkan efisiensi.
- Kerangka kerja pembersihan data: Dalam banyak kasus, tidak mungkin untuk mendapatkan grafik pembersihan data yang lengkap untuk memandu proses di muka. Hal ini membuat pembersihan data merupakan proses berulang yang melibatkan eksplorasi dan interaksi yang signifikan, yang mungkin membutuhkan kerangka kerja

dalam bentuk kumpulan metode untuk deteksi dan penghapusan kesalahan selain audit data. Ini dapat diintegrasikan dengan tahapan pemrosesan data lainnya seperti integrasi dan pemeliharaan.

6.8 Proses Mining/Penambangan

Proses mining (penambangan data) adalah teknik manajemen proses yang memungkinkan untuk analisis proses bisnis berdasarkan log peristiwa. Selama proses penambangan, algoritma penambangan data khusus diterapkan pada dataset log peristiwa untuk mengidentifikasi tren, pola, dan detail yang terkandung dalam log peristiwa yang direkam oleh sistem informasi. Proses mining bertujuan untuk meningkatkan efisiensi proses dan memahami proses. Proses penambangan data juga dikenal sebagai Penemuan Proses Bisnis Otomatis (*Automated Business Process Discovery*-ABPD).

Gambaran

Teknik penambangan data sering digunakan ketika tidak ada deskripsi formal dari proses dapat diperoleh dengan pendekatan lain, atau ketika kualitas dokumentasi yang ada dipertanyakan. Misalnya, penerapan metodologi proses penambangan ke jalur audit sistem manajemen alur kerja, log transaksi sistem perencanaan sumber daya perusahaan, atau catatan pasien elektronik di rumah sakit dapat menghasilkan model yang menggambarkan proses, organisasi, dan produk. Analisis log peristiwa juga dapat digunakan untuk membandingkan log peristiwa dengan model sebelumnya untuk memahami apakah pengamatan sesuai dengan model preskriptif atau deskriptif.

Tren manajemen kontemporer seperti BAM (*Business Activity Monitoring*), BOM (*Business Operations Management*), dan BPI (*Business Process Intelligence*) menggambarkan minat dalam mendukung fungsionalitas diagnosis dalam konteks teknologi Manajemen Proses Bisnis (misalnya, *Workflow Management Systems* dan proses lainnya - and other *process-aware information systems*).

Aplikasi

Proses penambangan mengikuti opsi yang ditetapkan dalam rekayasa proses bisnis, kemudian melampaui opsi tersebut dengan memberikan umpan balik untuk pemodelan proses bisnis:

- analisis proses menyaring, memesan, dan mengompres file log untuk wawasan lebih lanjut ke dalam hubungan operasi proses.
- desain proses dapat didukung oleh umpan balik dari pemantauan proses (rekaman aksi atau peristiwa atau logging)
- berlakunya proses menggunakan hasil dari proses penambangan berdasarkan logging untuk memicu operasi proses lebih lanjut

Klasifikasi

Ada tiga kelas teknik penambangan proses. Klasifikasi ini didasarkan pada apakah ada model sebelumnya dan, jika demikian, bagaimana model sebelumnya digunakan selama proses penambangan.

- Penemuan: Model sebelumnya (a priori) tidak ada. Berdasarkan log peristiwa, model baru dibangun atau ditemukan berdasarkan peristiwa tingkat rendah. Sebagai contoh, menggunakan algoritma alpha (pendekatan didaktis didaktik). Banyak teknik yang ada ada untuk model proses konstruksi otomatis (misalnya, Petri net, ekspresi pi-kalkulus) berdasarkan log peristiwa. Baru-baru ini, proses penambangan penelitian telah mulai menargetkan perspektif lain (mis., Data, sumber daya, waktu, dll.). Salah satu contoh adalah teknik yang dijelaskan dalam (Aalst, Reijers, & Song, 2005), yang dapat digunakan untuk membangun jejaring sosial.
- Pemeriksaan kesesuaian: Digunakan ketika ada model a priori. Model yang ada dibandingkan dengan log peristiwa proses; perbedaan antara log dan model dianalisis. Misalnya, mungkin ada model proses yang menunjukkan bahwa pesanan pembelian lebih dari 1 juta Euro memerlukan dua cek. Contoh lain adalah pengecekan prinsip “empat mata”. Pemeriksaan kesesuaian dapat digunakan untuk mendeteksi penyimpangan untuk memperkaya model. Contohnya adalah perluasan model proses dengan data kinerja, yaitu, beberapa model proses apriori digunakan untuk memproyeksikan kemacetan potensial. Contoh lain adalah penambang keputusan yang dijelaskan dalam (Rozinat & Aalst, 2006b) yang mengambil model proses a priori dan menganalisis setiap pilihan dalam model proses. Untuk setiap pilihan, log peristiwa dikonsultasikan untuk melihat informasi mana yang biasanya tersedia saat pilihan tersebut dibuat. Kemudian teknik penambangan data klasik digunakan untuk melihat elemen data mana yang memengaruhi pilihan. Akibatnya, pohon keputusan dihasilkan untuk setiap pilihan dalam proses.
- Ekstensi: Digunakan saat ada model a priori. Model ini diperluas dengan aspek atau perspektif baru, sehingga tujuannya bukan untuk memeriksa kesesuaian, tetapi untuk

meningkatkan model yang ada. Contohnya adalah perluasan model proses dengan data kinerja, mis., Beberapa model proses sebelumnya yang secara dinamis dijelaskan dengan data kinerja.

Perangkat Lunak untuk Proses Mining

Kerangka kerja perangkat lunak untuk evaluasi algoritma proses penambangan telah dikembangkan di Universitas Teknologi Eindhoven oleh Wil van der Aalst dan lainnya, dan tersedia sebagai toolkit open source.

- Process Mining
- ProM Framework
- ProM Import Framework

Fungsionalitas Proses Penambangan juga ditawarkan oleh vendor komersial berikut:

- Interstage Automated Process Discovery, layanan Proses Penambangan yang ditawarkan oleh Fujitsu, Ltd. sebagai bagian dari Interstage Integration Middleware Suite.
- Disco adalah perangkat lunak Proses Penambangan yang lengkap oleh Fluxicon.
- ARIS Process Performance Manage, Proses mining dan Alat Intelijen Proses yang ditawarkan oleh Software AG sebagai bagian dari Solusi Proses Inteligensi.
- QPR ProcessAnalyzer, perangkat lunak Proses Penambangan untuk Automated Business Process Discovery (ABPD).
- Perceptive Process Mining, solusi Proses Penambangan oleh Perangkat Lunak Perseptif (sebelumnya Futura Reflect / Pallas Athena Reflect).
- Proses Penambangan Celonis, solusi Proses Penambangan yang ditawarkan oleh Celonis
- Analisis Proses Bisnis SNP, solusi Proses Penambangan yang berfokus pada SAP oleh SNP Schneider-Neureither & Partner AG
- minit adalah perangkat lunak Proses Penambangan yang ditawarkan oleh Gradient ECM
- myInvenio cloud dan solusi di tempat oleh Cognitive Technology Ltd.
- LANA adalah alat penambangan proses yang menampilkan penemuan dan pemeriksaan kesesuaian.
- Platform ProcessGold Enterprise, sebuah integrasi Proses Penambangan & Intelejensi Bisnis.

6.9 Intelejensi Kompetitif

Intelejensi kompetitif (*Competiitive Intellegence*-CI) adalah tindakan mendefinisikan, mengumpulkan, menganalisis, dan mendistribusikan intelijen tentang produk, pelanggan, pesaing, dan segala aspek lingkungan yang diperlukan untuk mendukung para eksekutif dan manajer membuat keputusan strategis bagi suatu organisasi.

Intelijen kompetitif pada dasarnya berarti memahami dan mempelajari apa yang terjadi di dunia di luar bisnis sehingga orang dapat sekompetitif mungkin. Ini berarti mempelajari sebanyak mungkin, sesegera mungkin, tentang industri seseorang secara umum, pesaing seseorang, atau bahkan aturan zonasi khusus kabupaten tersebut. Singkatnya, ini memberdayakan antisipasi dan menghadapi tantangan secara langsung.

Poin-poin penting dari definisi ini:

1. Intelejensi kompetitif adalah praktik bisnis yang etis dan legal, berbeda dengan spionase industri, yang ilegal.
2. Fokusnya adalah pada lingkungan bisnis eksternal
3. Ada proses yang terlibat dalam mengumpulkan informasi, mengubahnya menjadi intelijen dan kemudian menggunakannya dalam pengambilan keputusan. Beberapa profesional CI keliru menekankan bahwa jika intelijen yang dikumpulkan tidak dapat digunakan atau ditindaklanjuti, itu bukan kecerdasan.

Definisi CI yang lebih terfokus menganggapnya sebagai fungsi organisasi yang bertanggung jawab untuk identifikasi awal risiko dan peluang di pasar sebelum menjadi jelas. Para ahli juga menyebut proses ini analisis sinyal awal. Definisi ini memfokuskan perhatian pada perbedaan antara penyebaran informasi faktual yang tersedia secara luas (seperti statistik pasar, laporan keuangan, kliping koran) yang dilakukan oleh fungsi-fungsi seperti perpustakaan dan pusat informasi, dan intelijen kompetitif yang merupakan perspektif tentang perkembangan dan acara yang ditujukan untuk menghasilkan keunggulan kompetitif.

Istilah CI sering dipandang sebagai sinonim dengan analisis pesaing, tetapi intelijen kompetitif lebih dari menganalisis pesaing: ini adalah tentang membuat organisasi lebih kompetitif relatif terhadap seluruh lingkungan dan pemangku kepentingan: pelanggan, pesaing, distributor, teknologi, dan data ekonomi makro.

6.9.1 Sejarah Perkembangan

Literatur yang terkait dengan bidang intelijen kompetitif paling baik dicontohkan oleh bibliografi terperinci yang diterbitkan dalam jurnal akademis wasit Society of Competitive Intelligence Professionals yang disebut Journal of Competitive Intelligence and Management. Meskipun elemen-elemen pengumpulan intelijen organisasi telah menjadi bagian dari bisnis selama bertahun-tahun, sejarah intelijen kompetitif dapat dibilang dimulai di AS pada tahun 1970-an, meskipun literatur di lapangan telah dipra-tanggalkan kali ini setidaknya oleh beberapa dekade. Pada tahun 1980, Michael Porter menerbitkan studi Competitive-Strategy: Teknik untuk Menganalisis Industri dan Pesaing yang secara luas dipandang sebagai dasar kecerdasan kompetitif modern. Sejak saat ini telah diperluas terutama oleh pasangan Craig Fleisher dan Babette Bensoussan, yang melalui beberapa buku populer tentang analisis kompetitif telah menambahkan 48 teknik analisis intelijen kompetitif yang umum diterapkan ke kotak alat praktisi. Pada tahun 1985, Leonard Fuld menerbitkan buku terlarisnya yang didedikasikan untuk intelijen pesaing. Namun, pelembagaan CI sebagai kegiatan formal di antara perusahaan-perusahaan Amerika dapat ditelusuri hingga tahun 1988, ketika Ben dan Tamar Gilad menerbitkan model organisasi pertama dari fungsi CI perusahaan formal, yang kemudian diadopsi secara luas oleh perusahaan-perusahaan AS. Program sertifikasi profesional pertama (CIP) dibuat pada tahun 1996 dengan berdirinya Akademi Kompetitif Kompetitif The Fuld-Gilad-Herring di Cambridge, Massachusetts.

Pada tahun 1986, Society of Competitive Intelligence Professionals (SCIP) didirikan di Amerika Serikat dan tumbuh pada akhir 1990-an menjadi sekitar 6.000 anggota di seluruh dunia, terutama di Amerika Serikat dan Kanada, tetapi dengan jumlah besar terutama di Inggris dan Australia. Karena kesulitan keuangan pada tahun 2009, organisasi bergabung dengan Frost & Sullivan di bawah Institut Frost & Sullivan. Sejak itu SCIP telah berganti nama menjadi “Profesional Intelijen Strategis & Kompetitif” untuk menekankan sifat strategis dari subjek, dan juga untuk memfokuskan kembali pendekatan umum organisasi, sambil tetap mempertahankan nama merek dan logo SCIP yang ada. Sejumlah upaya telah dilakukan untuk membahas kemajuan lapangan dalam pendidikan pasca-sekolah menengah (universitas), yang dicakup oleh beberapa penulis termasuk Blenkhorn & Fleisher, Fleisher, Fuld, Prescott, dan McGonagle. Meskipun pandangan umum adalah bahwa konsep intelijen kompetitif dapat dengan mudah ditemukan dan diajarkan di banyak sekolah bisnis di seluruh dunia, masih ada beberapa program akademik khusus, jurusan, atau gelar di lapangan, yang menjadi perhatian bagi akademisi di bidang yang akan ingin melihatnya diteliti lebih lanjut. Masalah-masalah ini dibahas secara luas oleh lebih dari selusin orang berpengetahuan luas dalam edisi khusus Majalah Intelijen Kompetitif yang

didedikasikan untuk topik ini. Di Prancis, seorang Ahli Khusus dalam Kecerdasan Ekonomi dan Manajemen Pengetahuan diciptakan pada tahun 1995 di dalam CERAM Business School, sekarang SKEMA Business School, di Paris, dengan tujuan memberikan pelatihan penuh dan profesional dalam Kecerdasan Ekonomi. Pusat Kecerdasan dan Pengaruh Global didirikan pada September 2011 di Sekolah yang sama.
Di sisi lain, praktisi dan perusahaan menganggap akreditasi profesional sangat penting dalam bidang ini. Pada tahun 2011, SCIP mengakui proses sertifikasi CIP dari Fuld-Gilad-Herring Academy of Competitive Intelligence sebagai program sertifikasi global dua tingkat (CIP-I dan CIP-II).

Perkembangan global juga tidak merata dalam kecerdasan bersaing. Beberapa jurnal akademis, khususnya Journal of Competitive Intelligence and Management dalam volume ketiganya, memberikan liputan tentang perkembangan global bidang ini. Misalnya, pada tahun 1997 École de guerre économique (fr) (Sekolah peperangan ekonomi) didirikan di Paris, Prancis. Ini adalah Euro pertama
lembaga kacang yang mengajarkan taktik perang ekonomi dalam dunia yang mengglobal. Di Jerman, intelijen kompetitif tidak dijaga hingga awal 1990-an. Istilah “intelijen kompetitif” pertama kali muncul dalam literatur Jerman pada tahun 1997. Pada tahun 1995 sebuah bab SCIP Jerman didirikan, yang sekarang kedua dalam hal keanggotaan di Eropa. Pada musim panas 2004, Institute for Competitive Intelligence didirikan, yang menyediakan program sertifikasi pascasarjana untuk Profesional Inteligensi Kompetitif. Jepang saat ini adalah satu-satunya negara yang secara resmi memiliki badan intelijen ekonomi (JETRO). Perusahaan ini didirikan oleh Kementerian Perdagangan dan Industri Internasional (MITI) pada tahun 1958.

Menerima pentingnya intelijen kompetitif, perusahaan multinasional besar, seperti ExxonMobil, Procter & Gamble, dan Johnson dan Johnson, telah menciptakan unit CI formal. Yang penting, organisasi melakukan kegiatan intelijen kompetitif tidak hanya sebagai perlindungan untuk melindungi terhadap ancaman dan perubahan pasar, tetapi juga sebagai metode untuk menemukan peluang dan tren baru.

Organisasi menggunakan intelijen kompetitif untuk membandingkan diri mereka dengan organisasi lain (“benchmarking kompetitif”), untuk mengidentifikasi risiko dan peluang di pasar mereka, dan untuk menguji tekanan rencana mereka terhadap respons pasar (war game), yang memungkinkan mereka membuat keputusan berdasarkan informasi. . Sebagian besar perusahaan saat ini menyadari pentingnya mengetahui apa yang dilakukan

pesaing mereka dan bagaimana industri ini berubah, dan informasi yang dikumpulkan memungkinkan organisasi untuk memahami kekuatan dan kelemahan mereka.

Salah satu kegiatan utama yang terlibat dalam intelijen persaingan perusahaan adalah penggunaan analisis rasio, menggunakan indikator kinerja utama (KPI). Organisasi membandingkan laporan tahunan pesaing mereka tentang KPI dan rasio tertentu, yang intrinsik untuk industri mereka. Ini membantu mereka melacak kinerja mereka, berhadap-hadapan dengan pesaing mereka.
Kepentingan aktual dari kategori-kategori informasi ini bagi sebuah organisasi tergantung pada con-testability dari pasarnya, budaya organisasi, kepribadian dan bias dari pembuat keputusan puncaknya, dan struktur pelaporan intelijen kompetitif di dalam perusahaan.

Strategic Intelligence (SI) berfokus pada jangka panjang, melihat masalah yang mempengaruhi daya saing perusahaan selama beberapa tahun. Cakrawala waktu aktual untuk SI pada akhirnya tergantung pada industri dan seberapa cepat itu berubah. Pertanyaan umum yang dijawab SI adalah, 'Di mana kita sebagai perusahaan dalam X tahun?' Dan 'Apa risiko dan peluang strategis yang dihadapi kita?' Jenis pekerjaan intelijen ini melibatkan antara lain identifikasi sinyal lemah dan penerapan metodologi dan proses yang disebut Strategic Early Warning (SEW), pertama kali diperkenalkan oleh Gilad, diikuti oleh Steven Shaker dan Victor Richardson, Alessandro Comai dan Joaquin Tena, dan lainnya. Menurut Gilad, 20% pekerjaan praktisi intelijen kompetitif harus didedikasikan untuk identifikasi awal strategis sinyal lemah dalam kerangka kerja SEW.

Tactical Intelligence: fokusnya adalah pada penyediaan informasi yang dirancang untuk meningkatkan keputusan jangka pendek, paling sering terkait dengan maksud meningkatkan pangsa pasar atau pendapatan. Secara umum, ini adalah jenis informasi yang Anda perlukan untuk mendukung proses penjualan dalam suatu organisasi. Ini menyelidiki berbagai aspek dari suatu produk / lini pemasaran produk:

- Produk - apa yang orang jual?
- Harga - berapa harga yang mereka tetapkan?
- Promosi - kegiatan apa yang mereka lakukan untuk mempromosikan produk ini?
- Tempat - di mana mereka menjual produk ini?
- Lainnya - struktur tenaga penjualan, desain uji klinis, masalah teknis, dll.

Dengan jumlah informasi yang tepat, organisasi dapat menghindari kejutan yang tidak menyenangkan dengan mengantisipasi gerakan pesaing dan mengurangi waktu respons.

Contoh-contoh penelitian intelijen kompetitif terbukti di surat kabar harian, seperti Wall Street Journal, Business Week, dan Fortune. Maskapai penerbangan besar mengubah ratusan tarif setiap hari sebagai tanggapan atas taktik pesaing. Mereka menggunakan informasi untuk merencanakan strategi pemasaran, harga, dan produksi mereka sendiri.

Sumber daya, seperti Internet, telah mempermudah pengumpulan informasi tentang pesaing. Dengan mengklik tombol, analis dapat menemukan tren masa depan dan persyaratan pasar. Namun kecerdasan kompetitif jauh lebih dari ini, karena tujuan utamanya adalah untuk mengarah pada keunggulan kompetitif. Karena Internet sebagian besar materi domain publik, informasi yang dikumpulkan cenderung menghasilkan wawasan yang akan unik bagi perusahaan. Bahkan ada risiko bahwa informasi yang dikumpulkan dari Internet akan menjadi informasi yang salah dan menyesatkan pengguna, sehingga peneliti intelijen kompetitif sering waspada menggunakan informasi tersebut.

Akibatnya, meskipun Internet dipandang sebagai sumber utama, sebagian besar profesional CI harus menghabiskan waktu dan intelijen pengumpulan anggaran mereka menggunakan penelitian utama - jaringan dengan para pakar industri, dari pameran dan konferensi perdagangan, dari pelanggan dan pemasok mereka sendiri, dan sebagainya. . Di mana Internet digunakan, itu adalah untuk mengumpulkan sumber-sumber untuk penelitian primer serta informasi tentang apa yang dikatakan perusahaan tentang dirinya dan keberadaan online-nya (dalam bentuk tautan ke perusahaan lain, strateginya mengenai mesin pencari dan iklan online, menyebutkan dalam forum diskusi dan di blog, dll.). Juga, yang penting adalah basis data berlangganan online dan sumber agregasi berita yang telah menyederhanakan proses pengumpulan sumber sekunder. Sumber-sumber media sosial juga menjadi penting — memberikan nama yang diwawancarai yang potensial, serta opini dan sikap, dan terkadang menyampaikan berita (mis., Melalui Twitter).

Organisasi harus berhati-hati untuk tidak menghabiskan terlalu banyak waktu dan upaya pada pesaing lama tanpa menyadari adanya pesaing baru. Mengetahui lebih banyak tentang pesaing Anda akan memungkinkan bisnis Anda tumbuh dan sukses. Praktik intelijen kompetitif berkembang setiap tahun, dan sebagian besar perusahaan dan mahasiswa bisnis sekarang menyadari pentingnya mengetahui pesaing mereka.

Menurut Arjan Singh dan Andrew Beurschgens dalam artikel mereka tahun 2006 di Competitive Intelligence Review, ada empat tahap pengembangan kemampuan intelijen kompetitif dengan perusahaan. Dimulai dengan “stick fetching”, di mana departemen CI

sangat reaktif, hingga “kelas dunia”, di mana ia sepenuhnya terintegrasi dalam proses pengambilan keputusan.

Tren Terbaru

Kemajuan teknis dalam pemrosesan paralel besar-besaran yang ditawarkan oleh arsitektur “data besar” Hadoop telah memungkinkan penciptaan beberapa platform untuk pengenalan entitas yang diberi nama seperti Proyek Apache OpenNLP dan Apache Stanbol. Yang pertama termasuk parser statistik pra-terlatih yang dapat membedakan elemen kunci untuk membangun tren dan mengevaluasi posisi kompetitif dan merespons dengan tepat. Penambangan informasi publik dari SEC.gov, Penghargaan Kontrak Federal, media sosial (Twitter, Reddit, Facebook, dan lainnya), vendor, dan situs web pesaing sekarang memungkinkan kontra-intelijen real-time sebagai strategi untuk perluasan pasar horizontal dan vertikal serta penentuan posisi produk. Ini terjadi secara otomatis di pasar besar-besaran seperti Amazon.com dan klasifikasi mereka serta prediksi asosiasi produk dan probabilitas pembelian.

Bidang Serupa

Intelejensi kompetitif telah dipengaruhi oleh intelijen strategis nasional. Meskipun intelijen nasional diteliti 50 tahun lalu, intelijen kompetitif diperkenalkan pada 1990-an. Profesional intelijen kompetitif dapat belajar dari pakar intelijen nasional, terutama dalam analisis situasi yang kompleks. Kecerdasan kompetitif dapat dikacaukan dengan (atau terlihat tumpang tindih) dengan pemindaian lingkungan, intelijen bisnis, dan riset pasar. Craig Fleisher mempertanyakan kesesuaian istilah, membandingkannya dengan intelijen bisnis, intelijen pesaing, manajemen pengetahuan, intelijen pasar, riset pemasaran, dan intelijen strategis.

Fleisher menyarankan bahwa intelijen bisnis memiliki dua bentuk. Bentuknya yang sempit (kontemporer) lebih fokus pada teknologi informasi dan fokus internal daripada CI, sementara definisi (historis) yang lebih luas lebih inklusif daripada CI. Manajemen pengetahuan (*Knowledge Management*-KM), ketika dicapai dengan tidak patut, dipandang sebagai praktik organisasi berbasis teknologi informasi yang mengandalkan penambangan data, intranet perusahaan, dan pemetaan aset organisasi untuk membuatnya dapat diakses oleh anggota organisasi untuk pengambilan keputusan. CI berbagi beberapa aspek KM; mereka berbasis kecerdasan manusia dan berbasis pengalaman untuk analisis kualitatif yang lebih canggih. KM sangat penting untuk perubahan yang efektif. Faktor

kunci yang efektif adalah sistem TI yang berdedikasi dan kuat yang menjalankan siklus kecerdasan penuh.

Market intelligence (MI) adalah intelijen bertarget industri yang dikembangkan dalam aspek waktu-nyata dari peristiwa kompetitif yang terjadi di antara empat P dari bauran pemasaran (harga, tempat, promosi dan produk) di pasar produk (atau layanan) untuk lebih baik memahami daya tarik pasar. Taktik kompetitif berbasis waktu, MI digunakan oleh manajer pemasaran dan penjualan untuk merespons konsumen lebih cepat di pasar. Fleisher menyarankan itu tidak didistribusikan seluas seperti beberapa bentuk CI, yang juga didistribusikan kepada pembuat keputusan non-pemasaran. Intelijen pasar memiliki horizon waktu yang lebih pendek daripada area intelijen lainnya, dan diukur dalam beberapa hari, minggu, atau (dalam industri yang lambat bergerak) beberapa bulan.

Riset pasar adalah bidang taktis yang digerakkan oleh metode yang terdiri dari penelitian data pelanggan netral (kepercayaan dan persepsi) netral yang dikumpulkan dalam survei atau kelompok fokus, dan dianalisis dengan teknik penelitian statistik. CI memanfaatkan beragam sumber (primer dan sekunder) dari berbagai pemangku kepentingan (pemasok, pesaing, distributor, pengganti dan media) untuk menjawab pertanyaan yang ada, mengajukan pertanyaan baru, dan memandu tindakan.

Ben Gilad dan Jan Herring meletakkan seperangkat prasyarat mendefinisikan CI, membedakannya dari disiplin ilmu yang kaya informasi lainnya seperti riset pasar atau pengembangan bisnis. Mereka menunjukkan bahwa tubuh pengetahuan umum dan seperangkat alat yang unik (topik intelijen utama, permainan perang bisnis dan analisis blindspot) membedakan CI; sementara aktivitas sensorik lainnya di perusahaan komersial fokus pada satu segmen pasar (pelanggan, pemasok atau target akuisisi), CI mensintesis data dari semua pelaku berdampak tinggi (*high-impact players*-HIP).

Gilad kemudian memfokuskan penggambaran CI-nya pada perbedaan antara informasi dan intelijen. Menurutnya, penyebut yang umum di antara fungsi-fungsi sensor organisasi (apakah mereka disebut riset pasar, intelijen bisnis atau intelijen pasar) adalah bahwa mereka memberikan informasi daripada intelijen. Kecerdasan, kata Gilad, adalah perspektif tentang fakta dan bukan fakta itu sendiri. Unik di antara fungsi-fungsi perusahaan, intelijen kompetitif memiliki perspektif risiko dan peluang untuk kinerja perusahaan; dengan demikian, itu (bukan aktivitas informasi) adalah bagian dari aktivitas manajemen risiko organisasi.

Etika

Etika telah menjadi isu diskusi yang telah lama berlangsung di kalangan praktisi CI. Pada dasarnya, pertanyaan berkisar pada apa yang bisa dan tidak diizinkan dalam hal aktivitas praktisi CI. Sejumlah perawatan ilmiah yang sangat baik telah dihasilkan pada topik ini, yang paling jelas dibahas melalui publikasi Society of Competitive Intelligence Professionals. Buku Competitive Intelligence Ethics: Navigating the Grey Zone memberikan hampir dua puluh pandangan terpisah tentang etika dalam CI, serta 10 kode lain yang digunakan oleh berbagai individu atau organisasi. Menggabungkan bahwa dengan lebih dari dua lusin artikel ilmiah atau studi yang ditemukan dalam berbagai entri bibliografi CI, jelas bahwa tidak ada kekurangan studi yang dilakukan untuk mengklasifikasikan, memahami dan menangani etika CI dengan lebih baik.

Informasi kompetitif dapat diperoleh dari sumber publik atau berlangganan, dari jaringan dengan staf pesaing atau pelanggan, pembongkaran produk pesaing atau dari wawancara penelitian lapangan. Penelitian intelijen kompetitif dapat dibedakan dari spionase industri, karena praktisi CI umumnya mematuhi pedoman hukum setempat dan norma-norma bisnis yang etis.

Outsourcing

Outsourcing telah menjadi bisnis besar bagi para profesional intelijen kompetitif. Perusahaan seperti Aperio Intelligence, Cast Intelligence, Fuld & Co, Black Cube, Securitas, dan Emissary Investigative Services menawarkan layanan intelijen perusahaan.

Bab 7 : Operational Intellegence : Komponen Tehnologi

Operational Intelligence (OI) memiliki sejumlah aspek yang telah dijelaskan dalam bab ini. Beberapa fitur ini adalah pemrosesan acara yang kompleks, manajemen proses bisnis, metadata, dan analisis akar penyebab. Komponen yang dibahas dalam teks ini sangat penting untuk memperluas pengetahuan yang ada tentang intelijen operasional.

7.1 Operational Intelligence

Operational Intelligence (OI) disebut juga kecerdasan operasional atau intelejensi oprasional, adalah kategori dinamis, analisis bisnis real-time yang memberikan visibilitas dan wawasan data, peristiwa streaming, dan operasi bisnis. Solusi Kecerdasan Operasional menjalankan kueri terhadap umpan data streaming dan data acara untuk memberikan hasil analisis waktu-nyata sebagai instruksi operasional. Kecerdasan Operasional memberi organisasi kemampuan untuk membuat keputusan dan segera bertindak berdasarkan wawasan analisis ini, melalui tindakan manual atau otomatis.

Tujuan

Tujuan OI adalah untuk memantau kegiatan bisnis dan mengidentifikasi dan mendeteksi situasi yang berkaitan dengan inefisiensi, peluang, dan ancaman serta memberikan solusi operasional. Beberapa definisi mendefinisikan intelijen operasional suatu pendekatan event-centric untuk memberikan informasi yang memberdayakan orang untuk membuat keputusan yang lebih baik. Selain itu, metrik ini bertindak sebagai titik awal untuk analisis lebih lanjut (menelusuri lebih dalam, melakukan analisis akar penyebab - mengaitkan anomali dengan transaksi spesifik dan aktivitas bisnis).

Sistem OI yang canggih juga menyediakan kemampuan untuk mengaitkan metadata dengan metrik, langkah proses, saluran, dll. Dengan ini, menjadi mudah untuk mendapatkan informasi terkait, misalnya, “mengambil informasi kontak orang yang mengelola aplikasi yang menjalankan langkah dalam transaksi bisnis yang memakan waktu 60% lebih banyak dari biasanya, ”atau“ melihat tren penerimaan / penolakan untuk pelanggan yang ditolak persetujuannya dalam transaksi ini, ”atau“ Luncurkan aplikasi yang berinteraksi dengan langkah proses ini. ”

Fitur-Fitur OI

Solusi intelijen operasional yang berbeda dapat menggunakan banyak teknologi yang berbeda dan dilaksanakan dengan cara yang berbeda. Bagian ini mencantumkan fitur-fitur umum dari solusi intelijen operasional:

- Real-time monitoring
- Real-time situation detection
- Real-time dashboards for different user roles
- Correlation of events
- Industry-specific dashboards
- Multidimensional analysis
    - Root cause analysis
    - Time Series and trend analysis
- Big Data Analytics: Kecerdasan Operasional sangat cocok untuk mengatasi tantangan yang melekat pada Big Data. Operational Intelligence secara terus-menerus memonitor dan menganalisis berbagai sumber Big Data dengan kecepatan tinggi dan volume tinggi. Sering dilakukan dalam memori, platform dan solusi OI kemudian menyajikan perhitungan dan perubahan tambahan, secara real-time, kepada pengguna akhir.

Komponen Teknologi

Solusi intelijen operasional berbagi banyak fitur, dan karenanya banyak juga berbagi komponen teknologi. Ini adalah daftar dari beberapa komponen teknologi yang umum ditemukan, dan fitur yang memungkinkan:

- Business activity monitoring(BAM) - Kustomisasi dan personalisasi Dashboard
- Complex event processing (CEP) - Analisis lanjutan yang berkelanjutan dari informasi real-time dan data historis
- Business process managemen (BPM) - Untuk melakukan eksekusi kebijakan dan proses yang digerakkan oleh model yang didefinisikan sebagai model Model Proses Bisnis dan Notasi (BPMN)
- Kerangka kerja metadata untuk memodelkan dan menghubungkan acara dengan sumber daya
- Penerbitan dan pemberitahuan multi-saluran
- Database dimensi
- Analisis Root cause
- Pengumpulan acara multi-protokol

Kecerdasan operasional adalah segmen pasar yang relatif baru (dibandingkan dengan segmen intelijen bisnis dan manajemen proses bisnis yang lebih matang). Selain perusahaan yang memproduksi produk khusus dan fokus pada area ini, ada banyak perusahaan di area yang berdekatan yang memberikan solusi dengan beberapa komponen OI.

Intelijen operasional menempatkan informasi lengkap di ujung jari seseorang, memungkinkan seseorang untuk membuat keputusan yang lebih cerdas pada waktunya untuk memaksimalkan dampak. Dengan mengkorelasikan berbagai macam peristiwa dan data dari umpan streaming dan silo data historis, intelijen operasional membantu organisasi mendapatkan visibilitas informasi secara real-time, dalam konteks, melalui dashboard canggih, wawasan real-time ke dalam kinerja bisnis, kesehatan dan status sehingga tindakan segera berdasarkan kebijakan dan proses bisnis dapat diambil. Kecerdasan operasional menerapkan manfaat analisis waktu-nyata, peringatan, dan tindakan untuk spektrum luas kasus penggunaan di dan di luar perusahaan.

Satu segmen teknologi tertentu adalah AIDC (Identifikasi Otomatis dan Pengambilan Data) yang diwakili oleh barcode, RFID, dan pengenalan suara.

Perbandingan dengan Teknologi atau Solusi Lain

Intelejensi Bisnis

OI sering dikaitkan dengan atau dibandingkan dengan intelijen bisnis (BI) atau intelijen bisnis real-time, dalam arti bahwa keduanya membantu masuk akal dari sejumlah besar informasi. Tetapi ada beberapa perbedaan mendasar: OI terutama aktivitas-sentris, sedangkan BI terutama data-sentris. Seperti halnya sebagian besar teknologi, masing-masing teknologi ini dapat dipaksa secara sub-optimal untuk melakukan tugas yang lain. OI, menurut definisi, real-time, tidak seperti BI atau BI “Sesuai Permintaan”, yang secara tradisional merupakan pendekatan berdasarkan fakta dan berdasarkan laporan untuk mengidentifikasi pola. BI real-time (mis., On-Demand BI) mengandalkan database sebagai satu-satunya sumber peristiwa.

OI menyediakan analisis berkelanjutan, real-time pada data saat istirahat dan data dalam penerbangan, sedangkan BI biasanya hanya melihat data historis saat istirahat. OI dan BI bisa saling melengkapi. OI paling baik digunakan untuk perencanaan jangka pendek, seperti memutuskan “tindakan terbaik berikutnya,” sementara BI paling baik digunakan untuk perencanaan jangka panjang (selama beberapa hari ke minggu berikutnya). BI

memerlukan pendekatan yang lebih reaktif, sering bereaksi terhadap peristiwa yang telah terjadi.

Jika semua yang diperlukan adalah sekilas kinerja historis selama periode waktu yang sangat spesifik, solusi BI yang ada harus memenuhi persyaratan. Namun, data historis perlu dianalisis dengan peristiwa yang terjadi sekarang, atau untuk mengurangi waktu antara ketika intelijen diterima dan ketika tindakan diambil, maka Intelijen Operasional adalah pendekatan yang lebih tepat.

Manajemen Sistem

Sistem Manajemen terutama mengacu pada ketersediaan dan pemantauan kemampuan infrastruktur TI. Pemantauan ketersediaan mengacu pada pemantauan status komponen infrastruktur TI seperti server, router, jaringan, dll. Ini biasanya memerlukan proses ping atau polling komponen dan menunggu untuk menerima tanggapan. Pemantauan kemampuan biasanya mengacu pada transaksi sintetis di mana aktivitas pengguna ditiru oleh program perangkat lunak khusus, dan respons yang diterima diperiksa kebenarannya.

Complex Event Processing

Ada hubungan yang kuat antara perusahaan pengolah peristiwa kompleks dan intelijen operasional, terutama karena CEP dianggap oleh banyak perusahaan OI sebagai komponen inti dari solusi OI mereka. Perusahaan CEP cenderung hanya berfokus pada pengembangan kerangka CEP untuk perusahaan lain untuk digunakan dalam organisasi mereka sebagai mesin CEP murni.

Business Activity Monitoring

Business activity monitoring (BAM) adalah perangkat lunak yang membantu dalam pemantauan proses bisnis, karena proses tersebut diimplementasikan dalam sistem komputer. BAM adalah solusi perusahaan yang terutama dimaksudkan untuk memberikan ringkasan proses bisnis secara real-time kepada manajer operasi dan manajemen tingkat atas. Perbedaan utama antara BAM dan OI tampaknya berada pada rincian implementasi - deteksi situasi real-time muncul dalam BAM dan OI dan sering diimplementasikan menggunakan CEP. Selain itu, BAM berfokus pada model proses tingkat tinggi sedangkan OI sebaliknya bergantung pada korelasi untuk menyimpulkan hubungan antara berbagai peristiwa.

Business Process Management

Rangkaian manajemen proses bisnis adalah lingkungan runtime tempat seseorang dapat melakukan eksekusi kebijakan dan proses yang ditentukan oleh model yang didefinisikan sebagai model BPMN. Sebagai bagian dari rangkaian intelijen operasional, rangkaian BPM dapat memberikan kemampuan untuk mendefinisikan dan mengelola kebijakan di seluruh perusahaan, menerapkan kebijakan tersebut pada peristiwa, dan kemudian mengambil tindakan sesuai dengan kebijakan yang telah ditentukan. BPM suite juga menyediakan kemampuan untuk mendefinisikan kebijakan seolah-olah pernyataan lalu menerapkannya pada acara.

7.2 Business Activity Monitoring

Business activity monitoring (BAM) adalah perangkat lunak yang membantu dalam pemantauan kegiatan bisnis, karena kegiatan tersebut diimplementasikan dalam sistem komputer.

Istilah ini awalnya diciptakan oleh analis di Gartner, Inc. dan mengacu pada agregasi, analisis, dan penyajian informasi waktu-nyata tentang kegiatan di dalam organisasi dan yang melibatkan pelanggan dan mitra. Aktivitas bisnis dapat berupa proses bisnis yang diatur oleh perangkat lunak manajemen proses bisnis (BPM), atau proses bisnis yang merupakan serangkaian kegiatan yang mencakup berbagai sistem dan aplikasi. BAM adalah solusi perusahaan yang terutama dimaksudkan untuk memberikan ringkasan aktivitas bisnis real-time kepada manajer operasi dan manajemen tingkat atas.

Tujuan dan Manfaat

Tujuan pemantauan kegiatan bisnis adalah untuk memberikan informasi real-time tentang status dan hasil berbagai operasi, proses, dan transaksi. Manfaat utama BAM adalah memungkinkan perusahaan untuk membuat keputusan bisnis yang lebih baik, dengan cepat menangani masalah, dan memposisikan ulang organisasi untuk mengambil keuntungan penuh dari peluang yang muncul.

Fitur Utama

Salah satu fitur yang paling terlihat dari solusi BAM adalah penyajian informasi di dashboard yang berisi indikator kinerja utama (KPI) yang digunakan untuk memberikan jaminan dan visibilitas aktivitas dan kinerja. Informasi ini digunakan oleh operasi teknis dan bisnis untuk memberikan visibilitas, pengukuran, dan jaminan kegiatan bisnis utama. Ini juga

dieksploitasi oleh korelasi peristiwa untuk mendeteksi dan memperingatkan masalah yang akan datang.

Meskipun sistem BAM biasanya menggunakan layar dasbor komputer untuk menyajikan data, BAM berbeda dari dasbor yang digunakan oleh intelijen bisnis (BI) sejauh peristiwa diproses dalam real-time atau mendekati real-time dan didorong ke dasbor dalam sistem BAM, sedangkan Dasbor BI menyegarkan pada interval yang telah ditentukan dengan polling atau query database. Bergantung pada interval penyegaran yang dipilih, BAM dan BI dashboard dapat serupa atau sangat bervariasi.

Beberapa solusi BAM juga menyediakan fungsi notifikasi masalah, yang memungkinkan mereka untuk berinteraksi secara otomatis dengan sistem pelacakan masalah. Misalnya, seluruh kelompok orang dapat dikirimi surel, pesan suara atau pesan teks, sesuai dengan sifat masalahnya. Pemecahan masalah otomatis, jika memungkinkan, dapat memperbaiki dan memulai kembali proses yang gagal.

Upaya Penempatan

Di hampir semua penyebaran BAM diperlukan penyesuaian yang luas untuk perusahaan tertentu. Banyak solusi BAM berusaha untuk mengurangi kustomisasi yang luas dan dapat menawarkan template yang ditulis untuk menyelesaikan masalah umum di sektor tertentu, misalnya perbankan, manufaktur, dan stockbroking. Karena tingginya tingkat integrasi sistem yang diperlukan untuk penyebaran awal, banyak perusahaan menggunakan pakar yang berspesialisasi dalam BAM untuk mengimplementasikan solusi.

BAM sekarang dianggap sebagai komponen penting dari solusi Kecerdasan Operasional (OI) untuk menghadirkan visibilitas ke dalam operasi bisnis. Berbagai sumber data dapat digabungkan dari silo organisasi yang berbeda untuk memberikan gambaran operasi umum yang menggunakan informasi saat ini. Di mana pun wawasan real-time memiliki nilai terbesar, solusi OI dapat diterapkan untuk memberikan informasi yang dibutuhkan. Memproses Acara

Semua acara proses solusi BAM. Sementara sebagian besar solusi BAM pertama terkait erat dengan solusi BPM dan oleh karena itu peristiwa yang diproses dikeluarkan saat proses sedang diatur, ini memiliki kelemahan mengharuskan perusahaan untuk berinvestasi di BPM sebelum dapat memperoleh dan menggunakan BAM. Solusi BAM generasi terbaru didasarkan pada teknologi pemrosesan peristiwa kompleks (CEP), dan

dapat memproses volume tinggi dari peristiwa teknis yang mendasari untuk menurunkan peristiwa bisnis tingkat yang lebih tinggi, sehingga mengurangi ketergantungan pada BPM, dan menyediakan BAM kepada khalayak yang lebih luas dari pelanggan.

Studi Kasus

Bank mungkin tertarik untuk meminimalkan jumlah uang yang dipinjamnya semalam dari bank sentral. Transfer antar bank harus dikomunikasikan dan diatur melalui otomatisasi dengan waktu yang ditentukan setiap hari kerja. Kegagalan komunikasi vital apa pun dapat membebani bank dalam jumlah besar dengan bunga yang dibebankan oleh bank sentral. Solusi BAM akan diprogram untuk mengetahui setiap pesan dan menunggu konfirmasi. Kegagalan untuk menerima konfirmasi dalam jumlah waktu yang wajar akan menjadi alasan bagi solusi BAM untuk meningkatkan alarm yang akan mengatur intervensi manual untuk menyelidiki penyebab keterlambatan dan untuk mendorong masalah ke arah penyelesaian sebelum menjadi mahal.

Contoh lain melibatkan perusahaan telekomunikasi seluler yang tertarik untuk mendeteksi situasi di mana pelanggan baru tidak diatur dengan cepat dan benar di jaringan mereka dan dalam berbagai solusi penagihan dan CRM. Peristiwa teknis tingkat rendah seperti pesan yang lewat dari satu aplikasi ke aplikasi lain melalui sistem middleware, atau transaksi yang terdeteksi dalam file log database, diproses oleh mesin CEP. Semua peristiwa yang berkaitan dengan pelanggan individu dikorelasikan untuk mendeteksi situasi yang tidak normal di mana pelanggan individu belum secara tepat atau benar ditetapkan, atau dibentuk. Peringatan dapat dibuat untuk memberi tahu operasi teknis atau untuk memberi tahu operasi bisnis, dan langkah penyediaan yang gagal dapat dimulai kembali secara otomatis.

7.3 Complex Event Processing

Complex Event Processing adalah metode pelacakan dan analisis (pemrosesan) aliran informasi (data) tentang hal-hal yang terjadi (peristiwa), dan memperoleh kesimpulan dari mereka. Pemrosesan peristiwa kompleks, atau CEP, adalah pemrosesan peristiwa yang menggabungkan data dari berbagai sumber untuk menyimpulkan peristiwa atau pola yang menunjukkan keadaan yang lebih rumit. Tujuan dari pemrosesan acara yang kompleks adalah untuk mengidentifikasi peristiwa yang bermakna (seperti peluang atau ancaman) dan menanggapinya secepat mungkin.

Peristiwa ini dapat terjadi di berbagai lapisan organisasi sebagai lead penjualan, pesanan atau panggilan layanan pelanggan. Atau, mereka dapat berupa item berita, pesan teks,

posting media sosial, umpan pasar saham, laporan lalu lintas, laporan cuaca, atau jenis data lainnya. Suatu peristiwa juga dapat didefinisikan sebagai "perubahan keadaan," ketika pengukuran melebihi ambang waktu, suhu, atau nilai lainnya yang telah ditentukan. Analis menyarankan bahwa CEP akan memberi organisasi cara baru untuk menganalisis pola secara real-time dan membantu sisi bisnis berkomunikasi lebih baik dengan departemen TI dan layanan.

Banyaknya informasi yang tersedia tentang acara kadang-kadang disebut sebagai peristiwa cloud.

Deskripsi Konseptual

Di antara ribuan peristiwa yang masuk, sistem pemantauan misalnya dapat menerima tiga kejadian berikut dari sumber yang sama:

- lonceng gereja berdering.
- penampilan seorang pria dalam tuksedo dengan seorang wanita dalam gaun putih yang mengalir.
- padi yang beterbangan di udara.

Dari peristiwa ini, sistem pemantauan dapat menyimpulkan peristiwa yang kompleks: pernikahan. CEP sebagai teknik membantu menemukan peristiwa kompleks dengan menganalisis dan menghubungkan peristiwa lain: lonceng, pria dan wanita dalam pakaian pernikahan dan nasi terbang di udara.

CEP bergantung pada sejumlah teknik, termasuk:

- Deteksi pola-peristiwa
- Mengabstraksi peristiwa
- Penyaringan peristiwa
- Agregasi dan transformasi peristiwa
- Pemodelan hierarki peristiwa
- Mendeteksi hubungan (seperti hubungan sebab akibat, keanggotaan atau waktu) antara peristiwa
- Mengabstraksi proses yang didorong oleh peristiwa

Aplikasi komersial CEP ada di berbagai industri dan termasuk perdagangan saham algoritmik, deteksi penipuan kartu kredit, pemantauan aktivitas bisnis, dan pemantauan keamanan.

Sejarah

Area CEP memiliki akar dalam simulasi kejadian diskrit, area basis data aktif dan beberapa bahasa pemrograman. Aktivitas dalam industri ini didahului oleh gelombang proyek penelitian pada 1990-an. Menurut proyek pertama yang membuka jalan ke bahasa CEP generik dan model eksekusi adalah proyek Rapide di Universitas Stanford, yang disutradarai oleh David Luckham. Secara paralel ada dua proyek penelitian lain: Infospheres di California Institute of Technology, disutradarai oleh K. Mani Chandy, dan Apama di University of Cambridge disutradarai oleh John Bates. Produk komersial adalah tanggungan dari konsep yang dikembangkan dalam ini dan beberapa proyek penelitian selanjutnya. Upaya masyarakat dimulai dalam serangkaian simposium pemrosesan acara yang diselenggarakan oleh Event Processing Technical Society, dan kemudian oleh seri konferensi ACM DEBS. Salah satu upaya masyarakat adalah memproduksi manifesto pemrosesan acara

Konsep Terkait

CEP digunakan dalam solusi Operational Intelligence (OI) untuk memberikan wawasan tentang operasi bisnis dengan menjalankan analisis kueri terhadap umpan langsung dan data peristiwa. Solusi OI mengumpulkan data real-time dan berkorelasi dengan data historis untuk memberikan wawasan dan analisis situasi saat ini. Berbagai sumber data dapat digabungkan dari silo organisasi yang berbeda untuk memberikan gambaran operasi umum yang menggunakan informasi saat ini. Di mana pun wawasan real-time memiliki nilai terbesar, solusi OI dapat diterapkan untuk memberikan informasi yang dibutuhkan.

Dalam manajemen jaringan, manajemen sistem, manajemen aplikasi dan manajemen layanan, orang biasanya merujuk pada korelasi peristiwa. Sebagai mesin CEP, mesin korelasi peristiwa (korelator peristiwa) menganalisis massa peristiwa, menunjukkan yang paling signifikan, dan memicu aksi. Namun, kebanyakan dari mereka tidak menghasilkan acara yang disimpulkan baru. Sebaliknya, mereka mengaitkan peristiwa tingkat tinggi dengan peristiwa tingkat rendah.

Mesin inferensi, mis. mesin penalaran berbasis aturan biasanya menghasilkan informasi yang disimpulkan dalam kecerdasan buatan. Namun, mereka biasanya tidak menghasilkan informasi baru dalam bentuk peristiwa yang kompleks (mis., Disimpulkan).

Studi Kasus

Contoh CEP yang lebih sistemik melibatkan mobil, beberapa sensor, dan berbagai peristiwa dan reaksi. Bayangkan sebuah mobil memiliki beberapa sensor — sensor yang mengukur tekanan ban, sensor yang mengukur kecepatan, dan sensor yang mendeteksi jika seseorang duduk di kursi atau meninggalkan kursi.

Pada situasi pertama, mobil bergerak dan tekanan dari salah satu ban bergerak dari 45 psi menjadi 41 psi selama 15 menit. Karena tekanan dalam ban berkurang, serangkaian peristiwa yang mengandung tekanan ban dihasilkan. Selain itu, serangkaian acara yang berisi kecepatan mobil dihasilkan. Event Processor mobil dapat mendeteksi situasi di mana hilangnya tekanan ban selama periode waktu yang relatif lama menghasilkan terciptanya peristiwa "lossOfTirePressure". Peristiwa baru ini dapat memicu proses reaksi untuk mencatat kehilangan tekanan ke dalam log perawatan mobil, dan memberi tahu pengemudi melalui portal mobil bahwa tekanan ban telah berkurang.

Dalam situasi kedua, mobil bergerak dan tekanan salah satu ban turun dari 45 psi menjadi 20 psi dalam 5 detik. Situasi yang berbeda terdeteksi — mungkin karena kehilangan tekanan terjadi selama periode waktu yang lebih pendek, atau mungkin karena perbedaan nilai antara setiap peristiwa lebih besar dari batas yang telah ditentukan. Situasi yang berbeda menghasilkan acara baru "blowOut- Tire" yang dihasilkan. Peristiwa baru ini memicu proses reaksi yang berbeda untuk segera memperingatkan pengemudi dan memulai rutinitas komputer di atas pesawat untuk membantu pengemudi menghentikan mobil tanpa kehilangan kendali melalui penyaradan.

Selain itu, peristiwa yang mewakili situasi yang terdeteksi juga dapat digabungkan dengan peristiwa lain untuk mendeteksi situasi yang lebih kompleks. Misalnya, dalam situasi terakhir mobil itu bergerak normal dan menderita ban pecah yang mengakibatkan mobil meninggalkan jalan dan menabrak pohon, dan pengemudi terlempar dari mobil. Serangkaian situasi berbeda terdeteksi dengan cepat. Kombinasi "blowOutTire", "zeroSpeed" dan "driverLeftSeat" dalam periode waktu yang sangat singkat menghasilkan situasi baru yang terdeteksi: "occupantThrownAccident". Meskipun tidak ada pengukuran langsung yang dapat menentukan secara meyakinkan bahwa pengemudi dilemparkan, atau bahwa

ada kecelakaan, kombinasi peristiwa memungkinkan situasi terdeteksi dan kejadian baru dibuat untuk menandakan situasi yang terdeteksi. Ini adalah inti dari acara yang kompleks (atau gabungan). Ini rumit karena seseorang tidak dapat secara langsung mendeteksi situasi; kita harus menyimpulkan atau menyimpulkan bahwa situasinya telah terjadi dari kombinasi peristiwa lain.

Jenis-Jenis CEP

Sebagian besar solusi dan konsep CEP dapat diklasifikasikan ke dalam dua kategori utama:

- CEP yang berorientasi pada agregasi
- CEP berorientasi deteksi

Solusi CEP yang berorientasi pada agregasi difokuskan pada mengeksekusi algoritma on-line sebagai respons terhadap data peristiwa yang memasuki sistem. Contoh sederhana adalah untuk terus menerus menghitung rata-rata berdasarkan data dalam acara masuk.

CEP berorientasi deteksi difokuskan pada deteksi kombinasi peristiwa yang disebut pola peristiwa atau situasi. Contoh sederhana untuk mendeteksi suatu situasi adalah dengan mencari urutan kejadian tertentu.

Ada juga pendekatan hybrid.

Proses Integrasi dengan Manajemen Proses Bisnis

Kecocokan alami untuk CEP adalah dengan Manajemen Proses Bisnis, atau BPM. BPM berfokus pada proses bisnis end-to-end, untuk terus mengoptimalkan dan menyelaraskan untuk lingkungan operasionalnya.

Namun, optimalisasi bisnis tidak hanya bergantung pada proses individu, ujung ke ujung. Tampaknya proses yang berbeda dapat saling mempengaruhi secara signifikan. Pertimbangkan skenario ini: Dalam industri dirgantara, adalah praktik yang baik untuk memantau kerusakan kendaraan untuk mencari tren (menentukan kelemahan potensial dalam proses pembuatan, bahan, dll.). Proses terpisah lainnya memonitor siklus hidup kendaraan operasional saat ini dan menonaktifkannya jika perlu. Satu kegunaan untuk

CEP adalah untuk menghubungkan proses-proses terpisah ini, sehingga dalam kasus proses awal (pemantauan kerusakan) menemukan kerusakan berdasarkan kelelahan logam (peristiwa penting), tindakan dapat dibuat untuk mengeksploitasi proses kedua (siklus hidup)) untuk mengeluarkan penarikan pada kendaraan menggunakan batch logam yang sama yang ditemukan rusak pada proses awal.

Integrasi CEP dan BPM harus ada pada dua tingkat, baik pada tingkat kesadaran bisnis (pengguna harus memahami potensi manfaat holistik dari proses individu mereka) dan juga pada tingkat teknologi (perlu ada metode dimana CEP dapat berinteraksi dengan Implementasi BPM). Untuk tinjauan terkini terkini tentang integrasi CEP dengan BPM, yang sering disebut sebagai Manajemen Proses Bisnis yang Didorong oleh Peristiwa, lihat.

Peran CEP yang berorientasi pada komputasi bisa dibilang tumpang tindih dengan teknologi Business Rule. Misalnya, pusat layanan pelanggan menggunakan CEP untuk analisis aliran klik dan manajemen pengalaman pelanggan. Perangkat lunak CEP dapat memasukkan informasi waktu-nyata tentang jutaan peristiwa (klik atau interaksi lainnya) per detik ke intelijen bisnis dan aplikasi pendukung keputusan lainnya. Agen bantuan "aplikasi rekomendasi" ini menyediakan layanan yang dipersonalisasi berdasarkan pengalaman masing-masing pelanggan. Aplikasi CEP dapat mengumpulkan data tentang apa yang pelanggan lakukan saat ini, atau bagaimana mereka baru-baru ini berinteraksi dengan perusahaan di berbagai saluran lain, termasuk dalam cabang, atau di Web melalui fitur layanan mandiri, pesan instan dan email. Aplikasi kemudian menganalisis total pengalaman pelanggan dan merekomendasikan skrip atau langkah selanjutnya yang memandu agen di telepon, dan mudah-mudahan membuat pelanggan senang.

Integrasi di Bidang Jasa Keuangan

Industri jasa keuangan adalah pengadopsi awal teknologi CEP, menggunakan pemrosesan peristiwa kompleks untuk menyusun dan mengontekstualisasikan data yang tersedia sehingga dapat menginformasikan perilaku perdagangan, khususnya perdagangan algoritmik, dengan mengidentifikasi peluang atau ancaman yang mengindikasikan pedagang (atau sistem perdagangan otomatis) harus membeli atau menjual. Misalnya, jika seorang pedagang ingin melacak saham yang memiliki lima gerakan naik diikuti dengan empat gerakan turun, teknologi CEP dapat melacak peristiwa semacam itu. Teknologi CEP juga dapat melacak kenaikan dan penurunan drastis dalam sejumlah perdagangan. Perdagangan algoritmik sudah merupakan praktik dalam perdagangan saham. Diperkirakan bahwa sekitar 60% perdagangan Ekuitas di Amerika Serikat adalah

melalui perdagangan algoritmik. CEP diharapkan terus membantu lembaga keuangan meningkatkan algoritme mereka dan menjadi lebih efisien.

Peningkatan terbaru dalam teknologi CEP membuatnya lebih terjangkau, membantu perusahaan kecil untuk membuat algoritma perdagangan sendiri dan bersaing dengan perusahaan besar. CEP telah berkembang dari teknologi yang muncul ke platform penting dari banyak pasar modal. Pertumbuhan teknologi yang paling konsisten adalah di perbankan, melayani deteksi penipuan, perbankan online, dan inisiatif pemasaran multichannel.

Saat ini, berbagai macam aplikasi keuangan menggunakan CEP, termasuk sistem laba, rugi, dan manajemen risiko, analisis urutan dan likuiditas, perdagangan kuantitatif dan sistem pembangkit sinyal, dan lainnya.

Integrasi dengan Database Deret Waktu

Database time series adalah sistem perangkat lunak yang dioptimalkan untuk penanganan data yang diatur oleh waktu. Rangkaian waktu adalah urutan item data yang terbatas atau tidak terbatas, di mana setiap item memiliki stempel waktu yang terkait dan urutan stempel waktu tidak menurun. Elemen dari deret waktu sering disebut kutu. Stempel waktu tidak harus naik (hanya tidak menurun) karena dalam praktiknya resolusi waktu dari beberapa sistem seperti sumber data keuangan bisa sangat rendah (milidetik, mikrodetik, atau bahkan nanodetik), sehingga kejadian berurutan dapat membawa stempel waktu yang sama.

Data deret waktu menyediakan konteks historis untuk analisis yang biasanya terkait dengan pemrosesan peristiwa yang kompleks. Ini dapat berlaku untuk semua industri vertikal seperti keuangan dan bekerja sama dengan teknologi lain seperti BPM.

Pertimbangkan skenario di bidang keuangan di mana ada kebutuhan untuk memahami volatilitas harga historis untuk menentukan ambang statistik dari pergerakan harga di masa depan. Ini berguna untuk model perdagangan dan analisis biaya transaksi.

Kasus ideal untuk analisis CEP adalah untuk melihat deret waktu historis dan data streaming real-time sebagai kontinum waktu tunggal. Apa yang terjadi kemarin, minggu lalu atau bulan lalu hanyalah pertentangan dari apa yang terjadi hari ini dan apa yang mungkin terjadi di masa depan. Sebuah contoh mungkin melibatkan membandingkan volume pasar

saat ini dengan volume historis, harga dan volatilitas untuk logika eksekusi perdagangan. Atau kebutuhan untuk bertindak berdasarkan harga pasar langsung mungkin melibatkan perbandingan to benchmark yang mencakup pergerakan sektor dan indeks, yang tren intra-hari dan historisnya mengukur volatilitas dan kelancaran outlier.

7.4 Manajemen Proses Bisnis

Manajemen Proses Bisnis (*Business Process Management*-BPM) adalah bidang dalam manajemen operasi yang berfokus pada peningkatan kinerja perusahaan dengan mengelola dan mengoptimalkan proses bisnis perusahaan. Karena itu dapat digambarkan sebagai "proses optimasi proses." Dikatakan bahwa BPM memungkinkan organisasi menjadi lebih efisien, lebih efektif dan lebih mampu berubah daripada pendekatan manajemen hirarkis tradisional yang terfokus secara fungsional. Proses-proses ini dapat berdampak pada biaya dan pendapatan dari suatu organisasi.

Sebagai pendekatan pembuatan kebijakan, BPM melihat proses sebagai aset penting dari suatu organisasi yang harus dipahami, dikelola, dan dikembangkan untuk mengumumkan produk dan layanan bernilai tambah kepada klien atau pelanggan. Pendekatan ini sangat mirip dengan manajemen kualitas total lainnya atau metodologi proses perbaikan berkelanjutan dan pendukung BPM juga mengklaim bahwa pendekatan ini dapat didukung, atau diaktifkan, melalui teknologi. Karena itu, banyak artikel dan sarjana BPM sering membahas BPM dari salah satu dari dua sudut pandang: orang dan / atau teknologi.
Definisi

BPMInstitute.org mendefinisikan Manajemen Proses Bisnis sebagai: definisi, peningkatan, dan manajemen proses bisnis perusahaan end-to-end dalam rangka mencapai tiga hasil yang penting bagi perusahaan berbasis pelanggan yang digerakkan oleh pelanggan: 1) kejelasan arah strategis , 2) penyelarasan sumber daya perusahaan, dan 3) peningkatan disiplin dalam operasi sehari-hari.

Koalisi Manajemen Alur Kerja, BPM.com dan beberapa sumber lain telah mencapai kesepakatan tentang definisi berikut:

> Business Process Management (BPM) adalah disiplin yang melibatkan setiap kombinasi pemodelan, otomatisasi, eksekusi, kontrol, pengukuran dan optimalisasi aliran aktivitas bisnis, dalam mendukung tujuan perusahaan, sistem spanning, karyawan, pelanggan dan mitra di dalam dan di luar batas perusahaan.

Asosiasi Profesional Manajemen Proses Bisnis mendefinisikan BPM sebagai:

Business Process Management (BPM) adalah pendekatan disiplin untuk mengidentifikasi, merancang, melaksanakan, mendokumentasikan, mengukur, memantau, dan mengendalikan proses bisnis otomatis dan non-otomatis untuk mencapai hasil yang konsisten dan bertarget yang selaras dengan tujuan strategis organisasi. BPM melibatkan definisi yang disengaja, kolaboratif dan semakin berbantuan teknologi, peningkatan, inovasi, dan manajemen proses bisnis ujung-ke-ujung yang mendorong hasil bisnis, menciptakan nilai, dan memungkinkan organisasi untuk memenuhi tujuan bisnisnya dengan lebih gesit. BPM memungkinkan perusahaan untuk menyelaraskan proses bisnisnya dengan strategi bisnisnya, yang mengarah pada kinerja perusahaan yang efektif secara keseluruhan melalui peningkatan aktivitas kerja tertentu baik dalam departemen tertentu, lintas perusahaan, atau antar organisasi.

Gartner mendefinisikan Business process management (BPM) sebagai:

"Disiplin proses pengelolaan (bukan tugas) sebagai sarana untuk meningkatkan hasil kinerja bisnis dan kelincahan operasional. Proses mencakup batasan organisasi, menghubungkan orang, arus informasi, sistem dan aset lainnya untuk menciptakan dan memberikan nilai kepada pelanggan dan konstituen. "

Adalah umum untuk membingungkan BPM dengan BPM Suite (BPMS). BPM adalah disiplin profesional yang dilakukan oleh orang-orang, sedangkan BPMS adalah seperangkat alat teknologi yang dirancang untuk membantu para profesional BPM mencapai tujuan mereka. BPM juga tidak perlu bingung dengan aplikasi atau solusi yang dikembangkan untuk mendukung proses tertentu. Suites dan solusi mewakili cara mengotomatisasi proses bisnis, tetapi otomatisasi hanya satu aspek dari BPM.

Perubahan dalam Manajemen Proses Bisnis

Konsep proses bisnis mungkin sama tradisionalnya dengan konsep tugas, departemen, produksi, dan output, yang timbul dari masalah penjadwalan job shop di awal abad ke-20. Pendekatan manajemen dan peningkatan pada 2010, dengan definisi formal dan pemodelan teknis, telah ada sejak awal 1990-an. Perhatikan bahwa istilah "proses bisnis" kadang-kadang digunakan oleh praktisi TI sebagai sinonim dengan pengelolaan proses middleware atau dengan mengintegrasikan tugas-tugas perangkat lunak aplikasi.

Meskipun BPM awalnya berfokus pada otomatisasi proses bisnis dengan menggunakan teknologi informasi, sejak itu telah diperluas untuk mengintegrasikan proses yang digerakkan manusia di mana interaksi manusia berlangsung secara seri atau paralel

dengan penggunaan teknologi. Misalnya, sistem manajemen alur kerja dapat menetapkan langkah-langkah individual yang membutuhkan penerapan intuisi atau penilaian manusia untuk manusia yang relevan dan tugas-tugas lain dalam alur kerja ke sistem otomatis yang relevan.

Variasi yang lebih baru seperti “manajemen interaksi manusia” prihatin dengan interaksi antara pekerja manusia yang melakukan tugas.

Pada 2010 teknologi telah memungkinkan kopling BPM dengan metodologi lain, seperti Six Sigma. Beberapa alat BPM seperti SIPOCs, aliran proses, RACIs, CTQs dan histogram memungkinkan pengguna untuk:

- memvisualisasikan - fungsi dan proses
- mengukur - menentukan ukuran yang tepat untuk menentukan kesuksesan
- menganalisis - membandingkan berbagai simulasi untuk menentukan peningkatan yang optimal
- meningkatkan - memilih dan terapkan peningkatan itu
- mengontrol - menggunakan implementasi ini dan dengan menggunakan dasbor yang ditentukan pengguna memantau peningkatan secara real time dan memasukkan informasi kinerja kembali ke model simulasi dalam persiapan untuk iterasi perbaikan berikutnya
- merekayasa ulang - mengubah proses dari awal untuk hasil yang lebih baik

Ini membawa manfaat untuk dapat mensimulasikan perubahan proses bisnis berdasarkan data dunia nyata (tidak hanya pada asumsi pengetahuan). Selain itu, penggabungan BPM ke metodologi industri memungkinkan pengguna untuk terus merampingkan dan mengoptimalkan proses untuk memastikan bahwa itu disesuaikan dengan kebutuhan pasarnya.

Pada 2012, penelitian tentang BPM telah memberikan perhatian yang semakin besar terhadap kepatuhan proses bisnis. Meskipun aspek kunci dari proses bisnis adalah fleksibilitas, karena proses bisnis terus-menerus perlu beradaptasi dengan perubahan dalam lingkungan, kepatuhan terhadap strategi bisnis, kebijakan dan peraturan pemerintah juga harus dipastikan. Aspek kepatuhan dalam BPM sangat penting bagi organisasi pemerintah. Pada 2010 pendekatan BPM dalam konteks pemerintah sebagian besar fokus pada proses operasional dan representasi pengetahuan. Meskipun ada banyak studi teknis tentang proses bisnis operasional di sektor publik dan swasta, peneliti jarang

memperhitungkan kegiatan kepatuhan hukum, misalnya proses implementasi hukum dalam badan administrasi publik.

Sikuls BPM

Kegiatan manajemen proses bisnis dapat secara sewenang-wenang dikelompokkan ke dalam kategori seperti desain, pemodelan, pelaksanaan, pemantauan, dan optimisasi.

Rancangan

Desain proses mencakup identifikasi proses yang ada dan desain proses "yang akan dilakukan". Area fokus termasuk representasi dari aliran proses, faktor-faktor di dalamnya, peringatan dan pemberitahuan, peningkatan, prosedur operasi standar, perjanjian tingkat layanan, dan mekanisme penyerahan tugas. Terlepas dari apakah proses yang ada dipertimbangkan atau tidak, tujuan langkah ini adalah untuk memastikan bahwa rancangan teoretis yang benar dan efisien disiapkan.

Perbaikan yang diusulkan bisa dalam alur kerja manusia ke manusia, manusia ke sistem atau sistem ke sistem, dan mungkin menargetkan regulasi, pasar, atau tantangan kompetitif yang dihadapi oleh bisnis. Proses yang ada dan desain proses baru untuk berbagai aplikasi harus disinkronkan dan tidak menyebabkan pemadaman besar atau gangguan proses.

Pemodelan

Pemodelan menggunakan desain teoretis dan memperkenalkan kombinasi variabel (mis., Perubahan biaya sewa atau bahan, yang menentukan bagaimana proses dapat beroperasi dalam keadaan yang berbeda).

Mungkin juga melibatkan menjalankan "analisis bagaimana-jika" (Ketentuan-kapan, jika, lain) pada proses: "Bagaimana jika saya memiliki 75% sumber daya untuk melakukan tugas yang sama?" "Bagaimana jika saya ingin melakukan pekerjaan yang sama untuk 80% dari biaya saat ini?".

Eksekusi

Salah satu cara untuk mengotomatiskan proses adalah dengan mengembangkan atau membeli aplikasi yang mengeksekusi langkah-langkah proses yang diperlukan; namun, dalam praktiknya, aplikasi ini jarang menjalankan semua langkah proses secara akurat atau sepenuhnya. Pendekatan lain adalah dengan menggunakan kombinasi perangkat

lunak dan intervensi manusia; Namun pendekatan ini lebih kompleks, membuat proses dokumentasi sulit.

Sebagai tanggapan terhadap masalah-masalah ini, perangkat lunak telah dikembangkan yang memungkinkan proses bisnis penuh (seperti yang dikembangkan dalam aktivitas desain proses) dapat didefinisikan dalam bahasa komputer yang dapat langsung dieksekusi oleh komputer. Model proses dapat dijalankan melalui mesin eksekusi yang mengotomatisasi proses langsung dari model (misalnya menghitung rencana pembayaran pinjaman) atau, ketika langkah terlalu rumit untuk diotomatisasi, Notasi Pemodelan Proses Bisnis (BPMN) menyediakan kemampuan akhir untuk input manusia. Dibandingkan dengan salah satu dari pendekatan sebelumnya, secara langsung mengeksekusi definisi proses dapat lebih mudah dan karena itu lebih mudah untuk ditingkatkan. Namun, mengotomatisasi definisi proses membutuhkan infrastruktur yang fleksibel dan komprehensif, yang biasanya mengesampingkan penerapan sistem ini dalam lingkungan TI yang legal.

Aturan bisnis telah digunakan oleh sistem untuk memberikan definisi untuk perilaku yang mengatur, dan mesin aturan bisnis dapat digunakan untuk mendorong pelaksanaan dan resolusi proses.

Pemantauan

Pemantauan mencakup pelacakan proses individu, sehingga informasi tentang negara mereka dapat dengan mudah dilihat, dan statistik tentang kinerja satu atau lebih proses dapat disediakan. Contoh pelacakan ini adalah dapat menentukan keadaan pesanan pelanggan (mis. Pesanan tiba, menunggu pengiriman, faktur dibayar) sehingga masalah dalam operasinya dapat diidentifikasi dan diperbaiki.

Selain itu, informasi ini dapat digunakan untuk bekerja dengan pelanggan dan pemasok untuk meningkatkan proses mereka yang terhubung. Contohnya adalah generasi tindakan pada seberapa cepat pesanan pelanggan diproses atau berapa banyak pesanan diproses dalam sebulan terakhir. Langkah-langkah ini cenderung masuk ke dalam tiga kategori: waktu siklus, tingkat cacat dan produktivitas.

Tingkat pemantauan tergantung pada informasi apa yang ingin dievaluasi dan dianalisis bisnis dan bagaimana bisnis ingin dipantau, secara real-time, mendekati real-time atau ad hoc. Di sini, pemantauan kegiatan bisnis (BAM) memperluas dan memperluas alat pemantauan yang umumnya disediakan oleh BPMS.

Proses penambangan adalah kumpulan metode dan alat yang terkait dengan proses pemantauan. Tujuan dari proses penambangan adalah untuk menganalisis log peristiwa yang diekstraksi melalui pemantauan proses dan membandingkannya dengan model proses apriori. Proses penambangan memungkinkan analis proses untuk mendeteksi perbedaan antara pelaksanaan proses aktual dan model a priori serta untuk menganalisis kemacetan.

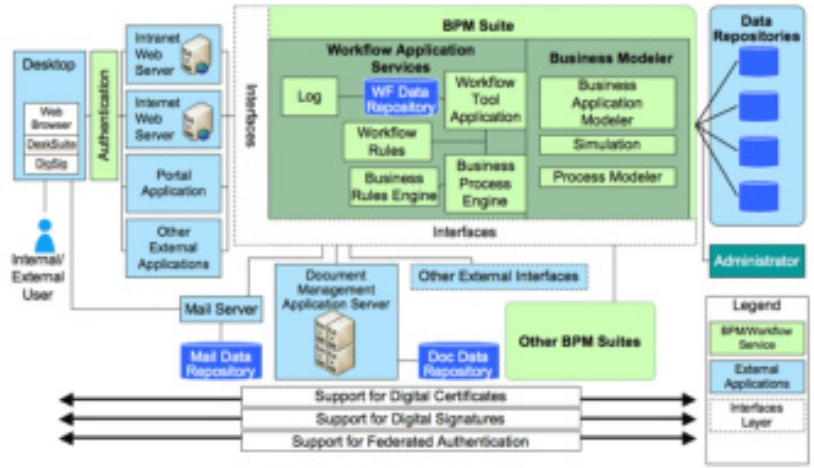
Optimasi

Optimalisasi proses mencakup pengambilan informasi kinerja proses dari tahap pemodelan atau pemantauan; mengidentifikasi potensi atau hambatan aktual dan peluang potensial untuk penghematan biaya atau peningkatan lainnya; dan kemudian, menerapkan peningkatan tersebut dalam desain proses. Alat-alat penambangan proses dapat menemukan kegiatan dan hambatan penting, menciptakan nilai bisnis yang lebih baik.

Rekayasa ulang

Ketika proses menjadi terlalu kompleks atau tidak efisien, dan optimasi tidak mengambil output yang diinginkan, biasanya direkomendasikan oleh komite pengarah perusahaan yang diketuai oleh presiden / CEO untuk merekayasa ulang seluruh siklus proses. Rekayasa ulang proses bisnis (BPR) telah digunakan oleh organisasi untuk berupaya mencapai efisiensi dan produktivitas di tempat kerja.

Penerapan



Contoh Pola *Layanan Manajemen Proses Bisnis (BPM): Pola ini menunjukkan bagaimana alat* manajemen proses *bisnis (BPM) dapat digunakan untuk mengimplementasikan proses bisnis* melalui orkestrasi *kegiatan antara orang dan sistem.*

Sementara langkah-langkah dapat dilihat sebagai siklus, kendala ekonomi atau waktu cenderung membatasi proses hanya beberapa iterasi. Ini sering terjadi ketika suatu organisasi menggunakan pendekatan untuk tujuan jangka pendek dan menengah daripada mencoba untuk mengubah budaya organisasi. Iterasi sejati hanya dimungkinkan melalui upaya kolaboratif dari peserta proses. Di sebagian besar organisasi, kompleksitas akan membutuhkan teknologi yang memungkinkan untuk mendukung peserta proses dalam tantangan manajemen proses harian ini.

Sampai saat ini, banyak organisasi sering memulai proyek atau program BPM dengan tujuan mengoptimalkan area yang telah diidentifikasi sebagai area untuk perbaikan.

Saat ini, standar internasional untuk tugas tersebut telah membatasi BPM untuk aplikasi di sektor TI, dan ISO / IEC 15944 mencakup aspek operasional bisnis. Namun, beberapa perusahaan dengan budaya praktik terbaik menggunakan prosedur operasi standar untuk mengatur proses operasional mereka. Standar lain saat ini sedang dikerjakan untuk membantu implementasi BPM (BPMN, Arsitektur Perusahaan, Model Motivasi Bisnis).

Teknologi BPM

BPM sekarang dianggap sebagai komponen penting dari solusi intelijen operasional (OI) untuk memberikan informasi real-time dan dapat ditindaklanjuti. Informasi real-time ini dapat ditindaklanjuti dengan berbagai cara - peringatan dapat dikirim atau keputusan eksekutif dapat dibuat menggunakan dasbor real-time. Solusi OI menggunakan informasi real-time untuk mengambil tindakan otomatis berdasarkan aturan yang telah ditentukan sebelumnya sehingga langkah-langkah keamanan dan atau proses manajemen pengecualian dapat dimulai.

Karena itu, beberapa orang melihat BPM sebagai “jembatan antara Teknologi Informasi (TI) dan Bisnis.” Bahkan, argumen dapat dibuat bahwa “pendekatan holistik” ini menjembatani silo organisasi dan teknologi.

Ada empat komponen penting dari BPM Suite:

- Mesin proses - platform yang kuat untuk pemodelan dan mengeksekusi aplikasi berbasis proses, termasuk aturan bisnis
- Analisis bisnis - memungkinkan manajer mengidentifikasi masalah bisnis, tren, dan peluang dengan laporan dan dasbor dan bereaksi sesuai itu

- Manajemen konten - menyediakan sistem untuk menyimpan dan mengamankan dokumen elektronik, gambar, dan file lainnya
- Alat kolaborasi - menghilangkan hambatan komunikasi intra dan interdepartemen melalui forum diskusi, ruang kerja yang dinamis, dan papan pesan

BPM juga membahas banyak masalah IT penting yang mendukung driver bisnis ini, termasuk:

- Mengelola proses end-to-end, menghadapi pelanggan
- Mengkonsolidasikan data dan meningkatkan visibilitas ke dalam dan akses ke data dan informasi terkait
- Meningkatkan fleksibilitas dan fungsionalitas infrastruktur dan data saat ini
- Mengintegrasikan dengan sistem yang ada dan meningkatkan arsitektur berorientasi layanan (SOA)
- Menetapkan bahasa umum untuk penyelarasan bisnis-TI

Validasi BPMS adalah masalah teknis lain yang perlu diperhatikan oleh vendor dan pengguna, jika kepatuhan terhadap peraturan wajib. Tugas validasi dapat dilakukan oleh pihak ketiga yang diautentikasi atau oleh pengguna sendiri. Either way, dokumentasi validasi perlu dibuat. Dokumen validasi biasanya dapat dipublikasikan secara resmi atau disimpan oleh pengguna.

Cloud Computing BPM

Manajemen proses bisnis komputasi awan adalah penggunaan alat (BPM) yang disampaikan sebagai layanan perangkat lunak (SaaS) melalui jaringan. Logika bisnis Cloud BPM digunakan pada server aplikasi dan data bisnis berada di penyimpanan cloud.

Pasar

Menurut Gartner, 20% dari semua “proses bisnis bayangan” akan didukung oleh platform cloud BPM. Gartner merujuk pada semua proses organisasi tersembunyi yang didukung oleh departemen TI sebagai bagian dari proses bisnis lama seperti spreadsheet Excel, perutean email menggunakan aturan, perutean panggilan telepon, dll. Ini dapat, tentu saja juga digantikan oleh teknologi lain seperti perangkat lunak alur kerja.

Manfaat

Manfaat menggunakan layanan cloud BPM termasuk menghilangkan kebutuhan dan biaya pemeliharaan set keterampilan teknis khusus di rumah dan mengurangi gangguan dari fokus utama perusahaan. Menawarkan penganggaran TI yang terkontrol dan memungkinkan mobilitas geografis.

Internet of Things (IoT)

Internet of Things yang muncul menimbulkan tantangan signifikan untuk mengontrol dan mengelola aliran informasi melalui sejumlah besar perangkat. Untuk mengatasinya, arah baru yang dikenal sebagai BPM Everywhere menunjukkan janji sebagai cara memadukan teknik proses tradisional, dengan kemampuan tambahan untuk mengotomatisasi penanganan semua perangkat independen.

7.5 Metadata

Metadata adalah “data [informasi] yang menyediakan informasi tentang data lain”. Ada tiga jenis metadata: metadata struktural, metadata deskriptif, dan metadata administratif. Metadata struktural adalah data tentang wadah data. Misalnya “buku” berisi data dan data tentang buku adalah metadata tentang wadah data itu.

Metadata deskriptif menggunakan setiap contoh data aplikasi atau konten data. Di banyak negara, metadata yang berkaitan dengan email, panggilan telepon, halaman web, lalu lintas video, koneksi IP dan lokasi ponsel secara rutin disimpan oleh organisasi pemerintah.

Sejarah

Metadata secara tradisional digunakan dalam katalog kartu perpustakaan sampai tahun 1980-an, ketika perpustakaan mengubah data katalog mereka menjadi basis data digital. Pada 2000-an, karena format digital menjadi cara yang lazim untuk menyimpan data dan informasi, metadata juga digunakan untuk menggambarkan data digital menggunakan standar metadata.

Ada standar metadata yang berbeda untuk setiap disiplin ilmu yang berbeda (mis., Koleksi museum, file audio digital, situs web, dll.). Menjelaskan isi dan konteks data atau file data meningkatkan kegunaannya. Misalnya, halaman web dapat menyertakan metadata yang menentukan bahasa perangkat lunak apa yang ditulis oleh halaman tersebut (mis., HTML), alat apa yang digunakan untuk membuatnya, subjek apa yang dimaksud halaman

tersebut, dan di mana menemukan informasi lebih lanjut tentang subjek tersebut. Metadata ini dapat secara otomatis meningkatkan pengalaman pembaca dan membuatnya lebih mudah bagi pengguna untuk menemukan halaman web online. CD dapat berisi metadata yang menyediakan informasi tentang musisi, penyanyi, dan penulis lagu yang karyanya muncul di disk.

Tujuan utama metadata adalah untuk membantu pengguna menemukan informasi yang relevan dan menemukan sumber daya. Metadata juga membantu mengatur sumber daya elektronik, menyediakan identifikasi digital, dan mendukung pengarsipan dan pelestarian sumber daya. Metadata membantu pengguna dalam penemuan sumber daya dengan "memungkinkan sumber daya ditemukan dengan kriteria yang relevan, mengidentifikasi sumber daya, menyatukan sumber daya yang sama, membedakan sumber daya yang berbeda, dan memberikan informasi lokasi." Metadata kegiatan telekomunikasi termasuk lalu lintas internet sangat banyak dikumpulkan oleh berbagai organisasi pemerintah nasional. Data ini digunakan untuk keperluan analisis lalu lintas dan dapat digunakan untuk pengawasan massal.

Pengertian

Metadata berarti "data tentang data". Meskipun awalan "meta" berarti «setelah» atau «di luar», ini digunakan untuk berarti «tentang» dalam epistemologi. Metadata didefinisikan sebagai data yang menyediakan informasi tentang satu atau lebih aspek data; ini digunakan untuk meringkas informasi dasar tentang data yang dapat membuat pelacakan dan bekerja dengan data spesifik lebih mudah. Beberapa contoh termasuk:

- Sarana pembuatan data
- Tujuan data
- Waktu dan tanggal pembuatan
- Pencipta atau penulis data
- Lokasi di jaringan komputer tempat data itu dibuat
- Standar yang digunakan
- Ukuran file

Misalnya, gambar digital dapat mencakup metadata yang menggambarkan seberapa besar gambar, kedalaman warna, resolusi gambar, saat gambar dibuat, kecepatan rana, dan data lainnya. Metadata dokumen teks dapat berisi informasi tentang berapa lama dokumen itu, siapa penulisnya, kapan dokumen itu ditulis, dan ringkasan pendek dokumen

itu. Meta-data dalam halaman web juga dapat berisi deskripsi konten halaman, serta kata-kata kunci yang terhubung ke konten. Tautan ini sering disebut “Metatag”, yang digunakan sebagai faktor utama dalam menentukan urutan pencarian web hingga akhir 1990-an. Ketergantungan metatag dalam pencarian web menurun pada akhir 1990-an karena “isian kata kunci”. Metatag sebagian besar disalahgunakan untuk mengelabui mesin pencari agar berpikir beberapa situs web memiliki relevansi lebih dalam pencarian daripada yang sebenarnya.

Metadata dapat disimpan dan dikelola dalam database, sering disebut metadata registry atau repositori metadata. Namun, tanpa konteks dan titik referensi, mungkin tidak mungkin untuk mengidentifikasi metadata hanya dengan melihatnya. Sebagai contoh: dengan sendirinya, sebuah basis data yang berisi beberapa angka, semua 13 digit panjangnya bisa merupakan hasil perhitungan atau daftar angka untuk dimasukkan ke dalam persamaan - tanpa konteks lain, angka itu sendiri dapat dianggap sebagai data. Tetapi jika diberikan konteks bahwa basis data ini adalah log dari koleksi buku, angka-angka 13 digit itu sekarang dapat diidentifikasi sebagai ISBN - informasi yang merujuk pada buku, tetapi bukan informasi itu sendiri dalam buku tersebut. Istilah “metadata” diciptakan pada tahun 1968 oleh Philip Bagley, dalam bukunya “Perpanjangan Konsep Bahasa Pemrograman” di mana jelas bahwa ia menggunakan istilah dalam pengertian “tradisional” ISO 11179, yang merupakan “metadata struktural” yaitu “data tentang wadah data ”; alih-alih arti alternatif “konten tentang contoh individual konten data” atau metacontent, jenis data biasanya ditemukan dalam katalog perpustakaan. Sejak itu bidang manajemen informasi, ilmu informasi, teknologi informasi, kepustakawanan, dan GIS telah secara luas mengadopsi istilah ini. Dalam bidang ini kata metadata didefinisikan sebagai “data tentang data”. Walaupun ini adalah definisi yang diterima secara umum, berbagai disiplin ilmu telah mengadopsi penjelasan dan penggunaan istilah mereka yang lebih spesifik.

Jenis-Jenis Metadata

Sementara aplikasi metadata bermacam-macam, mencakup berbagai bidang, ada model khusus dan diterima dengan baik untuk menentukan jenis metadata. Bretherton & Singley (1994) membedakan antara dua kelas yang berbeda: metadata struktural / kontrol dan metadata panduan. Meta- data struktural menggambarkan struktur objek basis data seperti tabel, kolom, kunci, dan indeks. Metadata panduan membantu manusia menemukan item tertentu dan biasanya dinyatakan sebagai serangkaian kata kunci dalam bahasa alami. Menurut Ralph Kimball, metadata dapat dibagi menjadi 2 kategori serupa: metadata teknis dan metadata bisnis. Metadata teknis sesuai dengan metadata internal,

dan metadata bisnis sesuai dengan metadata eksternal. Kimball menambahkan kategori ketiga, memproses metadata. Di sisi lain, NISO membedakan antara tiga jenis metadata: deskriptif, struktural, dan administrasi.

Metadata deskriptif biasanya digunakan untuk penemuan dan identifikasi, sebagai informasi untuk mencari dan menemukan objek, seperti judul, penulis, subjek, kata kunci, penerbit. Metadata struktural menggambarkan bagaimana komponen suatu objek diatur. Contoh metadata struktural adalah bagaimana halaman disusun untuk membentuk bab-bab sebuah buku. Akhirnya, metadata administratif memberikan informasi untuk membantu mengelola sumber. Metadata administratif mengacu pada informasi teknis, termasuk jenis file, atau kapan dan bagaimana file itu dibuat. Dua sub-jenis meta data administratif adalah metadata manajemen hak dan metadata pelestarian. Metadata manajemen hak menjelaskan hak kekayaan intelektual, sementara metadata pelestarian berisi informasi untuk melestarikan dan menghemat sumber daya.

Struktur Metadata

Metadata (metacontent) atau, lebih tepatnya, kosakata yang digunakan untuk merakit pernyataan metadata (metacon), biasanya terstruktur sesuai dengan konsep standar menggunakan skema metadata yang terdefinisi dengan baik, termasuk: standar metadata dan model metadata. Alat-alat seperti kosakata terkontrol, taksonomi, tesauri, kamus data, dan registrasi metadata dapat digunakan untuk menerapkan standardisasi lebih lanjut pada metadata. Kesamaan metadata struktural juga sangat penting dalam pengembangan model data dan dalam desain database.

Sintaksis

Sintaks metadata (metacontent) mengacu pada aturan yang dibuat untuk menyusun bidang atau elemen metadata (metacontent). Skema metadata tunggal dapat diekspresikan dalam sejumlah markup atau bahasa pemrograman yang berbeda, yang masing-masing membutuhkan sintaks yang berbeda. Misalnya, Dublin Core dapat diekspresikan dalam teks biasa, HTML, XML, dan RDF.

Contoh umum dari metacontent (panduan) adalah klasifikasi bibliografi, subjek, nomor kelas Desimal Dewey. Selalu ada pernyataan tersirat dalam “klasifikasi” beberapa objek. Untuk mengklasifikasikan objek sebagai, misalnya, kelas Dewey nomor 514 (Topologi) (yaitu buku-buku yang memiliki nomor 514 di tulang belakang mereka) pernyataan tersirat

adalah: “<book> <judul subjek> <514>. Ini adalah triple-subjek-objek-predikat, atau yang lebih penting, triple-atribut-nilai tiga. Dua elemen pertama dari triple (kelas, atribut) adalah potongan-potongan dari beberapa metadata struktural yang memiliki semantik yang ditentukan. Elemen ketiga adalah nilai, lebih disukai dari beberapa kosakata terkontrol, beberapa data referensi (master). Kombinasi dari metadata dan elemen data master menghasilkan pernyataan yang merupakan pernyataan metacontent yaitu “metacontent = metadata + data master”. Semua elemen ini dapat dianggap sebagai “kosakata”. Baik metadata dan data master adalah kosa kata yang dapat dirangkai menjadi pernyataan metacontent. Ada banyak sumber dari vocabularies ini, baik meta dan data master: UML, EDIFACT, XSD, Dewey / UDC / LoC, SKOS, ISO-25964, Pantone, Nomenklatur Binomial Linnaean, dll. Menggunakan kosakata terkontrol untuk komponen metacontent pernyataan, apakah untuk pengindeksan atau temuan, disahkan oleh ISO 25964: “Jika pengindeks dan pencari dipandu untuk memilih istilah yang sama untuk konsep yang sama, maka dokumen yang relevan akan diambil.” Ini sangat relevan ketika mempertimbangkan mesin pencari internet, seperti Google. Proses mengindeks halaman kemudian mencocokkan string teks menggunakan algoritma yang kompleks; tidak ada kecerdasan atau “kesimpulan” yang terjadi, hanya ilusi di sana.

Skema Hierarkis, Linear, dan Planar

Skema metadata bisa bersifat hierarkis di mana hubungan ada antara elemen metadata dan elemen bersarang sehingga hubungan orangtua-anak ada di antara elemen. Contoh dari skema metadata hirarkis adalah skema IEEE LOM, di mana elemen metadata dapat menjadi milik elemen metadata induk. Skema metadata juga bisa satu dimensi, atau linier, di mana setiap elemen terpisah dari elemen lain dan diklasifikasikan menurut satu dimensi saja. Contoh skema metadata linier adalah skema Dublin Core, yang satu dimensi. Skema metadata seringkali dua dimensi, atau planar, di mana setiap elemen sepenuhnya terpisah dari elemen lain tetapi diklasifikasikan menurut dua dimensi ortogonal.

Hypermapping

Dalam semua kasus di mana skema metadata melebihi penggambaran planar, beberapa jenis hypermapping diperlukan untuk memungkinkan tampilan dan tampilan metadata sesuai dengan aspek yang dipilih dan untuk melayani pandangan khusus. Hypermapping sering berlaku untuk pelapisan hamparan informasi geografis dan geologis.

Granularitas

Sejauh mana data atau metadata terstruktur disebut sebagai “granularity” -nya. “Granularity” mengacu pada seberapa banyak detail yang disediakan. Metadata dengan granularitas tinggi memungkinkan informasi yang lebih dalam, lebih terperinci, dan lebih terstruktur serta memungkinkan manipulasi teknis yang lebih tinggi. Tingkat granularitas yang lebih rendah berarti bahwa metadata dapat dibuat dengan biaya yang jauh lebih rendah tetapi tidak akan memberikan informasi yang terperinci. Dampak utama granularitas tidak hanya pada penciptaan dan penangkapan, tetapi lebih lanjut pada biaya pemeliharaan. Segera setelah struktur metadata usang, demikian juga akses ke data yang dirujuk. Karenanya granularity harus memperhitungkan upaya untuk menciptakan metadata serta upaya untuk mempertahankannya.

Standar

Standar internasional berlaku untuk metadata. Banyak pekerjaan yang sedang dilakukan dalam komunitas standar nasional dan internasional, khususnya ANSI (American National Standards Institute) dan ISO (Organisasi Internasional untuk Standardisasi) untuk mencapai konsensus tentang standardisasi metadata dan registrasi. Standar registri metadata inti adalah ISO / IEC 11179 Metadata Registries (MDR), kerangka kerja untuk standar tersebut dijelaskan dalam ISO / IEC 11179-1: 2004. Edisi baru dari Bagian 1 sedang dalam tahap akhir untuk dipublikasikan pada tahun 2015 atau awal 2016. Ini telah direvisi agar sesuai dengan edisi saat ini dari Bagian 3, ISO / IEC 11179-3: 2013 yang memperluas MDR untuk mendukung pendaftaran Sistem Konsep. Standar ini menetapkan skema untuk merekam makna dan struktur teknis data untuk penggunaan yang jelas oleh manusia dan komputer. Standar ISO / IEC 11179 mengacu pada metadata sebagai objek informasi tentang data, atau “data tentang data”. Dalam ISO / IEC 11179 Bagian-3, objek informasi adalah data tentang Elemen Data, Domain Nilai, dan objek informasi semantik dan representasional yang dapat digunakan kembali yang menjelaskan arti dan rincian teknis dari item data. Standar ini juga menentukan rincian untuk registri metadata, dan untuk mendaftarkan dan mengelola objek informasi dalam Meta-data Registry. ISO / IEC 11179 Bagian 3 juga memiliki ketentuan untuk menggambarkan struktur senyawa yang merupakan turunan dari elemen data lain, misalnya melalui perhitungan, koleksi satu atau lebih elemen data, atau bentuk lain dari data turunan. Walaupun standar ini menggambarkan dirinya sendiri awalnya sebagai “elemen data” registri, tujuannya adalah untuk mendukung menggambarkan dan mendaftarkan konten metadata secara independen dari aplikasi tertentu, meminjamkan deskripsi untuk ditemukan dan digunakan kembali oleh manusia

atau komputer dalam mengembangkan aplikasi baru, database, atau untuk analisis data yang dikumpulkan sesuai dengan konten metadata terdaftar. Standar ini telah menjadi dasar umum untuk jenis lain dari registrasi metadata, menggunakan kembali dan memperluas bagian registrasi dan administrasi standar.

Istilah metadata Dublin Core adalah seperangkat istilah kosakata yang dapat digunakan untuk menggambarkan sumber daya untuk tujuan penemuan. Rangkaian asli dari 15 istilah metadata klasik, yang dikenal sebagai Dublin Core Metadata Element Set didukung dalam dokumen standar berikut:

- IETF RFC 5013
- Standar ISO 15836-2009
- NISO Standar Z39.85.

Meskipun bukan standar, Microformat (juga disebutkan dalam bagian metadata di internet di bawah) adalah pendekatan berbasis web untuk markup semantik yang berupaya untuk menggunakan kembali tag HTML / XHTML yang ada untuk menyampaikan metadata. Microformat mengikuti standar XHTML dan HTML tetapi bukan standar itu sendiri. Salah satu penganjur dari Microformats, Tantek Çelik, menandai masalah dengan pendekatan alternatif:

Ini bahasa baru yang kami ingin Anda pelajari, dan sekarang Anda perlu meng-output file-file tambahan ini di server Anda. Ini merepotkan. (Microformats) menurunkan penghalang untuk masuk.

Penggunaan

Foto-foto

Metadata dapat ditulis ke dalam file foto digital yang akan mengidentifikasi siapa yang memilikinya, hak cipta dan informasi kontak, merek atau model kamera apa yang dibuat file, bersama dengan informasi eksposur (kecepatan rana, f-stop, dll.) Dan informasi deskriptif, seperti kata kunci tentang foto, membuat file atau gambar dapat dicari di komputer dan / atau Internet. Beberapa metadata dibuat oleh kamera dan beberapa input oleh fotografer dan / atau perangkat lunak setelah mengunduh ke komputer. Sebagian besar kamera digital menulis metadata tentang nomor model, kecepatan rana, dll., Dan beberapa memungkinkan Anda untuk mengeditnya; fungsi ini telah tersedia di sebagian besar DSLR Nikon sejak Nikon D3, pada sebagian besar kamera Canon baru sejak Canon

EOS 7D, dan pada sebagian besar Pentax DSLR sejak Pentax K-3. Metadata dapat digunakan untuk mempermudah pengorganisasian pasca produksi dengan penggunaan kata kunci. Filter dapat digunakan untuk menganalisis serangkaian foto tertentu dan membuat pilihan pada kriteria seperti peringkat atau waktu pengambilan.

Standar Metadata Fotografi diatur oleh organisasi yang mengembangkan standar berikut. Mereka termasuk, tetapi tidak terbatas pada:

- Model Pertukaran Informasi IPTC IIM (Dewan Telekomunikasi Pers Internasional),
- Skema Inti IPTC untuk XMP
- XMP - Platform Metadata yang Dapat Diperpanjang (standar ISO)
- Exif - Format file gambar yang dapat dipertukarkan, Dikelola oleh CIPA (Asosiasi Kamera & Produk Pencitraan) dan diterbitkan oleh JEITA (Asosiasi Industri Teknologi Informasi dan Elektronik Jepang)
- Dublin Core (Dublin Core Metadata Initiative - DCMI)
- PLUS (Sistem Universal Lisensi Gambar).
- VRA Core (Asosiasi Sumber Daya Visual)

Telekomunikasi

Informasi tentang waktu, asal dan tujuan panggilan telepon, pesan elektronik, pesan instan, dan mode telekomunikasi lainnya, yang bertentangan dengan konten pesan, adalah bentuk lain dari metadata. Pengumpulan sebagian besar catatan detail metadata panggilan ini oleh badan intelijen telah menimbulkan kontroversi setelah pengungkapan oleh Edward Snowden Badan intelijen seperti NSA menjaga metadata online jutaan pengguna internet hingga satu tahun, terlepas dari apakah mereka orang atau tidak. menarik bagi agensi.

Video

Metadata sangat berguna dalam video, di mana informasi tentang kontennya (seperti transkrip percakapan dan deskripsi teks dari adegannya) tidak dapat dipahami secara langsung oleh komputer, tetapi di mana pencarian konten yang efisien diperlukan. Ada dua sumber di mana metadata video diturunkan: (1) metadata yang dikumpulkan secara operasional, yaitu informasi tentang konten yang dihasilkan, seperti jenis peralatan, perangkat lunak, tanggal, dan lokasi; (2) metadata yang ditulis manusia, untuk meningkatkan visibilitas mesin pencari, kemampuan menemukan, keterlibatan pemirsa, dan memberikan peluang iklan kepada penerbit video. Dalam masyarakat saat ini, sebagian besar perangkat

lunak pengeditan video profesional memiliki akses ke metadata. Avid's MetaSync dan Adobe's Bridge adalah dua contoh utama dari ini.

Halaman web

Halaman web sering menyertakan metadata dalam bentuk meta tag. Deskripsi dan kata kunci dalam tag meta biasanya digunakan untuk menggambarkan konten halaman Web. Elemen meta juga menentukan deskripsi halaman, kata-kata kunci, penulis dokumen, dan kapan dokumen terakhir diubah. Metadata halaman web membantu mesin pencari dan pengguna untuk menemukan jenis halaman web yang mereka cari.

Penciptaan

Metadata dapat dibuat baik dengan pemrosesan informasi otomatis atau dengan pekerjaan manual. Metadata elementer yang ditangkap oleh komputer dapat mencakup informasi tentang kapan objek dibuat, siapa yang membuatnya, kapan terakhir diperbarui, ukuran file, dan ekstensi file. Dalam konteks ini objek mengacu pada salah satu dari yang berikut:

- Benda fisik seperti buku, CD, DVD, peta kertas, kursi, meja, pot bunga, dll.
- File elektronik seperti gambar digital, foto digital, dokumen elektronik, file program, tabel database, dll.

Virtualisasi data telah muncul di tahun 2000-an sebagai teknologi perangkat lunak baru untuk melengkapi "tumpukan" virtualisasi di perusahaan. Metadata digunakan dalam server virtualisasi data yang merupakan komponen infrastruktur perusahaan, bersama dengan database dan server aplikasi. Metadata di server ini disimpan sebagai repositori persisten dan menggambarkan objek bisnis di berbagai sistem dan aplikasi perusahaan. Kesamaan metadata struktural juga penting untuk mendukung virtualisasi data.

Layanan Statistik dan Sensus

Pekerjaan standardisasi memiliki dampak besar pada upaya membangun sistem metadata dalam komunitas statistik. Beberapa standar metadata dijelaskan, dan pentingnya mereka untuk lembaga statistik dibahas. Penerapan standar di Biro Sensus, Badan Perlindungan Lingkungan, Biro Statistik Tenaga Kerja, Statistik Kanada, dan banyak lainnya dijelaskan. Penekanan pada dampak yang dapat dimiliki oleh registri metadata di lembaga statistik.

Perpustakaan dan Informasi

Metadata telah digunakan dengan berbagai cara sebagai cara katalogisasi item di perpustakaan dalam format digital dan analog. Data tersebut membantu mengklasifikasikan, mengagregasi, mengidentifikasi, dan menemukan buku, DVD, majalah atau objek apa pun yang mungkin dimiliki oleh perpustakaan dalam koleksinya. Sampai tahun 1980-an, banyak katalog perpustakaan menggunakan kartu 3x5 inci dalam laci file untuk menampilkan judul buku, penulis, materi pelajaran, dan string alfa-numerik singkat (nomor panggilan) yang menunjukkan lokasi fisik buku di rak perpustakaan. Sistem Desimal Dewey yang digunakan oleh perpustakaan untuk klasifikasi bahan pustaka menurut subjek adalah contoh awal penggunaan metadata. Mulai tahun 1980-an dan 1990-an, banyak perpustakaan mengganti kartu file kertas ini dengan database komputer. Basis data komputer ini memudahkan dan lebih cepat bagi pengguna untuk melakukan pencarian kata kunci. Bentuk lain dari koleksi metadata yang lebih lama adalah penggunaan oleh Biro Sensus AS tentang apa yang dikenal sebagai "Formulir Panjang." Formulir Panjang mengajukan pertanyaan yang digunakan untuk membuat data demografis untuk menemukan pola distribusi. Perpustakaan menggunakan metadata dalam katalog perpustakaan, paling umum sebagai bagian dari Sistem Manajemen Perpustakaan Terpadu. Metadata diperoleh dengan membuat katalog sumber daya seperti buku, majalah, DVD, halaman web atau gambar digital. Data ini disimpan dalam sistem manajemen perpustakaan terintegrasi, ILMS, menggunakan standar metadata MARC. Tujuannya adalah untuk mengarahkan pelanggan ke lokasi fisik atau elektronik dari barang atau area yang mereka cari serta untuk memberikan deskripsi barang yang dimaksud.

Contoh yang lebih baru dan khusus dari metadata perpustakaan termasuk pembentukan perpustakaan digital termasuk repositori e-print dan perpustakaan gambar digital. Meskipun sering didasarkan pada prinsip-prinsip perpustakaan, fokus pada penggunaan non-pustakawan, terutama dalam menyediakan metadata, berarti mereka tidak mengikuti pendekatan katalog tradisional atau umum. Mengingat sifat khusus dari materi yang disertakan, bidang metadata sering kali dibuat secara khusus, mis. bidang klasifikasi taksonomi, bidang lokasi, kata kunci, atau pernyataan hak cipta. Informasi file standar seperti ukuran dan format file biasanya disertakan secara otomatis. Operasi perpustakaan selama beberapa dekade telah menjadi topik utama dalam upaya menuju standardisasi internasional. Standar untuk metadata di perpustakaan digital termasuk Dublin Core, METS, MODS, DDI, DOI, URN, skema PREMIS, EML, dan OAI-PMH. Perpustakaan terkemuka di dunia memberikan petunjuk tentang strategi standar metadata mereka.

Di Museum

Metadata dalam konteks museum adalah informasi yang melatih spesialis dokumentasi budaya, seperti arsiparis, pustakawan, pendaftar dan kurator museum, membuat indeks, struktur, menggambarkan, mengidentifikasi, atau menentukan karya seni, arsitektur, objek budaya dan gambar mereka. Metadata deskriptif paling umum digunakan dalam konteks museum untuk identifikasi objek dan tujuan pemulihan sumber daya.

Pemakaian

Metadata dikembangkan dan diterapkan dalam lembaga dan museum pengumpul untuk:

- Memfasilitasi penemuan sumber daya dan menjalankan permintaan pencarian.
- Membuat arsip digital yang menyimpan informasi yang berkaitan dengan berbagai aspek koleksi museum dan benda budaya, dan berfungsi untuk keperluan arsip dan manajerial.
- Menyediakan audiensi publik akses ke objek budaya melalui penerbitan konten digital online.

Standar

Banyak museum dan pusat warisan budaya mengakui bahwa mengingat keragaman karya seni dan benda budaya, tidak ada model atau standar tunggal yang cukup untuk menggambarkan dan membuat katalog karya budaya. Misalnya, artefak Pribumi yang terpahat dapat digolongkan sebagai karya seni, artefak arkeologis, atau benda warisan Pribumi. Tahap awal standardisasi dalam pengarsipan, deskripsi dan katalogisasi dalam komunitas museum dimulai pada akhir 1990-an dengan pengembangan standar seperti Kategori untuk Deskripsi Karya Seni (CDWA), Spektrum, Model Referensi Konseptual (CIDOC), Katalogisasi Objek Budaya (CCO) dan skema XML CDWA Lite. Standar-standar ini menggunakan bahasa markup HTML dan XML untuk pemrosesan, publikasi, dan implementasi mesin. The Anglo-American Cataloging Rules (AACR), yang awalnya dikembangkan untuk mengkarakterisasi buku, juga telah diterapkan pada objek budaya, karya seni dan arsitektur. Standar, seperti CCO, diintegrasikan ke dalam Sistem Manajemen Koleksi Museum (CMS), basis data di mana museum dapat mengelola koleksi, akuisisi, pinjaman, dan konservasi mereka. Para sarjana dan profesional di lapangan mencatat bahwa “lanskap standar dan teknologi yang berkembang dengan cepat” menciptakan tantangan bagi para dokumenter budaya, khususnya para profesional yang tidak terlatih secara teknis. Sebagian besar lembaga dan museum pengumpul menggunakan basis

data relasional untuk mengkategorikan karya budaya dan gambar mereka. Basis data dan metadata relasional berfungsi untuk mendokumentasikan dan menggambarkan hubungan kompleks antara objek budaya dan karya seni beragam, serta antara objek dan tempat, orang dan gerakan artistik. Struktur basis data relasional juga bermanfaat dalam mengumpulkan lembaga dan museum karena memungkinkan arsiparis untuk membuat perbedaan yang jelas antara benda budaya dan gambar-gambarnya; perbedaan yang tidak jelas dapat menyebabkan pencarian membingungkan dan tidak akurat.

Benda Budaya dan Karya Seni

Materialitas suatu benda, fungsi dan tujuan, serta ukuran (misalnya, pengukuran, seperti tinggi, lebar, berat), persyaratan penyimpanan (misalnya, lingkungan yang dikendalikan iklim) dan fokus museum dan koleksi, memengaruhi kedalaman deskriptif dari data yang dikaitkan dengan objek oleh dokumenter budaya. Praktek katalogisasi kelembagaan yang mapan, tujuan dan keahlian para dokumenter budaya dan struktur database juga memengaruhi informasi yang dianggap berasal dari objek budaya, dan cara-cara di mana objek budaya dikategorikan. Selain itu, museum sering menggunakan perangkat lunak manajemen koleksi komersial standar yang mengatur dan membatasi cara para arsiparis dapat menggambarkan karya seni dan benda budaya. Selain itu, lembaga dan museum pengumpul menggunakan Kosakata Terkendali untuk menggambarkan objek budaya dan karya seni dalam koleksi mereka. Getty Vocabularies dan Library of Congress Controlled Vocabularies memiliki reputasi baik dalam komunitas museum dan direkomendasikan oleh standar CCO. Museum didorong untuk menggunakan kosakata terkontrol yang kontekstual dan relevan dengan koleksi mereka dan meningkatkan fungsionalitas sistem informasi digital mereka. Controlled Vocabularies bermanfaat dalam database karena memberikan tingkat konsistensi yang tinggi, meningkatkan pengambilan sumber daya. Struktur metadata, termasuk kosakata terkontrol, mencerminkan ontologi sistem dari mana mereka diciptakan. Seringkali proses melalui mana benda-benda budaya dijelaskan dan dikategorikan melalui metadata di museum tidak mencerminkan perspektif masyarakat pembuat.

Museum dan Internet

Metadata telah berperan dalam penciptaan sistem informasi digital dan arsip di dalam museum, dan membuatnya lebih mudah bagi museum untuk menerbitkan konten digital secara online. Ini telah memungkinkan audiens yang mungkin tidak memiliki akses ke objek budaya karena hambatan geografis atau ekonomi untuk memiliki akses ke mereka. Pada tahun 2000-an, karena semakin banyak museum yang mengadopsi standar

kearsipan dan menciptakan basis data yang rumit, diskusi tentang Linked Data antara basis data museum muncul di komunitas museum, arsip, dan ilmu perpustakaan. Sistem Manajemen Pengumpulan (CMS) dan alat Manajemen Aset Digital dapat berupa sistem lokal atau bersama. Sarjana Humaniora Digital mencatat banyak manfaat interoperabilitas antara database dan koleksi museum, sementara juga mengakui kesulitan mencapai interoperabilitas tersebut.

Hukum

Amerika Serikat

Masalah yang melibatkan metadata dalam litigasi di Amerika Serikat semakin meluas. Pengadilan telah melihat berbagai pertanyaan yang melibatkan metadata, termasuk kemampuan menemukan metadata oleh para pihak. Meskipun Aturan Federal Prosedur Sipil hanya menetapkan aturan tentang dokumen elektronik, hukum kasus berikutnya telah diuraikan tentang persyaratan pihak untuk mengungkapkan metadata. Pada Oktober 2009, Mahkamah Agung Arizona memutuskan bahwa catatan metadata adalah catatan publik. Metadata dokumen telah terbukti sangat penting dalam lingkungan hukum di mana litigasi telah meminta metadata, yang dapat mencakup informasi sensitif yang merugikan pihak tertentu di pengadilan. Menggunakan alat penghapus metadata untuk “membersihkan” atau mengurangi dokumen dapat mengurangi risiko tanpa disadari mengirim data sensitif. Proses ini secara parsial melindungi firma hukum dari kemungkinan merusak kebocoran data sensitif melalui penemuan elektronik.

Australia

Di Australia, kebutuhan untuk memperkuat keamanan nasional telah menghasilkan pengenalan hukum penyimpanan metadata baru. Undang-undang baru ini berarti bahwa kedua lembaga keamanan dan kepolisian akan diizinkan untuk mengakses hingga dua tahun metadata individu, yang seharusnya membuatnya lebih mudah untuk menghentikan serangan teroris dan kejahatan serius terjadi. Pada tahun 2000-an, undang-undang tidak mengizinkan akses ke konten pesan orang, panggilan telepon atau email dan riwayat penelusuran web, tetapi ketentuan ini dapat diubah oleh pemerintah.

Di bidang Kesehatan

Penelitian medis Australia memelopori definisi metadata untuk aplikasi dalam perawatan kesehatan. Pendekatan itu menawarkan upaya pertama yang diakui untuk mematuhi standar internasional dalam ilmu kedokteran alih-alih mendefinisikan standar kepemilikan di bawah payung Organisasi Kesehatan Dunia (WHO). Komunitas medis belum menyetujui

perlunya mengikuti standar metadata meskipun penelitian yang mendukung standar ini.

Pergudangan Data

Data warehouse (DW) adalah repositori dari data yang disimpan secara elektronik dari suatu organisasi. Gudang data dirancang untuk mengelola dan menyimpan data. Gudang data berbeda dari sistem intelijen bisnis (BI), karena sistem BI dirancang untuk menggunakan data untuk membuat laporan dan menganalisis informasi, untuk memberikan panduan strategis bagi manajemen. Metadata adalah alat penting dalam bagaimana data disimpan di gudang data. Tujuan dari gudang data adalah untuk menampung data yang terstandarisasi, terstruktur, konsisten, terintegrasi, benar, "dibersihkan" dan tepat waktu, diekstraksi dari berbagai sistem operasional dalam suatu organisasi. Data yang diekstraksi terintegrasi dalam lingkungan data warehouse untuk memberikan perspektif enterprisewide. Data disusun dengan cara melayani persyaratan pelaporan dan analisis. Desain kesamaan struktural metadata menggunakan metode pemodelan data seperti diagram model hubungan entitas penting dalam setiap upaya pengembangan data warehouse. Mereka merinci metadata pada setiap bagian data di gudang data. Komponen penting dari data warehouse / sistem intelijen bisnis adalah metadata dan alat untuk mengelola dan mengambil metadata. Ralph Kimball menggambarkan metadata sebagai DNA dari data warehouse ketika metadata mendefinisikan elemen-elemen dari data warehouse dan bagaimana mereka bekerja bersama.

Kimball dkk. merujuk pada tiga kategori utama metadata: Metadata teknis, metadata bisnis, dan metadata proses. Metadata teknis terutama definisi, sedangkan metadata bisnis dan metadata proses terutama deskriptif. Kategori terkadang tumpang tindih.

- Metadata teknis mendefinisikan objek dan proses dalam sistem DW / BI, seperti yang terlihat dari sudut pandang teknis. Metadata teknis mencakup metadata sistem, yang mendefinisikan struktur data seperti tabel, bidang, tipe data, indeks dan partisi dalam mesin relasional, serta basis data, dimensi, ukuran, dan model penambangan data. Metadata teknis mendefinisikan model data dan cara ditampilkan untuk pengguna, dengan laporan, jadwal, daftar distribusi, dan hak keamanan pengguna.
- Metadata bisnis adalah konten dari data warehouse yang dijelaskan dalam istilah yang lebih ramah pengguna. Metadata bisnis memberi tahu Anda data apa yang Anda miliki, dari mana asalnya, apa artinya dan apa hubungannya dengan data lain di gudang

data. Metadata bisnis juga dapat berfungsi sebagai dokumentasi untuk sistem DW / BI. Pengguna yang menelusuri gudang data terutama melihat metadata bisnis.

- Metadata proses digunakan untuk menggambarkan hasil berbagai operasi di gudang data. Dalam proses ETL, semua data utama dari tugas dicatat saat eksekusi. Ini termasuk waktu mulai, waktu selesai, detik CPU digunakan, disk membaca, menulis disk, dan baris diproses. Saat memecahkan masalah ETL atau proses kueri, data semacam ini menjadi berharga. Metadata proses adalah pengukuran fakta saat membangun dan menggunakan sistem DW / BI. Beberapa organisasi mencari nafkah dengan mengumpulkan dan menjual data semacam ini kepada perusahaan - dalam hal ini metadata proses menjadi metadata bisnis untuk tabel fakta dan dimensi. Metadata proses pengumpulan adalah untuk kepentingan pelaku bisnis yang dapat menggunakan data untuk mengidentifikasi pengguna produk mereka, produk mana yang mereka gunakan, dan tingkat layanan apa yang mereka terima.

Di internet

Format HTML yang digunakan untuk mendefinisikan halaman web memungkinkan untuk dimasukkannya berbagai jenis metadata, dari teks deskriptif dasar, tanggal dan kata kunci hingga skema metadata lanjutan seperti Dublin Core, e-GMS, dan standar AGLS. Halaman juga dapat di-geotag dengan koordinat. Metadata dapat dimasukkan dalam tajuk halaman atau dalam file terpisah. Microformats memungkinkan metadata ditambahkan ke data di halaman dengan cara yang tidak dilihat oleh pengguna web biasa, tetapi komputer, perayap web, dan mesin pencari dapat dengan mudah mengakses. Banyak mesin pencari yang berhati-hati dalam menggunakan metadata dalam algoritma peringkat mereka karena eksploitasi metadata dan praktik optimasi mesin pencari, SEO, untuk meningkatkan peringkat. Lihat artikel elemen Meta untuk diskusi lebih lanjut. Sikap berhati-hati ini dapat dibenarkan karena orang, menurut Doctorow, tidak melakukan perawatan dan ketekunan saat membuat metadata mereka sendiri dan bahwa metadata adalah bagian dari lingkungan kompetitif di mana metadata digunakan untuk mempromosikan tujuan pembuat metadata itu sendiri. Studi menunjukkan bahwa mesin pencari merespons halaman web dengan implementasi metadata, dan Google memiliki pengumuman di situsnya yang menunjukkan tag meta yang dimengerti oleh mesin pencariannya. Startup pencarian perusahaan, Swiftype, mengenali metadata sebagai sinyal relevansi yang dapat diterapkan oleh webmaster untuk mesin pencari khusus situs web mereka, bahkan merilis ekstensi mereka sendiri, yang dikenal sebagai Meta Tag 2.

Dalam Industri Siaran

Dalam industri penyiaran, metadata ditautkan ke media penyiaran audio dan video ke:

- mengidentifikasi media: nama klip atau daftar putar, durasi, kode waktu, dll.
- menjelaskan konten: catatan mengenai kualitas konten video, peringkat, deskripsi (misalnya, selama acara olahraga, kata kunci seperti gol, kartu merah akan dikaitkan dengan beberapa klip)
- mengklasifikasikan media: metadata memungkinkan untuk menyortir media atau untuk dengan mudah dan cepat menemukan konten video (berita TV bisa sangat membutuhkan beberapa konten arsip untuk subjek). Sebagai contoh, BBC memiliki sistem klasifikasi subjek besar, Lonclass, versi khusus dari Klasifikasi Desimal Universal Universal yang lebih umum.

Metadata ini dapat ditautkan ke media video berkat server video. Kebanyakan acara olahraga siaran utama seperti FIFA World Cup atau Olympic Games menggunakan metadata ini untuk mendistribusikan konten video mereka ke stasiun TV melalui kata kunci. Seringkali host penyiar yang bertugas mengatur metadata melalui International Broadcast Center dan server videonya. Metadata ini direkam dengan gambar dan dimasukkan oleh operator metadata (penebang) yang mengaitkannya dalam metadata langsung yang tersedia di kisi metadata melalui perangkat lunak (seperti Multicam (LSM) atau IPDirector yang digunakan selama Piala Dunia FIFA atau Olimpiade).

Geospasial

Metadata yang menggambarkan objek geografis dalam penyimpanan atau format elektronik (seperti kumpulan data, peta, fitur, atau dokumen dengan komponen geospasial) memiliki sejarah sejak setidaknya tahun 1994 (lihat halaman Perpustakaan MIT di FGDC Metadata). Kelas metadata ini dijelaskan lebih lengkap pada artikel metadata geospasial.

Ekologis dan Lingkungan

Metadata ekologi dan lingkungan dimaksudkan untuk mendokumentasikan “siapa, apa, kapan, di mana, mengapa, dan bagaimana” pengumpulan data untuk studi tertentu. Ini biasanya berarti organisasi atau lembaga mana yang mengumpulkan data, jenis data apa, tanggal mana data itu dikumpulkan, alasan pengumpulan data, dan metodologi yang digunakan untuk pengumpulan data. Metadata harus dihasilkan dalam format yang umum digunakan oleh komunitas sains yang paling relevan, seperti Darwin Core, Ecological Metadata Language, atau Dublin Core. Alat pengeditan metadata ada untuk memfasilitasi pembuatan metadata (mis. Metavist, Mercury: Metadata Search System,

Morpho). Metadata harus mendeskripsikan asal-usul data (dari mana asalnya, serta setiap transformasi yang dilakukan data) dan bagaimana memberikan kredit untuk (mengutip) produk data.

Musik Digital

Ketika pertama kali dirilis pada tahun 1982, Compact Disc hanya berisi Table Of Contents (TOC) dengan jumlah trek pada disk dan panjangnya dalam sampel. Empat belas tahun kemudian pada tahun 1996, revisi standar Buku Merah CD menambahkan CD-Teks untuk membawa metadata tambahan. Tetapi CD-Text tidak diadopsi secara luas. Tak lama kemudian, menjadi umum bagi komputer pribadi untuk mengambil metadata dari sumber eksternal (mis. CDDB, Gracenote) berdasarkan TOC.

Format audio digital seperti file audio digital menggantikan format musik seperti kaset dan CD pada tahun 2000-an. File audio digital dapat diberi label dengan informasi lebih banyak daripada yang dapat terkandung hanya dalam nama file. Informasi deskriptif itu disebut tag audio atau metadata audio secara umum. Program komputer yang mengkhususkan diri dalam menambah atau memodifikasi informasi ini disebut editor tag. Metadata dapat digunakan untuk memberi nama, menggambarkan, katalog, dan menunjukkan kepemilikan atau hak cipta untuk file audio digital, dan keberadaannya membuatnya lebih mudah untuk menemukan file audio tertentu dalam grup, biasanya melalui penggunaan mesin pencari yang mengakses metadata. Ketika berbagai format audio digital dikembangkan, berbagai upaya dilakukan untuk menstandarkan lokasi tertentu dalam file digital tempat informasi ini dapat disimpan.

Akibatnya, hampir semua format audio digital, termasuk mp3, broadcast wav dan file AIFF, memiliki lokasi standar serupa yang dapat diisi dengan metadata. Metadata untuk musik digital terkompresi dan tidak terkompresi sering kali dikodekan dalam tag ID3. Editor umum seperti TagLib mendukung MP3, Ogg Vorbis, FLAC, MPC, Speex, WavPack TrueAudio, WAV, AIFF, MP4, dan format file ASF.

Aplikasi Cloud

Dengan ketersediaan aplikasi Cloud, yang termasuk yang menambahkan metadata ke konten, metadata semakin tersedia melalui Internet.

Administrasi dan Manajemen

Penyimpanan

Metadata dapat disimpan baik secara internal, dalam file atau struktur yang sama dengan data (ini juga disebut metadata tertanam), atau secara eksternal, dalam file atau bidang terpisah dari data yang dijelaskan. Penyimpanan data biasanya menyimpan metadata yang terlepas dari data, tetapi dapat dirancang untuk mendukung pendekatan metadata yang disematkan. Setiap opsi memiliki kelebihan dan kekurangan:

- Penyimpanan internal berarti metadata selalu bergerak sebagai bagian dari data yang mereka gambarkan; dengan demikian, metadata selalu tersedia dengan data, dan dapat dimanipulasi secara lokal. Metode ini menciptakan redundansi (menghalangi normalisasi), dan tidak memungkinkan mengelola semua metadata sistem di satu tempat. Ini bisa dibilang meningkatkan konsistensi, karena metadata mudah diubah setiap kali data diubah.
- Penyimpanan eksternal memungkinkan kolokasi metadata untuk semua konten, misalnya dalam database, untuk pencarian dan manajemen yang lebih efisien. Redundansi dapat dihindari dengan menormalkan organisasi metadata. Dalam pendekatan ini, metadata dapat disatukan dengan konten saat informasi ditransfer, misalnya dalam media streaming; atau dapat dirujuk (misalnya, sebagai tautan web) dari konten yang ditransfer. Di sisi bawah, pembagian metadata dari konten data, terutama dalam file mandiri yang merujuk ke metadata sumber mereka di tempat lain, meningkatkan peluang untuk ketidaksejajaran antara keduanya, karena perubahan pada keduanya mungkin tidak tercermin di yang lain.

Metadata dapat disimpan dalam bentuk yang bisa dibaca manusia atau biner. Menyimpan metadata dalam format yang dapat dibaca manusia seperti XML dapat bermanfaat karena pengguna dapat memahami dan mengeditnya tanpa alat khusus. Di sisi lain, format ini jarang dioptimalkan untuk kapasitas penyimpanan, waktu komunikasi, dan kecepatan pemrosesan. Format metadata biner memungkinkan efisiensi dalam semua hal ini, tetapi membutuhkan perpustakaan khusus untuk mengubah informasi biner menjadi konten yang dapat dibaca manusia.

Manajemen Basis Data

Setiap sistem basis data relasional memiliki mekanisme sendiri untuk menyimpan metadata. Contoh metadata database relasional meliputi:

- Tabel semua tabel dalam database, nama, ukuran, dan jumlah baris di setiap tabel.

- Tabel kolom di setiap basis data, tabel apa yang digunakan, dan jenis data yang disimpan di setiap kolom.

Dalam terminologi basis data, kumpulan metadata ini disebut sebagai katalog. Standar SQL menentukan cara seragam untuk mengakses katalog, yang disebut skema informasi, tetapi tidak semua basis data mengimplementasikannya, bahkan jika mereka menerapkan aspek lain dari standar SQL. Untuk contoh metode akses metadata khusus basis data, lihat Oracle metadata. Akses terprogram ke metadata dimungkinkan menggunakan API seperti JDBC, atau SchemaCrawler.

7.6 Root-cause Analysis

Root cause analysis (RCA) adalah metode penyelesaian masalah yang digunakan untuk mengidentifikasi akar penyebab kesalahan atau masalah. Suatu faktor dianggap sebagai penyebab utama jika penghapusannya dari urutan masalah-kesalahan mencegah acara yang tidak diinginkan dari berulang; sedangkan faktor penyebab adalah faktor yang mempengaruhi hasil acara, tetapi bukan merupakan penyebab utama. Meskipun menghilangkan faktor penyebab dapat menguntungkan hasil, itu tidak mencegah terulangnya dengan pasti.

Misalnya, bayangkan segmen fiksi siswa yang menerima nilai ujian buruk. Setelah penyelidikan awal, diverifikasi bahwa siswa yang mengambil tes pada periode akhir hari sekolah mendapat skor lebih rendah. Penyelidikan lebih lanjut mengungkapkan bahwa pada sore hari, para siswa tidak memiliki kemampuan untuk fokus. Bahkan penyelidikan lebih lanjut mengungkapkan bahwa alasan kurangnya fokus adalah kelaparan. Jadi, akar penyebab dari skor pengujian yang buruk adalah kelaparan, diatasi dengan memindahkan waktu pengujian segera setelah makan siang.

Sebagai contoh lain, bayangkan investigasi terhadap mesin yang berhenti karena kelebihan beban dan sekring meledak. Investigasi menunjukkan bahwa mesin kelebihan beban karena memiliki bantalan yang tidak cukup dilumasi. Investigasi berlanjut lebih lanjut dan menemukan bahwa mekanisme pelumasan otomatis memiliki pompa yang tidak memompa cukup, sehingga kurangnya pelumasan. Investigasi pompa menunjukkan bahwa ia memiliki poros yang aus. Investigasi mengapa poros itu dipakai menemukan bahwa tidak ada mekanisme yang memadai untuk mencegah masuknya logam ke pompa. Memo ini memungkinkan masuk ke pompa, dan merusaknya. Akar penyebab masalah adalah karena itu, potongan logam dapat mencemari sistem pelumasan. Memperbaiki

masalah ini seharusnya mencegah seluruh urutan kejadian berulang. Bandingkan ini dengan penyelidikan yang tidak menemukan akar penyebabnya: mengganti sekring, bantalan, atau pompa pelumasan mungkin akan memungkinkan mesin untuk kembali beroperasi untuk sementara waktu. Tetapi ada risiko bahwa masalah itu hanya akan muncul kembali, sampai akar permasalahan diselesaikan.

Menyusul pengantar analisis Kepner – Tregoe — yang memiliki keterbatasan dalam arena desain, pengembangan, dan peluncuran roket yang sangat kompleks — RCA muncul pada 1950-an sebagai studi formal oleh Badan Penerbangan dan Antariksa Nasional (NASA) di Amerika Serikat. Metode baru analisis masalah yang dikembangkan oleh NASA termasuk praktik penilaian tingkat tinggi yang disebut MORT (Manajemen Risiko Pengawasan Pohon). MORT berbeda dari RCA dengan menetapkan penyebab untuk kelas umum dari penyebab kekurangan yang dapat diringkas ke dalam daftar pendek. Ini termasuk praktik kerja, prosedur, manajemen, kelelahan, tekanan waktu, dan beberapa lainnya. Misalnya: jika kecelakaan pesawat terjadi sebagai akibat dari kondisi cuaca buruk ditambah tekanan untuk pergi tepat waktu; kegagalan untuk mengamati tindakan pencegahan cuaca dapat mengindikasikan masalah manajemen atau pelatihan; dan kurangnya kekhawatiran cuaca yang tepat dapat mendakwa praktik kerja. Karena beberapa tindakan (metode) dapat secara efektif mengatasi akar penyebab masalah, RCA adalah proses berulang dan alat perbaikan terus menerus.

RCA diterapkan untuk mengidentifikasi dan mengoreksi akar penyebab peristiwa secara metodis, daripada sekadar mengatasi hasil simptomatik. Koreksi fokus pada akar penyebab memiliki tujuan sepenuhnya mencegah terulangnya masalah. Sebaliknya, RCFA (Root Cause Failure Analysis) mengakui bahwa pencegahan kekambuhan dengan satu tindakan korektif tidak selalu memungkinkan.

RCA biasanya digunakan sebagai metode reaktif untuk mengidentifikasi penyebab, mengungkap masalah dan menyelesaikannya. Analisis dilakukan setelah suatu peristiwa terjadi. Wawasan dalam RCA membuatnya berpotensi berguna sebagai metode pencegahan. Dalam peristiwa itu, RCA dapat digunakan untuk memperkirakan atau memprediksi peristiwa yang mungkin terjadi bahkan sebelum terjadi. Sementara satu mengikuti yang lain, RCA adalah proses yang sepenuhnya terpisah untuk manajemen insiden.

Daripada satu metodologi yang didefinisikan secara tajam, RCA terdiri dari banyak alat, proses, dan filosofi yang berbeda. Namun, beberapa pendekatan atau “sekolah” yang

didefinisikan secara luas dapat diidentifikasi dengan pendekatan dasar atau bidang asalnya: berbasis keselamatan, berbasis produksi, berbasis perakitan, berbasis proses, berbasis kegagalan, dan berbasis sistem.

- RCA berbasis keselamatan muncul dari bidang analisis kecelakaan dan keselamatan dan kesehatan kerja.
- RCA berbasis produksi memiliki akar di bidang kontrol kualitas untuk manufaktur industri.
- RCA berbasis proses, tindak lanjut dari RCA berbasis produksi, memperluas ruang lingkup RCA untuk mencakup proses bisnis.
- RCA berbasis kegagalan berasal dari praktik analisis kegagalan seperti yang digunakan dalam rekayasa dan pemeliharaan.
- RCA berbasis sistem telah muncul sebagai campuran dari sekolah-sekolah sebelumnya, memasukkan unsur-unsur dari bidang lain seperti manajemen perubahan, manajemen risiko dan analisis sistem.

Terlepas dari pendekatan yang berbeda di antara berbagai aliran analisis akar penyebab, semuanya berbagi beberapa prinsip umum. Beberapa proses umum untuk melakukan RCA juga dapat didefinisikan.

Prinsip-Prinsip Umum

1. Tujuan utama dari analisis akar permasalahan adalah: untuk mengidentifikasi faktor-faktor yang mengakibatkan sifat, besarnya, lokasi, dan waktu dari hasil (konsekuensi) berbahaya dari satu atau lebih peristiwa di masa lalu; untuk menentukan perilaku, tindakan, kelambanan, atau kondisi apa yang perlu diubah; untuk mencegah terulangnya hasil berbahaya yang serupa; dan untuk mengidentifikasi pelajaran yang dapat mendorong pencapaian konsekuensi yang lebih baik. (“Sukses” didefinisikan sebagai pencegahan kekambuhan yang hampir pasti.)
2. Agar efektif, analisis akar masalah harus dilakukan secara sistematis, biasanya sebagai bagian dari penyelidikan, dengan kesimpulan dan akar penyebab yang diidentifikasi didukung oleh bukti yang terdokumentasi. Upaya tim biasanya diperlukan.
3. Mungkin ada lebih dari satu penyebab utama untuk suatu peristiwa atau masalah, karenanya bagian yang sulit menunjukkan kegigihan dan mempertahankan upaya yang diperlukan untuk menentukannya.
4. Tujuan mengidentifikasi semua solusi untuk masalah adalah untuk mencegah terulangnya dengan biaya terendah dengan cara paling sederhana. Jika ada alternatif

yang sama efektifnya, maka pendekatan biaya yang paling sederhana atau paling rendah lebih disukai.

5. Penyebab utama yang diidentifikasi akan tergantung pada cara masalah atau peristiwa tersebut didefinisikan. Pernyataan masalah dan deskripsi peristiwa yang efektif (seperti kegagalan, misalnya) sangat membantu dan biasanya diperlukan untuk memastikan pelaksanaan analisis yang sesuai.
6. Salah satu cara logis untuk melacak akar penyebabnya adalah dengan memanfaatkan solusi penambangan data hierarkis clustering (seperti penambangan data berbasis teori-grafik). Penyebab root didefinisikan dalam konteks itu sebagai "kondisi yang memungkinkan satu atau lebih penyebab". Penyebab root dapat secara deduktif disortir dari kelompok atas yang di dalamnya kelompok memasukkan penyebab spesifik.
7. Agar efektif, analisis harus menetapkan urutan peristiwa atau garis waktu untuk memahami hubungan antara faktor-faktor penyebab (penyebab), penyebab utama (s) dan masalah atau peristiwa yang ditentukan untuk dicegah.
8. *Root cause analysis* dapat membantu mengubah budaya reaktif (yang bereaksi terhadap masalah) menjadi budaya yang berwawasan ke depan (yang memecahkan masalah sebelum terjadi atau meningkat). Lebih penting lagi, RCA mengurangi frekuensi masalah yang terjadi dari waktu ke waktu di lingkungan tempat proses itu digunakan.
9. *Root cause analysis* sebagai kekuatan untuk perubahan adalah ancaman bagi banyak budaya dan lingkungan. Ancaman terhadap budaya seringkali menemui perlawanan. Bentuk lain dari dukungan manajemen mungkin diperlukan untuk mencapai efektivitas dan keberhasilan dengan analisis akar masalah. Misalnya, kebijakan "non-hukuman" terhadap pengidentifikasi masalah mungkin diperlukan.

Proses Umum untuk Melakukan dan Mendokumentasikan Korektif berbasis RCA

RCA (dalam langkah 3, 4 dan 5) membentuk bagian paling kritis dari tindakan korektif yang berhasil, mengarahkan tindakan korektif pada akar penyebab masalah sebenarnya. Mengetahui akar penyebab adalah sekunder dari tujuan pencegahan, karena tidak mungkin untuk menentukan tindakan korektif yang benar-benar efektif untuk masalah yang ditentukan tanpa mengetahui akar penyebabnya.

1. Mendefinisikan masalah atau gambarkan peristiwa untuk mencegah di masa depan. Sertakan atribut kualitatif dan kuantitatif (properti) dari hasil yang tidak diinginkan. Biasanya ini termasuk menentukan sifat, besarnya, lokasi, dan waktu kejadian. Dalam

beberapa kasus, “menurunkan risiko reoccurrences” mungkin menjadi target yang masuk akal. Misalnya, “mengurangi risiko” kecelakaan mobil di masa depan tentu saja merupakan tujuan yang lebih ekonomis daripada “mencegah semua” kecelakaan mobil di masa depan.

2. Mengumpulkan data dan bukti, mengklasifikasikannya sepanjang garis waktu kejadian hingga kegagalan atau krisis akhir. Untuk setiap perilaku, kondisi, tindakan, dan tidak adanya tindakan, tentukan dalam “garis waktu” apa yang seharusnya dilakukan ketika berbeda dari apa yang dilakukan.
3. Dalam data mining model Hierarchical Clustering, gunakan kelompok-kelompok pengelompokan alih-alih mengklasifikasikan: (a) puncak kelompok yang menunjukkan penyebab spesifik; (B) menemukan kelompok atas mereka; (c) menemukan karakteristik kelompok yang konsisten; (d) tanyakan kepada ahli dan validasi.
4. Menanyakan “mengapa” dan identifikasi penyebab yang terkait dengan setiap langkah berurutan terhadap masalah atau peristiwa yang ditentukan. “Mengapa” diartikan sebagai “Apa faktor yang secara langsung menghasilkan efeknya?”
5. Mengklasifikasikan penyebab ke dalam dua kategori: faktor-faktor penyebab yang berhubungan dengan suatu peristiwa dalam urutan tersebut; dan akar penyebab yang mengganggu langkah rantai urutan ketika dihilangkan.
6. Mengidentifikasi semua faktor berbahaya lainnya yang memiliki klaim yang sama atau lebih baik disebut “penyebab utama.” Jika ada beberapa penyebab root, yang sering terjadi, ungkapkan dengan jelas untuk pemilihan optimal nanti.
7. Identifikasi tindakan korektif yang akan, dengan pasti, mencegah terulangnya setiap efek berbahaya dan hasil atau faktor terkait. Periksa bahwa setiap tindakan korektif akan, jika dilakukan sebelum peristiwa, mengurangi atau mencegah efek berbahaya tertentu.
8. Identifikasi solusi yang, ketika efektif dan dengan persetujuan kelompok: cegah pengulangan dengan kepastian yang wajar; berada dalam kendali institusi; memenuhi tujuan dan sasarannya; dan jangan menyebabkan atau memperkenalkan masalah baru yang tidak terduga lainnya.
9. Menerapkan koreksi penyebab akar yang direkomendasikan.
10. Memastikan efektivitas dengan mengamati solusi yang diterapkan dalam operasi.
11. Mengidentifikasi metodologi lain yang mungkin berguna untuk pemecahan masalah dan penghindaran masalah.
12. Mengidentifikasi dan atasi contoh lain dari setiap hasil yang berbahaya dan faktor berbahaya.

Daftar Pustaka

"A National Strategy to Eliminate Child Abuse and Neglect Fatalities" (PDF). Commission to Eliminate Child Abuse and Neglect Fatalities. (2016). Retrieved April 4, 2016.

Arikunto, Suharsimi. Prosedur penelitian suatu pendekatan praktik. Edisi revisi VI. 2006. Jakarta : Rineka Cipta

Bates, John, John Bates of Progress explains how complex event processing works and how it can simplify the use of algorithms for finding and capturing trading opportunities, Fix Global Trading, retrieved May 14, 2012

Coker, Frank (2014). Pulse: Understanding the Vital Signs of Your Business (1st ed.). Bellevue, WA: Ambient Light Publishing. pp. 30, 39, 42,more. ISBN 978-0-9893086-0-1.

"Eckerd Rapid Safety Feedback® Highlighted in National Report of Commission to Eliminate Child Abuse and Neglect Fatalities". Eckerd Kids. Retrieved 2016-04-04.

Finlay, Steven (2014). Predictive Analytics, Data Mining and Big Data. Myths, Misconceptions and Methods (1st ed.). Basingstoke: Palgrave Macmillan. p. 237. ISBN 1137379278.

Hooland, Seth Van; Verborgh, Ruben (2014). Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata. London: Facet.

Kimball, Ralph (2008). The Data Warehouse Lifecycle Toolkit (Second ed.). New York: Wiley. pp. 10, 115–117, 131–132, 140, 154–155. ISBN 978-0-470-14977-5.

Library of Congress Network Development and MARC Standards Office (2005-09-08). "Library of Congress Washington DC on metadata". Loc.gov. Retrieved 2011-12-23

Lindert, Bryan (October 2014). "Eckerd Rapid Safety Feedback Bringing Business Intelligence to Child Wel- fare" (PDF). Policy & Practice. Retrieved March 3, 2016.

Luckham, David C. (2012). Event Processing for Business: Organizing the Real-Time Enterprise Hoboken, New Jersey: John Wiley & Sons, Inc.,. p. 3. ISBN 978-0-470-53485-4.

National Information Standards Organization; Rebecca Guenther; Jaqueline Radebaugh (2004). Understanding Metadata (PDF). Bethesda, MD: NISO Press. ISBN 1-880124-62-9. Retrieved 2 April 2014.

"New Strategies Long Overdue on Measuring Child Welfare Risk - The Chronicle of Social Change". The Chron- icle of Social Change. Retrieved 2016-04-04.

Nigrini, Mark (June 2011). "Forensic Analytics: Methods and Techniques for Forensic Accounting Investiga- tions". Hoboken, NJ: John Wiley & Sons Inc. ISBN 978-0-470-89046-2.

Patrick Setiawan, dkk. (2013). Business Intelligence pada PT. Sinarmas Asset Management. Jakarta: Universitas Bina Nusantara.

Ponniah, P. (2001). Data Warehouse Fundamental for IT Proffesionals (second edition ed.). New York: Jhon Willy and Sons, Inc.

"Predictive Big Data Analytics: A Study of Parkinson's Disease using Large, Complex, Heterogeneous, Incon- gruent, Multi-source and Incomplete Observations". PLoS ONE. Retrieved 2016-08-10.

Rainardi, V. (2010). Building a Data Warehouse With Examples in SQL Server. New York: Apress.

Ralph Kimbal dan Joe Caserta. (2004). The Data Warehouse ETL Toolkit. Indianapolis: Wiley Publishing Inc.

Restia Rezalini, dkk. (2010). Perancangan dan Pembuatan Data Warehouse untuk Kebutuhan Sistem Pendukung Keputusan di Bidang Akademik pada Jurusan Sistem Informasi, ITS, Surabaya. SISFO-Jurnal Sistem Informasi.

Retrieved from msi.binus.ac.id: Diakses tanggal 17 Maret 2017 pada msi.binus.ac.id/files/2013/05/0401-09-SI-Suparto-Business-IntelligenceKonsep.pdf

Siegel, Eric (2013). Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die (1st ed.). Wiley. ISBN 978-1-1183-5685-2.
Suprapto Darudiato, dkk. (2013). Business Intelligence : Konsep dan Metoda.
Taiichi Ohno (1988). Toyota Production System: Beyond Large-Scale Production. Portland, Oregon: Productivity Press. p.17. ISBN 0-915299-14-3.
"VRA Core Support Pages". Visual Resource Association Foundation. Visual Resource Association Foundation. Retrieved 27 February 2016.
Wilson, Paul F.; Dell, Larry D.; Anderson, Gaylord F. (1993). Root Cause Analysis: A Tool for Total Quality Management. Milwaukee, Wisconsin: ASQ Quality Press. pp. 8–17. ISBN 0-87389-163-5.

PENGANTAR INTELEGENSI BISNIS

Dr. Joseph Teguh Santoso, S.Kom., M.Kom.

Dr. Joseph Teguh Santoso, S.Kom., M.Kom adalah Rektor dari Universitas Sains & Teknologi Komputer (Universitas STEKOM) Semarang yang memiliki banyak pengalaman praktisi dalam bidang *e-commerce* sejak tahun 2002. Beliau mempunyai 3 (tiga) toko *Official Online Store* di China untuk merek sepeda Raleigh, dengan omzet tahunan pada 2019 mencapai lebih dari Rp. 35 Milyar rupiah dan terus meningkat. Dr. Joseph T.S memiliki lisensi tunggal merk sepeda "Releigh" untuk penjualan *Online* di seluruh China. di samping itu beliau juga memiliki pabrik sepeda dan sepeda listrik merek "Fengjiu", yaitu Pabrik Sepeda Listrik yang masih tergolong kecil di China seperti Alibaba, Tmall, Taobao, JD, Aliexpress sangat membantu mahasiswa untuk memiliki pengalaman teknis dan praktis membuka toko *Online* bersama beliau

Diterbitkan Kerjasama Penerbit Yayasan Prima Agus Teknik
dengan UNIVERSITAS STEKOM

YAYASAN PRIMA AGUS TEKNIK

UNIVERSITAS STEKOM

ISBN 978-623-95042-6-7 (PDF)