

ILMU DATA PEMASARAN

Dr. Agus Wibowo, M.Kom, M.Si, MM.



YAYASAN PRIMA AGUS TEKNIK





ILMU DATA PEMASARAN

Dr. Agus Wibowo, M.Kom, M.Si, MM.



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

ISBN 978-634-7695-27-7 (PDF)



9

786347

695277

ILMU DATA PEMASARAN

Penulis :

Dr. Agus Wibowo, M.Kom, M.Si, MM.

ISBN : 978-634-7695-27-7 (PDF)

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 Yuniarto, S.Ds., M.Kom

Penebit :

Yayasan Prima Agus Teknik Bekerja sama dengan
Universitas Sains & Teknologi Komputer (Universitas STEKOM)

Anggota IKAPI No: 279 / ALB / JTE / 2023

Redaksi :

Jl. Majapahit no 605 Semarang

Telp. 08122925000

Fax. 024-6710144

Email : penerbit_ypat@stekom.ac.id

Distributor Tunggal :

Universitas STEKOM

Jl. Majapahit no 605 Semarang

Telp. 08122925000

Fax. 024-6710144

Email : info@stekom.ac.id

Hak cipta dilindungi undang-undang

Dilarang memperbanyak karya tulis ini dalam bentuk dan dengan cara
apapun tanpa ijin dari penulis

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat, hidayah, taufiq, dan karunia-Nya sehingga buku *Mastering Marketing Data Science: Panduan Komprehensif untuk Pemasar Masa Kini* dapat diselesaikan dengan baik sebagai bentuk kontribusi terhadap pengembangan ilmu pemasaran modern di era digital yang semakin sarat dengan data, algoritma, dan kecerdasan buatan.

Perkembangan teknologi informasi telah mengubah cara perusahaan memahami pelanggan, menyusun strategi, mengelola komunikasi, serta mengukur keberhasilan pemasaran. Jika dahulu keputusan pemasaran banyak bergantung pada intuisi dan pengalaman, kini pemasaran menuntut pendekatan yang lebih ilmiah, terukur, dan berbasis bukti. Data telah menjadi sumber daya utama yang memungkinkan pemasar mengenali kebutuhan pelanggan, memprediksi perilaku konsumen, mengoptimalkan kampanye, serta menciptakan keunggulan kompetitif yang berkelanjutan.

Buku ini disusun untuk menjembatani kebutuhan antara teori pemasaran modern dan praktik analisis data yang relevan dengan kebutuhan industri saat ini. Pembahasan di dalamnya dimulai dari dasar-dasar ilmu data pemasaran, pengolahan data, analisis statistik, pembelajaran mesin, pemrosesan bahasa alami, analisis perilaku pelanggan, hingga topik lanjutan seperti big data, analisis real-time, generative AI, serta etika dan privasi data. Dengan demikian, buku ini diharapkan dapat menjadi panduan yang utuh bagi mahasiswa, dosen, peneliti, praktisi bisnis, dan siapa pun yang ingin memahami pemasaran berbasis data secara lebih mendalam.

Bab 1 memperkenalkan konsep dasar ilmu data pemasaran dan menjelaskan bagaimana data berperan dalam mendukung keputusan pemasaran. Bab ini juga menguraikan pentingnya pendekatan berbasis data dalam meningkatkan penjualan silang, memperbaiki kampanye pemasaran, serta membantu perusahaan memahami pola perilaku pelanggan secara lebih akurat. Melalui bab ini, pembaca diajak memahami bahwa data bukan sekadar informasi tambahan, melainkan fondasi utama dalam strategi pemasaran modern.

Bab 2 membahas sumber-sumber data tradisional, metode pengumpulan data, serta langkah-langkah persiapan data sebelum dianalisis lebih lanjut. Pada tahap ini, pembaca akan mempelajari pentingnya pembersihan data, transformasi data, dan penyesuaian format agar data siap digunakan untuk analisis yang lebih kompleks. Bab ini juga menampilkan contoh koleksi dan olah data churn pelanggan sebagai ilustrasi praktis penerapan data dalam konteks bisnis nyata.

Bab 3 mengarahkan pembaca untuk memahami data historis dan parameter deskriptif yang sering digunakan dalam pemasaran. Selain itu, bab ini membahas teknik visualisasi data untuk membantu pemasar membaca tren, pola, dan anomali secara lebih jelas. Pembahasan juga mencakup pengukuran keberhasilan kampanye berbasis data serta evaluasi data pemasaran digital, sehingga pembaca dapat memahami bagaimana data dipakai untuk memantau efektivitas strategi pemasaran.

Bab 4 menjelaskan cara mengukur ketidakpastian dalam pemasaran melalui interval kepercayaan, pengujian hipotesis, dan eksperimen A/B. Bab ini sangat penting karena membantu pembaca memahami bagaimana asumsi pemasaran dapat diuji secara ilmiah sebelum diambil menjadi keputusan bisnis. Selain itu, bab ini juga membahas segmentasi pelanggan, pengolahan data pelanggan, dan integrasi data untuk strategi masa depan.

Bab 5 membahas peran analisis prediktif dalam pemasaran serta tantangan peramalan dalam konteks bisnis. Di dalamnya dijelaskan berbagai teknik pembelajaran mesin, evaluasi model, dan pemilihan model yang paling sesuai. Bab ini juga menguraikan dampak churn terhadap bisnis serta penggunaan regresi logistik untuk memprediksi pengurangan pelanggan. Dengan demikian, pembaca memperoleh gambaran bagaimana machine learning dapat membantu pemasar mengambil langkah antisipatif.

Bab 6 memperkenalkan pemrosesan bahasa alami atau NLP sebagai alat penting dalam memahami komunikasi pelanggan. Bab ini menjelaskan teknik-teknik NLP yang dapat digunakan dalam pemasaran, mulai dari pengolahan teks hingga analisis sentimen media sosial. Selain itu, pembaca juga diperkenalkan pada penggunaan bot dan asisten virtual yang mampu meningkatkan interaksi pelanggan secara instan dan efisien.

Bab 7 membahas bagaimana perilaku pasar dapat dipahami melalui media sosial dan situs web. Topik yang dibahas meliputi social network analysis, indikator utama kinerja situs web, social listening, serta pelacakan respons audiens. Bab ini juga mengulas optimasi tingkat konversi, sehingga pembaca dapat melihat hubungan antara interaksi digital dengan hasil pemasaran yang nyata.

Bab 8 kedelapan menjelaskan konsep pemodelan bauran pemasaran dan model atribusi berbasis data. Pembahasan ini penting untuk mengetahui saluran pemasaran mana yang paling berkontribusi terhadap hasil akhir. Di dalam bab ini juga dibahas multi-touch attribution, efektivitas strategi pemasaran perusahaan, dan bagaimana pengukuran kontribusi tiap saluran dapat membantu pengambilan keputusan yang lebih akurat.

Bab 9 kesembilan berfokus pada customer journey, yaitu perjalanan pelanggan dari tahap kesadaran hingga pembelian dan loyalitas. Bab ini menguraikan bagaimana memetakan touchpoint pelanggan, mengoptimalkan pemasaran lintas saluran, dan menilai jalur pembelian secara lebih terstruktur. Dengan memahami perjalanan pelanggan, perusahaan dapat menyusun pengalaman yang lebih relevan dan personal bagi setiap segmen pasar.

Bab 10 kesepuluh membahas pentingnya eksperimen dalam pengambilan keputusan pemasaran. Pembaca akan mempelajari desain eksperimen, fractional factorial design, dan multi-armed bandit sebagai metode untuk menguji strategi pemasaran secara efisien. Bab ini membantu pemasar memahami bagaimana eksperimen dapat digunakan untuk menemukan pendekatan terbaik dalam kondisi pasar yang dinamis.

Bab 11 kesebelas menguraikan peran big data dalam pemasaran modern. Topik yang dibahas mencakup komputasi terdistribusi, penggunaan teknologi canggih, serta personalisasi dan pemasaran waktu nyata. Bab ini menunjukkan bahwa kemampuan mengelola data dalam volume besar dan kecepatan tinggi sangat penting untuk menciptakan respons pemasaran yang cepat, tepat, dan relevan.

Bab 12 kedua belas membahas kecerdasan buatan generatif dan bagaimana teknologi ini mulai membentuk masa depan pemasaran. Pembahasan meliputi pemanfaatan generative AI untuk personalisasi konten, prediksi perilaku pelanggan, dan berbagai peluang baru dalam strategi pemasaran digital. Bab ini juga menyoroti pertimbangan etika yang perlu diperhatikan agar teknologi digunakan secara bertanggung jawab.

Bab 13, terakhir menutup pembahasan dengan isu yang sangat penting, yaitu etika, privasi, bias, keadilan, dan transparansi dalam penggunaan data pemasaran. Bab ini menegaskan bahwa kemajuan teknologi harus diimbangi dengan tanggung jawab moral dan perlindungan terhadap hak konsumen. Selain itu, bab ini juga mengulas tren masa depan ilmu data pemasaran yang akan semakin menuntut integrasi antara kemampuan analitik, empati, dan integritas.

Melalui buku ini, penulis berharap pembaca tidak hanya memahami konsep-konsep utama dalam marketing data science, tetapi juga mampu menerapkannya dalam situasi nyata. Buku ini diharapkan menjadi referensi yang bermanfaat bagi dunia akademik maupun industri, serta menjadi bekal bagi para pemasar masa kini untuk menghadapi persaingan yang semakin ketat dan berbasis data.

Semoga buku ini dapat memberikan manfaat sebesar-besarnya bagi pembaca dan menjadi bagian kecil dari upaya membangun praktik pemasaran yang lebih cerdas, adaptif, dan berorientasi pada masa depan.

Semarang, Mei 2026

Penulis

Dr. Agus Wibowo, M.Kom, M.Si, MM.

DAFTAR ISI

KATA PENGANTAR	i
DAFTAR ISI	iv
BAB 1 PENGANTAR ILMU DATA PEMASARAN	1
1.1 Ilmu Data Pemasaran.....	1
1.2 Peran Ilmu Data Dalam Pemasaran	3
1.3 Keputusan Berbasis Data	4
1.4 Pengantar Ilmu Data Pemasaran	6
1.5 Meningkatkan Penjualan Silang Perbankan Dengan Ilmu Data	8
1.6 Dampak Ilmu Data Pada Kampanye Pemasaran.....	9
1.7 Kesimpulan	11
BAB 2 PENGUMPULAN DAN PERSIAPAN DATA	13
2.1 Pendahuluan.....	13
2.2 Sumber Data Tradisional.....	15
2.3 Metode Pengumpulan Data	18
2.4 Persiapan Data.....	21
2.5 Koleksi & Olah Data Churn Pelanggan	33
2.6 Kesimpulan	35
BAB 3 NAVIGASI DATA.....	39
3.1 Pendahuluan.....	39
3.2 Pemahaman Komprehensif Tentang Data Historis	40
3.3 Parameter Deskriptif Metrik Pemasaran	42
3.4 Teknik Visualisasi Data	45
3.6 Parameter Keberhasilan Kampanye Berbasis Data	54
3.7 Evaluasi Data Pemasaran Digital Produk Minuman.....	56
3.8 Kesimpulan	58
BAB 4 MENGUKUR KEPASTIAN DAN MENGUJI HIPOTESIS PEMASARAN	63
4.1 Pendahuluan.....	63
4.2 Proyeksi Populasi Dan Validasi Asumsi Pasar	64
4.3 Pentingnya Ilmu Data Pemasaran.....	69
4.4 Interval Kepercayaan	72
4.5 Pengujian A/B Dalam Pemasaran	75
4.6 Pengujian Hipotesis Dalam Pemasaran	81
4.7 Segmentasi Dan Pengolahan Pelanggan.....	85
4.8 Menentukan Atribut Kunci Pelanggan.....	95
4.9 Integrasi Data Untuk Strategi Masa Depan	98
BAB 5 PEMBAHASAN PREDIKTIF DAN PEMBELAJARAN MESIN	105
5.1 Gambaran Umum Pembahasan Prediktif.....	105
5.2 Tantangan Peramalan Dan Pertimbangan Bisnis	108
5.3 Teknik Pembelajaran Mesin.....	110
5.4 Evaluasi Dan Seleksi Model	118

5.5	Dampak Churn Pada Bisnis	124
5.6	Lanskap Komersial	128
5.7	Memprediksi Pengurangan Pelanggan Dengan Regresi Logistik	132
5.8	Pembelajaran Mesin Dalam Pemasaran.....	138
BAB 6	PEMROSESAN BAHASA ALAMI DALAM PEMASARAN	143
6.1	Pengantar NLP Untuk Pemasar.....	143
6.2	Gambaran Umum Pemrosesan Bahasa Alami	143
6.3	Teknik NLP Dalam Pemasaran	147
6.4	Teknik Pemrosesan Bahasa Alami Utama Untuk Pemasaran	150
6.5	Komunikasi Instan Melalui Bot Dan Asisten Virtual	156
6.6	Pembahasan Sentimen Media Sosial	160
6.6	Kesimpulan	164
BAB 7	MEMBACA PERILAKU PASAR DI MEDIA SOSIAL DAN SITUS WEB	168
7.1	Pendahuluan.....	168
7.2	Jaringan Sosial (SNA)	168
7.3	Indikator Utama Kinerja Situs Web	176
7.4	Mendengarkan Dan Melacak Media Sosial	184
7.5	Optimasi Tingkat Konversi	191
7.6	Kesimpulan	196
BAB 8	PEMODELAN BAURAN PEMASARAN DAN ATRIBUSI	201
8.1	Pendahuluan.....	201
8.2	Konsep Pemodelan Bauran Pemasaran.....	201
8.3	Model Atribusi Berbasis Data	208
8.4	Atribusi Multi-Sentuh (MTA).....	212
8.5	Efektivitas Strategi Pemasaran Perusahaan	217
8.6	Kesimpulan	222
BAB 9	PEMBAHASAN PERJALANAN PELANGGAN	226
9.1	Pendahuluan.....	226
9.2	Pemetaan Perjalanan Pelanggan	226
9.3	Navigasi Touchpoint Pelanggan	230
9.4	Optimasi Pemasaran Lintas Saluran	235
9.5	Optimasi Atribusi Perjalanan Pembelian	241
9.6	Kesimpulan	245
BAB 10	DESAIN EKSPERIMENTAL DALAM PEMASARAN	250
10.1	Pendahuluan.....	250
10.2	Desain Eksperimen	250
10.3	Desain Faktorial Fraksional	254
10.4	Multi-Armed Bandit.....	259
10.5	Desain Eksperimental	263
10.6	Kesimpulan	268
BAB 11	TEKNOLOGI BIG DATA DAN ANALISIS REAL-TIME.....	271
11.1	Pendahuluan.....	271
11.2	Big Data.....	271
11.3	Kerangka Kerja Komputasi Terdistribusi	275

11.4	Penggunaan Teknologi Dan Metode Canggih.....	281
11.5	Personalisasi Dan Pemasaran Waktu Nyata	286
11.6	Kesimpulan	291
BAB 12	KECERDASAN BUATAN GENERATIF DAN APLIKASINYA DALAM PEMASARAN..	293
12.1	Pendahuluan.....	293
12.2	Memahami Kecerdasan Buatan Generatif	294
12.3	Generative AI Untuk Personalisasi Konten	299
12.4	Prediksi Perilaku Pelanggan Lewat Generative AI	301
12.5	Etika & Masa Depan Generative AI Dalam Pemasaran	306
12.6	Kesimpulan	310
BAB 13	ETIKA, PRIVASI, DAN MASA DEPAN ILMU DATA PEMASARAN.....	311
13.1	Pendahuluan.....	311
13.2	Pertimbangan Etika Dalam Ilmu Data Pemasaran	311
13.3	Peraturan Privasi Data	317
13.4	Bias, Keadilan, Dan Transparansi	323
13.5	Tren Yang Muncul Dan Masa Depan Ilmu Data Pemasaran	326
13.6	Kesimpulan	330
DAFTAR PUSTAKA.....		331
LAMPIRAN: KUNCI LATIHAN SOAL.....		335

BAB 1

PENGANTAR ILMU DATA PEMASARAN

1.1 ILMU DATA PEMASARAN

Dalam lanskap modern, ilmu data pemasaran berada di persimpangan yang menarik, secara rumit menjalin metodologi dan instrumen ilmu data yang canggih dengan ranah kebijaksanaan pemasaran yang mendalam. Apa yang menjadi inti dari persimpangan ini? Upaya untuk menggali wawasan yang mendalam, mengkatalisasi pertumbuhan organisasi, dan menyempurnakan cetak biru pemasaran. Karena data terus mengalir dari berbagai sumber—meliputi keterlibatan pelanggan, luasnya media sosial, dan metrik web yang rumit—ada kebutuhan mendesak untuk navigasi dan interpretasi yang cerdas.

Dalam ranah pemasaran, ilmu data memainkan peran penting dalam membuka wawasan berharga dan mendorong pengambilan keputusan strategis. Bidang dinamis ini mencakup berbagai faktor kunci yang secara kolektif berkontribusi pada kekuatan dan efektivitasnya. Faktor-faktor ini meliputi pengumpulan dan persiapan data berkualitas tinggi dari berbagai sumber, penerapan teknik pembelajaran canggih seperti pembelajaran deskriptif, prediktif, dan preskriptif, serta kemampuan untuk mengkomunikasikan temuan dengan cara yang jelas dan dapat ditindaklanjuti.

Lebih lanjut, ilmu data dalam pemasaran membutuhkan pemahaman tentang perilaku konsumen, tren pasar, dan lanskap persaingan, serta kemampuan untuk memanfaatkan pengetahuan ini untuk menginformasikan dan mengoptimalkan strategi pemasaran. Akibatnya, perpaduan keahlian pemasaran dan kemampuan ilmu data menciptakan kombinasi ampuh yang dapat secara signifikan meningkatkan keunggulan kompetitif perusahaan dan mendorong pertumbuhan bisnis.

Faktor-faktor kunci tersebut meliputi hal-hal berikut, yang akan dibahas secara rinci dalam buku ini:

1. **Pengumpulan data.** Mengumpulkan data yang relevan, yang diekstrak dari berbagai sumber seperti basis data internal, sistem manajemen hubungan pelanggan, lanskap media sosial, instrumen pembelajaran web, dan penyedia pihak ketiga (Bab 2: Pengumpulan dan Persiapan Data).
2. **Persiapan data.** Membersihkan, memproses awal, dan mengubah data mentah menjadi format yang siap untuk dipembahasan. Tahap ini sering kali bergulat dengan tantangan data yang hilang atau tidak sesuai, rekayasa fitur, dan normalisasi atau standardisasi data (Bab 2: Pengumpulan dan Persiapan Data).
3. **Pembahasan data.** Menggunakan teknik pembelajaran deskriptif, inferensial, dan prediktif untuk meneliti data, mengungkap wawasan, pola, dan tren yang dapat memandu strategi pemasaran dan proses pengambilan keputusan (Bab 3: Pembelajaran Deskriptif dalam Pemasaran dan Bab 4: Pembelajaran Inferensial dan Pengujian Hipotesis).

4. **Pengembangan model.** Merancang, memeriksa, dan memvalidasi model pembelajaran mesin, yang mencakup algoritma klasifikasi, regresi, atau pengelompokan, dengan tujuan untuk memprediksi perilaku pelanggan, melakukan segmentasi pelanggan, atau mengoptimalkan upaya pemasaran (Bab 5: Pembelajaran Prediktif dan Pembelajaran Mesin).
5. **Visualisasi dan komunikasi.** Menyampaikan temuan dan wawasan yang diperoleh dari pembahasan data dan model melalui visualisasi, laporan, dan presentasi yang jelas dan menarik, yang dirancang dengan cermat untuk berbagai pemangku kepentingan, baik itu eksekutif pemasaran, manajer produk, atau ilmuwan data (Bab 3: Pembahasan Deskriptif dalam Pemasaran).
6. **Implementasi dan optimasi.** Menggabungkan wawasan dan model ke dalam strategi, kampanye, dan proses pemasaran untuk mendorong pertumbuhan bisnis dan meningkatkan kinerja pemasaran. Pada fase ini, siklus berkelanjutan pemantauan, evaluasi, dan penyempurnaan model dan strategi berlangsung, responsif terhadap umpan balik, hasil, dan pasar yang terus berkembang (di seluruh bab).

Dalam perjalanan menerapkan ilmu data untuk masalah pemasaran, para praktisi menghadapi berbagai tantangan pada berbagai tahap, mulai dari pengumpulan data hingga implementasi. Tabel 1.1 menguraikan tantangan-tantangan ini dan mengusulkan solusi dan pendekatan umum, menyajikannya bukan sebagai langkah-langkah berurutan, tetapi sebagai aspek-aspek yang saling terkait dari proses ilmu data.

Ilmu data pemasaran membekali organisasi dengan kekuatan untuk membuat keputusan berbasis data, mengoptimalkan pengeluaran pemasaran, meningkatkan pengalaman pelanggan, dan mengamankan keunggulan kompetitif. Dengan memanfaatkan teknik-teknik canggih, seperti pembelajaran mesin (lihat Bab 5), pemrosesan bahasa alami (NLP) (lihat Bab 6), dan pembelajaran big data (lihat Bab 11), ilmuwan data pemasaran dapat menemukan peluang laten, meramalkan perilaku pelanggan, dan merancang strategi pemasaran yang dipersonalisasi yang sesuai dengan target audiens.

Tabel 1.1 Tantangan dan Solusi dalam Proses Ilmu Data.

Tahap	Tantangan	Solusi dan Pendekatan Umum
Pengumpulan data	• Sumber data yang terfragmentasi • Ketidakkonsistenan dalam data • Data tidak terstruktur	• Alat dan platform integrasi • Pemeriksaan validasi data • Pengikis dan pengurai web
Persiapan data	• Data hilang • Data bising • Catatan duplikat	• Teknik imputasi • Penyaringan dan pembersihan data • Metode penghapusan duplikasi
Pembahasan data	• Asumsi yang salah • Overfitting atau underfitting • Fitur yang tidak relevan	• Pengujian hipotesis • Validasi silang • Pemilihan dan ekstraksi fitur
Pengembangan model	• Memilih tipe model yang salah	• Pembandingan model

	• Tantangan validasi model • Masalah skalabilitas	• Validasi K-fold • Solusi komputasi awan dan terdistribusi
Visualisasi dan komunikasi	• Visual yang menyesatkan • Kompleksitas yang berlebihan • Hilangnya nuansa dalam penyederhanaan	• Penggunaan pedoman visualisasi standar • Desain iteratif • Anotasi dan konteks
Implementasi dan optimasi	• Kesulitan dalam penerapan waktu nyata • Tantangan siklus umpan balik • Integrasi dengan sistem yang ada	• Solusi data streaming • Alat pemantauan berkelanjutan • Middleware dan API

1.2 PERAN ILMU DATA DALAM PEMASARAN

Dunia ilmu data telah berkembang pesat sebagai katalisator ekspansi dan inovasi yang sangat diperlukan dalam lanskap pemasaran. Di tengah evolusi teknologi dan labirin perilaku pelanggan yang rumit, pemasar harus memanfaatkan wawasan berbasis data untuk mengungguli persaingan. Di sini, kita akan mengeksplorasi peran penting ilmu data dalam pemasaran:

- **Wawasan dan preferensi pelanggan.** Mengpembahasan data pelanggan, yang mencakup riwayat pembelian, detail demografis, dan perilaku online, memberdayakan ilmuwan data untuk membedakan tren, selera, dan pola, yang kemudian menginformasikan strategi pemasaran yang disesuaikan untuk memenuhi kebutuhan pelanggan.
- **Segmentasi dan pembuatan profil pelanggan.** Dengan menggunakan algoritma pengelompokan dan teknik pembelajaran mesin lainnya, ilmuwan data mengukir segmen pelanggan yang bermakna berdasarkan karakteristik yang sama, memfasilitasi kampanye yang ditargetkan, pesan yang dipersonalisasi, dan penawaran khusus yang meningkatkan keterlibatan dan tingkat konversi.
- **Optimalisasi pengeluaran pemasaran.** Metodologi ilmu data mengungkap efektivitas berbagai saluran pemasaran, kampanye, dan taktik. Dengan mengidentifikasi aktivitas pemasaran yang berdampak, organisasi mengoptimalkan pengeluaran pemasaran dan mengalokasikan sumber daya dengan lebih bijak.
- **Efektivitas kampanye dan pengujian A/B.** Efektivitas kampanye mengacu pada ukuran seberapa sukses kampanye pemasaran mencapai tujuannya, yang sering dievaluasi melalui indikator kinerja utama (KPI) seperti tingkat konversi atau pengembalian investasi. Salah satu metode utama yang digunakan oleh ilmuwan data untuk menilai efektivitas kampanye adalah pengujian A/B. Pengujian A/B, juga dikenal sebagai pengujian terpisah, melibatkan perbandingan dua versi variabel pemasaran (misalnya, materi iklan, baris subjek email, desain halaman arahan) untuk menentukan mana yang berkinerja lebih baik. Melalui eksperimen tersebut, ilmuwan data dapat mengpembahasan efektivitas berbagai strategi pemasaran, memungkinkan pemasar untuk terus menyempurnakan kampanye mereka dan membuat keputusan

berdasarkan data. Pendekatan ini sangat penting dalam lanskap pemasaran berbasis data saat ini.

- **Pembahasan sentimen dan pemantauan media sosial.** Teknik NLP meng pembahasan sentimen pelanggan, umpan balik, dan percakapan daring seputar merek atau produk. Hal ini membekali organisasi untuk memahami persepsi pelanggan, mengidentifikasi potensi masalah, dan menemukan peluang untuk perbaikan atau inovasi.
- **Sistem rekomendasi dan personalisasi.** Ilmuwan data dapat mengembangkan algoritma yang merekomendasikan produk atau konten berdasarkan preferensi pelanggan, riwayat penelusuran, dan data perilaku lainnya. Hal ini meningkatkan keterlibatan pelanggan, memperkuat penjualan, dan meningkatkan pengalaman pelanggan secara keseluruhan.
- **Peramalan dan perencanaan permintaan.** Dengan memanfaatkan pembahasan deret waktu dan teknik pemodelan prediktif, ilmuwan data dapat meramalkan penjualan, permintaan pelanggan, dan metrik pemasaran penting lainnya, memberdayakan organisasi untuk merencanakan strategi pemasaran, manajemen inventaris, dan alokasi sumber daya secara efektif.
- **Prediksi churn dan retensi pelanggan.** Dengan meng pembahasan perilaku pelanggan dan mengidentifikasi faktor-faktor yang berkontribusi terhadap churn, ilmuwan data dapat membuat model yang memprediksi risiko kehilangan pelanggan. Hal ini memungkinkan organisasi untuk secara proaktif mempertahankan pelanggan berharga dan meningkatkan kepuasan pelanggan secara keseluruhan.
- **Pemodelan dan atribusi bauran pemasaran.** Ilmuwan data mengukur pengaruh berbagai variabel pemasaran terhadap penjualan atau tujuan pemasaran lainnya dan mengaitkan keberhasilan pemasaran dengan saluran atau taktik tertentu. Hal ini membimbing organisasi dalam membuat keputusan yang tepat tentang bauran pemasaran mereka dan mengoptimalkan strategi untuk dampak maksimal.

Singkatnya, ilmu data telah menjadi aspek penting dalam pemasaran, membantu organisasi dalam memahami pelanggan, mengoptimalkan pendekatan pemasaran, dan mendorong pertumbuhan bisnis. Seiring dengan terus bertambahnya volume, variasi, dan kecepatan data, peran ilmu data dalam pemasaran akan semakin penting dan meluas.

1.3 KEPUTUSAN BERBASIS DATA

Di tengah peran utama pembelajaran pemasaran dan ilmu data dalam mengarahkan organisasi menuju keputusan berbasis data, fungsi-fungsi ini berbeda dalam cakupan, teknik, tujuan, keahlian, dan integrasi dengan strategi pemasaran. Pada bagian ini, kita akan membahas perbedaan-perbedaan ini secara lebih rinci.

Meskipun pembelajaran pemasaran terutama berfokus pada pengukuran, pembahasan, dan pelaporan data pemasaran untuk memahami efektivitas upaya pemasaran, mengidentifikasi tren, dan menginformasikan keputusan pemasaran, ilmu data mencakup serangkaian teknik dan metodologi yang lebih luas yang melampaui pembelajaran tradisional untuk menciptakan wawasan, prediksi, dan rekomendasi yang mendalam.

- **Pembelajaran pemasaran.** Misalnya, sebuah merek olahraga besar mungkin menggunakan pembelajaran pemasaran untuk mengukur efektivitas kampanye iklan Super Bowl-nya dengan memantau metrik seperti jumlah penayangan, rasio klik-tayang, dan penjualan langsung yang dihasilkan dari iklan tersebut.
- **Ilmu data.** Merek olahraga yang sama mungkin menggunakan ilmu data untuk meng pembahasan perilaku pembelian pelanggan, sentimen media sosial, dan sumber data kompleks lainnya untuk memprediksi jenis produk mana yang akan menjadi hit besar berikutnya atau dukungan selebriti mana yang mungkin paling sesuai dengan target audiens mereka.

Pembahasan pemasaran biasanya bergantung pada statistik deskriptif dan inferensial untuk meng pembahasan data dan menarik kesimpulan, menggunakan teknik seperti visualisasi data, statistik ringkasan, pengujian hipotesis, dan pembahasan regresi. Namun, ilmu data menggali teknik yang lebih canggih seperti pembelajaran mesin, NLP, dan pembahasan jaringan untuk mengungkap pola tersembunyi, membuat prediksi, dan merancang solusi berbasis data untuk teka-teki pemasaran yang kompleks.

- **Pembahasan pemasaran.** Pengecer e-commerce mungkin menggunakan statistik deskriptif untuk memahami produk mana yang paling laris, berdasarkan data historis, dan memvisualisasikan tren ini menggunakan grafik batang atau peta panas.
- **Ilmu data.** Untuk pengecer e-commerce yang sama, ilmu data dapat digunakan untuk mengembangkan sistem rekomendasi menggunakan pembelajaran mesin. Sistem ini dapat memprediksi dan menampilkan produk yang mungkin diminati pelanggan berdasarkan riwayat penelusuran mereka, sehingga secara signifikan meningkatkan peluang upsell dan cross-sell.

Pembahasan pemasaran terutama bertujuan untuk memahami dan mengevaluasi kinerja pemasaran masa lalu dan saat ini, yang melibatkan pengukuran KPI, identifikasi tren, dan evaluasi pengembalian investasi (ROI) kampanye pemasaran. Ilmu data, sebaliknya, bertujuan untuk memahami kinerja masa lalu sekaligus memprediksi hasil di masa depan dan mengoptimalkan strategi pemasaran. Hal ini dapat melibatkan pembuatan model untuk memprediksi perilaku pelanggan, segmentasi pelanggan berdasarkan preferensi, dan pembuatan sistem rekomendasi untuk upaya pemasaran yang dipersonalisasi.

- **Pembahasan pemasaran.** Sebuah jaringan kedai kopi dapat mengevaluasi kinerja program loyalitas baru dengan meng pembahasan frekuensi pelanggan tetap dan rata-rata penjualan per kunjungan setelah program tersebut diperkenalkan.
- **Ilmu data.** Jaringan kedai kopi yang sama dapat menggunakan ilmu data untuk memprediksi permintaan stok untuk minuman tertentu, memprediksi waktu puncak berdasarkan data historis dan pola cuaca, atau mengelompokkan pelanggan ke dalam kluster untuk menyesuaikan penawaran pemasaran dengan preferensi individu.

Analisis pemasaran seringkali memiliki latar belakang di bidang pemasaran, bisnis, atau ekonomi dan memiliki keterampilan analitis dan kuantitatif yang kuat. Mereka mahir dalam pembahasan statistik, visualisasi data, dan alat pelaporan, seperti Excel, Tableau, dan Google Analytics. Sebaliknya, ilmuwan data umumnya memiliki latar belakang di bidang ilmu

komputer, statistik, atau bidang terkait dan mahir dalam bahasa pemrograman (misalnya, SAS, Python), pustaka pembelajaran mesin (misalnya, scikit-learn, TensorFlow), dan platform big data (misalnya, Hadoop, Spark).

- **Pembahasan pemasaran.** Sebuah merek perawatan kulit mungkin mempekerjakan seorang analis pemasaran dengan latar belakang pembelajaran bisnis untuk menafsirkan data penjualan, memahami produk mana yang berkinerja baik di wilayah tertentu, dan mengidentifikasi tren pasar menggunakan alat seperti Excel dan Tableau.
- **Ilmu data.** Merek perawatan kulit juga dapat mempekerjakan seorang ilmuwan data dengan latar belakang pembelajaran mesin untuk membuat model yang memprediksi produk baru mana yang akan menjadi terlaris berdasarkan tren bahan, ulasan pelanggan, dan kumpulan data terkait lainnya.

Pembahasan pemasaran sering kali memberikan informasi bagi strategi pemasaran dengan menawarkan wawasan tentang preferensi pelanggan, kinerja kampanye, dan tren pasar. Ilmu data melampaui sekadar penyediaan wawasan, secara aktif terlibat dalam pengembangan dan optimalisasi strategi pemasaran. Ilmuwan data sering berkolaborasi dengan tim pemasaran untuk merancang eksperimen, mengembangkan model prediktif, dan mengimplementasikan solusi berbasis data.

- **Pembahasan pemasaran.** Sebuah toko fesyen online dapat mengpembahasan data tentang pakaian terlaris dan menggunakan wawasan ini untuk memandu desain koleksi musim berikutnya, memastikan keselarasan dengan preferensi pelanggan.
- **Ilmu data.** Toko fesyen yang sama dapat menggunakan teknik ilmu data untuk melakukan pengujian A/B pada berbagai tata letak situs web, mengoptimalkan pengalaman pengguna untuk mendorong penjualan.

Mereka juga dapat menggunakan model prediktif untuk mengidentifikasi pelanggan yang kemungkinan akan berhenti berlangganan, dan kemudian mengirimkan penawaran promosi yang ditargetkan kepada pelanggan tersebut. Terlepas dari perbedaan ini, pembahasan pemasaran dan ilmu data berfungsi sebagai disiplin ilmu yang saling melengkapi yang, secara bersamaan, memungkinkan organisasi untuk mencapai pemahaman komprehensif tentang pelanggan, pasar, dan kinerja pemasaran mereka. Dengan memanfaatkan kekuatan kedua disiplin ilmu tersebut, pemasar dapat membuat keputusan yang lebih tepat, mengoptimalkan strategi, dan mendorong pertumbuhan bisnis.

1.4 PENGANTAR ILMU DATA PEMASARAN

Seiring perkembangan lanskap bisnis, demikian pula kecanggihan dan kompleksitas teknik pemasaran. Kini, lebih dari sebelumnya, pemasaran sangat terkait erat dengan paradigma data, teknologi, dan algoritma yang terus berkembang. Menavigasi labirin pemasaran modern membutuhkan bukan hanya kesadaran, tetapi juga pemahaman mendalam tentang bahasa ilmu data sebagaimana diterapkan dalam pemasaran. Ini bukan hanya tentang menguasai jargon, tetapi lebih tentang memastikan Anda memiliki pengetahuan dasar untuk memanfaatkan kekuatan luar biasa dari strategi pemasaran berbasis data.

Istilah-istilah seperti pembelajaran mesin dan pembelajaran prediktif bukanlah sekadar kata-kata kunci—istilah-istilah tersebut mewakili metodologi transformatif yang telah merevolusi cara bisnis berinteraksi dengan konsumen, membentuk produk, dan merencanakan strategi masa depan mereka. Bagi siapa pun yang memulai perjalanan dalam ilmu data pemasaran, peta jalan dimulai dengan pemahaman yang jelas tentang istilah dan konsep fundamental. Di bagian ini, kami mengidentifikasi beberapa istilah paling penting yang akan Anda temui, yang berfungsi sebagai fondasi untuk perjalanan Anda ke kedalaman ilmu data pemasaran.

Ilmu Data

Ilmu data adalah bidang interdisipliner yang menggabungkan ilmu komputer, statistik, dan keahlian domain untuk menyaring pengetahuan dan wawasan dari data terstruktur dan tidak terstruktur. Teknik ilmu data membantu mengidentifikasi pola, tren, dan hubungan yang menginformasikan pengambilan keputusan dan mendukung pertumbuhan bisnis. Topik ini akan dieksplorasi secara detail di seluruh teks ini.

Visualisasi Data

Visualisasi data mengacu pada penggunaan representasi grafis, termasuk bagan, grafik, dan peta, untuk menampilkan data dan menyederhanakan pemahaman, eksplorasi, dan pembahasan. Alat visualisasi data seperti Tableau dan Power BI memungkinkan pemasar untuk menyampaikan wawasan, tren, dan pola dengan cara yang menarik secara visual dan mudah dicerna.

Segmentasi Pelanggan

Segmentasi pelanggan adalah tindakan memisahkan pelanggan ke dalam kelompok atau segmen berdasarkan karakteristik yang sama, seperti demografi, perilaku, atau preferensi. Segmentasi pelanggan memungkinkan organisasi untuk membuat kampanye pemasaran yang ditargetkan, pesan yang dipersonalisasi, dan penawaran yang disesuaikan agar sesuai dengan setiap kelompok pelanggan.

Pembahasan Prediktif

Pembahasan prediktif adalah penggunaan teknik statistik dan pembelajaran mesin untuk mengpembahasan data historis dan memprediksi peristiwa atau tren di masa mendatang. Pembahasan prediktif membantu organisasi dalam mengantisipasi perilaku pelanggan, mengoptimalkan strategi pemasaran, dan mengidentifikasi peluang atau risiko potensial.

Pembelajaran Mesin

Pembelajaran mesin adalah bagian dari ilmu data dan kecerdasan buatan (AI) yang menggunakan algoritma untuk belajar dari data, membedakan pola, dan membuat prediksi atau keputusan. Pembelajaran mesin mencakup pembelajaran terawasi (misalnya, regresi, klasifikasi), pembelajaran tak terawasi (misalnya, pengelompokan, pengurangan dimensi), dan pembelajaran penguatan.

Pemrosesan Bahasa Alami

Cabang AI yang membahas interaksi antara komputer dan bahasa manusia, memberdayakan komputer untuk memahami, menafsirkan, dan menghasilkan bahasa

manusia. Teknik NLP melayani berbagai aplikasi pemasaran, seperti pembahasan sentimen, pemodelan topik, dan pengembangan chatbot.

Pemodelan Bauran Pemasaran

Pemodelan bauran pemasaran adalah teknik yang mengukur dampak variabel pemasaran yang berbeda (misalnya, harga, promosi, produk, tempat) terhadap penjualan atau tujuan pemasaran lainnya. Pemodelan bauran pemasaran membantu organisasi dalam menilai efektivitas upaya pemasaran mereka, mengalokasikan sumber daya secara efisien, dan mengoptimalkan strategi pemasaran untuk dampak maksimal.

Big Data

Kumpulan data besar dan rumit yang sulit dikelola secara efisien oleh teknik pemrosesan data tradisional. Big data sering dicirikan oleh volume (jumlah data), variasi (jenis data), dan kecepatan (kecepatan pembuatan dan pemrosesan data). Teknologi big data, seperti Hadoop dan Spark, memfasilitasi pemrosesan dan pembahasan data dalam jumlah besar secara real-time.

Membangun pemahaman yang kuat tentang konsep dan terminologi kunci ini akan mempersiapkan Anda lebih baik untuk mempelajari berbagai teknik dan metodologi yang digunakan dalam ilmu data pemasaran dan aplikasi praktisnya dalam skenario pemasaran dunia nyata.

1.5 MENINGKATKAN PENJUALAN SILANG PERBANKAN DENGAN ILMU DATA

Sebuah bank ritel, yang untuk tujuan ilustrasi akan kita sebut NexaTrust Bank, ingin meningkatkan upaya penjualan silangnya dengan menawarkan produk keuangan yang ditargetkan kepada pelanggan yang sudah ada. Departemen pemasaran memutuskan untuk menggunakan teknik ilmu data untuk meningkatkan pendekatan mereka, dengan tujuan meningkatkan kepuasan pelanggan dan meningkatkan pendapatan.

NexaTrust Bank mengumpulkan data relevan dari berbagai sumber, termasuk demografi pelanggan, jenis rekening, riwayat transaksi, skor kredit, dan interaksi layanan pelanggan.

Data mentah dibersihkan, diproses terlebih dahulu, dan diubah menjadi format yang sesuai untuk pembahasan. Langkah ini melibatkan penanganan data yang hilang atau tidak konsisten, rekayasa fitur, dan normalisasi atau standardisasi data.

Menggunakan algoritma pengelompokan, NexaTrust Bank melakukan segmentasi pelanggannya berdasarkan perilaku keuangan, penggunaan produk, dan informasi demografis. Hal ini menghasilkan segmen pelanggan yang berbeda, seperti profesional muda, keluarga, individu dengan kekayaan bersih tinggi, dan pensiunan.

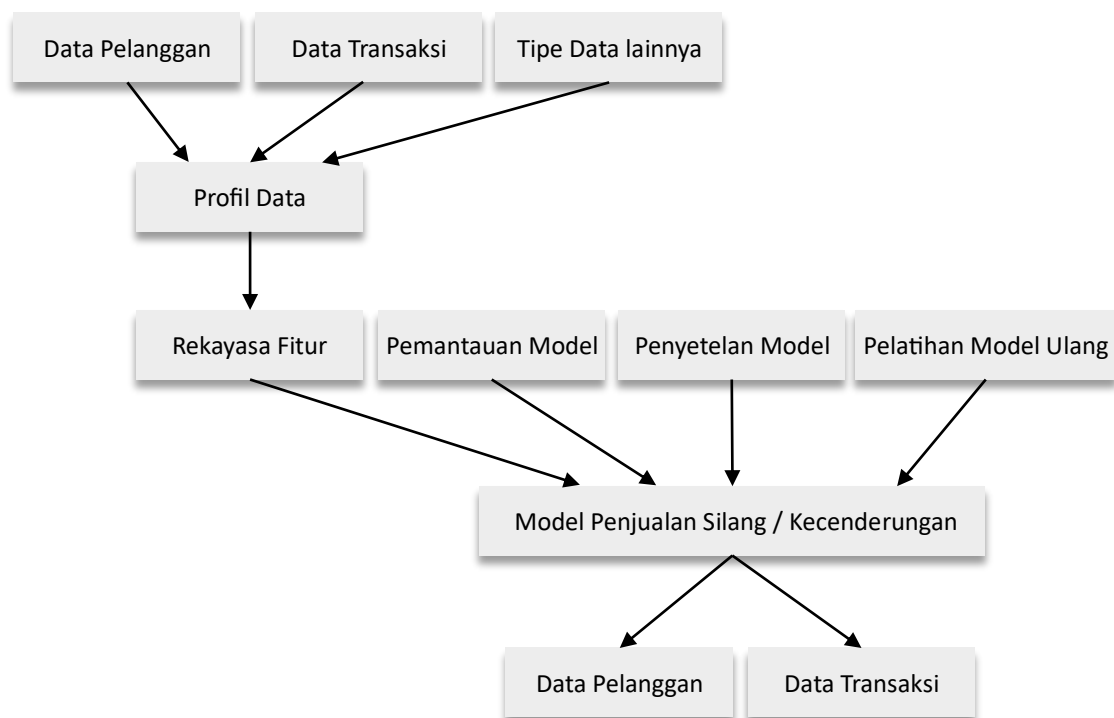
NexaTrust Bank mengembangkan sistem pemberi rekomendasi menggunakan algoritma pembelajaran mesin, seperti penyaringan kolaboratif atau penyaringan berbasis konten, untuk mengidentifikasi produk keuangan yang paling relevan untuk setiap segmen pelanggan berdasarkan preferensi, kebutuhan, dan perilaku keuangan mereka.

NexaTrust Bank mendesain kampanye pemasaran yang ditargetkan untuk setiap segmen pelanggan, dengan fokus pada produk keuangan yang direkomendasikan. Kampanye

ini mencakup pesan yang dipersonalisasi, penawaran yang disesuaikan, dan saluran komunikasi yang dirancang khusus (misalnya, email, SMS, notifikasi dalam aplikasi).

NexaTrust Bank melakukan pengujian A/B untuk mengevaluasi efektivitas berbagai variabel pemasaran, seperti materi iklan, jenis penawaran, dan saluran komunikasi. Hal ini memungkinkan NexaTrust Bank untuk terus mengoptimalkan kampanyenya berdasarkan wawasan berbasis data.

NexaTrust Bank memantau secara cermat kinerja kampanye penjualan silangnya, melacak metrik seperti tingkat konversi, kepuasan pelanggan, dan pendapatan. Berdasarkan wawasan ini, bank menyempurnakan model rekomendasi produk, segmentasi pelanggan, dan strategi pemasarannya untuk memaksimalkan efektivitas upaya penjualan silangnya (lihat Gambar 1.1). Dengan memanfaatkan teknik ilmu data, NexaTrust Bank dapat menawarkan produk keuangan yang lebih relevan dan personal kepada pelanggannya, meningkatkan kepuasan pelanggan, dan meningkatkan keberhasilan upaya penjualan silang.



Gambar 1.1 Menggunakan Ilmu Data untuk Meningkatkan Penjualan Silang di Departemen Pemasaran NexaTrust Bank.

1.6 DAMPAK ILMU DATA PADA KAMPANYE PEMASARAN

Mari kita pertimbangkan LuxeVogue Retailers, sebuah perusahaan ritel hipotetis yang ingin meningkatkan kinerja upaya pemasaran emailnya. Secara historis, seperti banyak perusahaan lain di sektor ritel, LuxeVogue Retailers menggunakan strategi email yang seragam, mengirimkan promosi yang sama kepada semua pelanggan. Namun, untuk tetap kompetitif dan meningkatkan efektivitas kampanye pemasaran mereka, tim di LuxeVogue Retailers beralih ke ilmu data.

Langkah pertama melibatkan pengumpulan data yang cermat, yang menjadi inti dari strategi berbasis data mereka. Data ini bukan hanya kumpulan angka dan fakta acak; ini adalah jalinan kisah pelanggan yang kaya yang diceritakan melalui profil demografis mereka, perilaku pembelian mereka di masa lalu, dan interaksi mereka dengan email sebelumnya. Dengan banyaknya data berharga yang tersedia, analis pemasaran LuxeVogue menggunakan algoritma pembelajaran mesin yang canggih untuk menyaring informasi ini, mengidentifikasi pola yang mungkin terlewatkan oleh mata manusia.

Gambar 1.2 memberikan gambaran grafis dari proses yang rumit ini. Gambar tersebut mengilustrasikan bagaimana data mengalir dari fase pengumpulan ke mesin pembelajaran mesin, yang kemudian menghasilkan segmen pelanggan yang dapat ditindaklanjuti.

Setelah segmen-segmen ini diidentifikasi, LuxeVogue Retailers memulai perjalanan pemasaran yang dipersonalisasi. Ini bukan hanya tentang menyapa pelanggan dengan nama depan mereka; ini tentang membuat penawaran yang sesuai dengan preferensi unik mereka dan mengirim email yang selaras dengan perilaku spesifik mereka. Misalnya, satu segmen mungkin terdiri dari pelanggan yang telah menunjukkan minat pada produk premium, dan segmen lain lebih sensitif terhadap harga dan responsif terhadap penawaran diskon.

Peritel LuxeVogue tidak berhenti sampai di situ. Dengan kemampuan personalisasi, mereka melangkah lebih jauh dengan pengujian A/B. Mereka membuat berbagai kampanye email dengan baris subjek yang berbeda—beberapa lugas, beberapa menggunakan daya tarik, dan yang lainnya dengan rasa urgensi. Tata letak email disesuaikan, beberapa dengan gambar yang menarik dan yang lainnya berfokus pada teks dan kejelasan. Konten itu sendiri bervariasi untuk melihat gaya bercerita mana yang paling sesuai dengan audiens mereka.

Setiap iterasi kampanye dipantau dengan cermat. Tim mengukur berapa banyak pelanggan yang membuka email (tingkat buka), berapa banyak yang mengklik tautan di dalamnya (tingkat klik-tayang), dan, yang terpenting, berapa banyak yang melakukan tindakan yang diinginkan, seperti melakukan pembelian (tingkat konversi). Proses ini bukan sekali jalan; ini adalah siklus berkelanjutan dari pembuatan hipotesis, pengujian, pembelajaran, dan penyempurnaan.

Melalui proses iteratif pengujian dan pembahasan ini, Peritel LuxeVogue tidak hanya mengirim email; Mereka sedang mengembangkan pemahaman yang lebih dalam tentang basis pelanggan mereka. Mereka mempelajari apa yang menginspirasi pelanggan untuk bertindak, waktu mana dalam sehari mereka paling mungkin berinteraksi dengan email, dan konten apa yang mendorong bukan hanya klik tetapi juga keterlibatan yang bermakna yang berkontribusi pada keuntungan.



Gambar 1.2 Menggunakan Ilmu Data untuk Mengoptimalkan Kampanye Pemasaran Email untuk Peritel LuxeVogue.

Hasilnya? Tim pemasaran yang lebih terinformasi yang dapat menunjukkan hubungan yang jelas antara elemen kampanye tertentu dan respons pelanggan. Kampanye pemasaran email menjadi lebih dari sekadar alat promosi—mereka menjadi percakapan dinamis antara Peritel LuxeVogue dan pelanggan mereka. Pendekatan strategis ini, yang didukung oleh ilmu data, pada akhirnya menghasilkan tingkat keterlibatan yang lebih tinggi, mendorong peningkatan penjualan dan ROI yang kuat untuk upaya pemasaran mereka.

1.7 KESIMPULAN

Pada bab pembuka buku ini, kami telah menyiapkan landasan dengan mengungkap seluk-beluk ilmu data pemasaran, mengklarifikasi esensinya, dan membedakannya dari pembelajaran pemasaran yang lebih konvensional. Melalui dua contoh ilustratif, kami bertujuan untuk menjelaskan manfaat luar biasa yang dapat diperoleh ketika mengintegrasikan pendekatan ilmu data untuk mengatasi dilema pemasaran yang rumit.

Seiring kita melangkah lebih jauh ke bab-bab selanjutnya, fokus kita akan beralih ke eksplorasi yang lebih dalam tentang metodologi, instrumen, dan teknik spesifik yang membentuk tulang punggung ilmu data pemasaran. Perjalanan kita akan mencakup berbagai macam subjek. Kita akan menyelami mekanisme pengumpulan data, prinsip-prinsip dasar berbagai bentuk pembelajaran, seni menafsirkan bahasa manusia melalui mesin, ranah pembahasan data media sosial dan situs web, dan tarian rumit strategi pemasaran, di antara yang lainnya.

Sepanjang buku ini, pembaca akan disajikan dengan contoh-contoh konkret yang disertai dengan penggambaran ilustratif, yang bertujuan untuk menjelaskan aplikasi nyata dari teknik yang dibahas di dunia bisnis nyata. Untuk meningkatkan pemahaman dan relevansi kontekstual, setiap bab akan diselingi dengan skenario dan studi kasus dunia nyata, yang dikurasi dengan cermat untuk menjembatani kesenjangan antara konsep teoritis dan manifestasi praktisnya.

Pada saat Anda membalik halaman terakhir buku ini, penulis berharap Anda akan memiliki seperangkat pengetahuan yang komprehensif, yang memungkinkan Anda untuk secara mahir menggunakan ilmu data pemasaran. Hal ini, pada gilirannya, akan memberdayakan Anda untuk menggali wawasan bisnis penting yang dapat menginformasikan dan memperkaya upaya pemasaran Anda, yang selanjutnya mendorong ekspansi dan kesuksesan bisnis.

BAB 2

PENGUMPULAN DAN PERSIAPAN DATA

2.1 PENDAHULUAN

Perjalanan data dalam ranah pemasaran mengingatkan kita pada evolusi sebuah berlian. Sama seperti berlian yang bermula dari seongkah batu bara, yang dipanaskan dan diberi tekanan hingga menjadi permata yang cemerlang, data pun bermula sebagai sumber daya mentah yang luas dan seringkali kacau. Dengan proses yang tepat, data berubah menjadi harta karun berupa wawasan. Namun, perbandingan ini lebih jauh lagi. Sama seperti nilai sebuah berlian bukan hanya terletak pada penemuannya tetapi juga pada pemotongan dan pemolesannya yang cermat, kekuatan sejati data bukan hanya terletak pada pengumpulannya tetapi juga pada persiapannya yang teliti.

Namun, dalam perlombaan untuk mengikuti perkembangan pasar yang semakin digital, pentingnya proses pengumpulan dan persiapan yang cermat ini seringkali diabaikan. Pemasar ingin segera terjun ke dalam pembahasan data, terkadang mengabaikan langkah-langkah dasar yang memastikan kualitas hasilnya. Tetapi seperti membangun gedung pencakar langit di atas tanah yang goyah, melewati atau mengabaikan fase pengumpulan dan persiapan data dapat mengakibatkan hasil yang tidak stabil dan tidak dapat diandalkan.

Orang mungkin bertanya-tanya, mengapa proses ini sering diabaikan? Alasannya bermacam-macam:

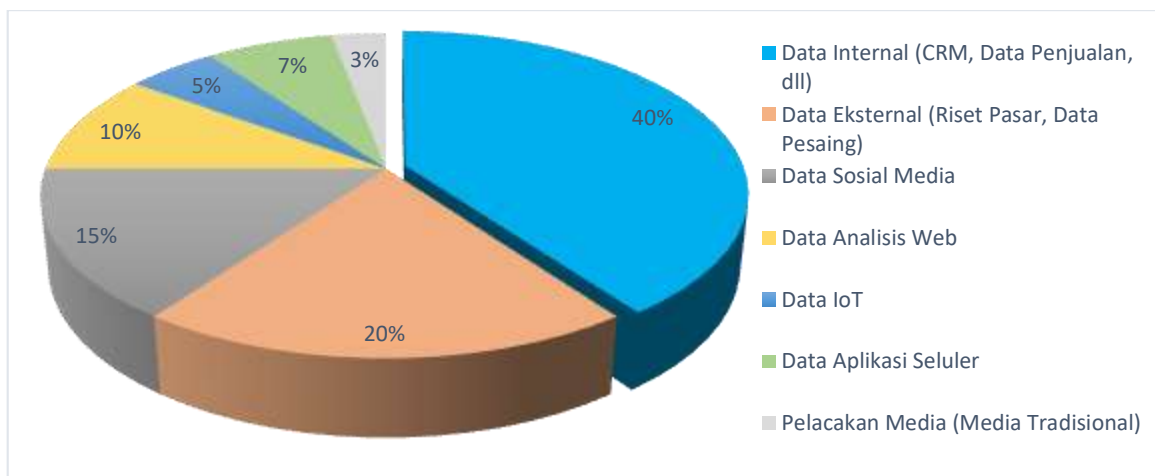
- **Ketersediaan data yang melimpah.** Keberadaan data yang merata mungkin memberikan ilusi bahwa data selalu siap digunakan. Dengan setiap klik, suka, berbagi, pembelian, dan bahkan gulir yang direkam, ada kesalahpahaman bahwa data ini dapat langsung ditindaklanjuti.
- **Fokus yang salah tempat.** Ada daya tarik tertentu pada alat dan algoritma pembelajaran canggih yang menjanjikan wawasan dan hasil instan. Sorotan seringkali lebih terang pada alat-alat ini, meninggalkan proses data fundamental dalam bayangan.
- **Meremehkan kompleksitas.** Pengumpulan dan persiapan data bukan hanya tentang mengumpulkan sejumlah besar data. Ini tentang mendapatkan data yang tepat dan memastikan data tersebut dalam bentuk yang tepat. Ini melibatkan keputusan dan langkah-langkah rumit yang sering diremehkan banyak orang.

Namun, kelalaian ini dapat memiliki dampak yang signifikan. Pengumpulan data yang salah atau persiapan yang tidak memadai dapat menyebabkan bias, ketidakakuratan, dan kesalahpahaman, mengaburkan wawasan dan berpotensi menyebabkan kesalahan pemasaran yang mahal. Oleh karena itu, sangat penting untuk memastikan bahwa data yang tersedia tidak hanya berlimpah tetapi juga dikelola dengan baik dan dipersiapkan dengan tepat untuk pembahasan.

Dalam bab ini, kita akan membahas inti dari proses ilmu data pemasaran—pengumpulan dan persiapan data. Kualitas, akurasi, dan relevansi data merupakan landasan

yang mendasari strategi pemasaran yang kuat dan efektif. Dan meskipun data tersedia melimpah di dunia digital kita, volume dan keragamannya yang besar dapat menjadi keuntungan sekaligus tantangan. Keragaman dan volume data yang tersedia di dunia digital saat ini menawarkan banyak peluang untuk pengambilan keputusan yang tepat dalam pemasaran.

Namun, memahami lanskap sumber data sangat penting untuk pengumpulan dan persiapan data yang efektif. Gambar 2.1 memberikan representasi visual dari perkiraan distribusi sumber data yang digunakan oleh organisasi saat ini. Gambar ini membantu menggambarkan variasi dan prevalensi berbagai jenis data, menyoroti di mana pemasar sering mengumpulkan informasi yang menjadi dasar strategi berbasis data mereka.



Gambar 2.1 Estimasi Distribusi Sumber Data yang Digunakan oleh Organisasi Saat Ini.

Dalam bab ini, kita akan membahas topik-topik berikut:

- **Keragaman sumber data.** Selami lebih dalam berbagai reservoir data yang dapat diakses oleh ilmuwan data pemasaran, termasuk catatan internal organisasi, basis data pelanggan, riset pasar eksternal, umpan media sosial, dan banyak lagi.
- **Teknik pengumpulan data.** Pahami beragam metode yang tersedia untuk pengumpulan data—dari survei tradisional hingga alat yang lebih modern seperti web scraping, antarmuka pemrograman aplikasi (API), dan bahkan pembelian data strategis.
- **Praktik terbaik persiapan data.** Temukan cara mengubah data mentah menjadi permata yang dipoles, siap untuk ekstraksi wawasan. Ini mencakup pembersihan data, integrasi data dari berbagai sumber, transformasi untuk kompatibilitas yang lebih baik, dan pengurangan data untuk fokus pada hal yang benar-benar penting.

Sepanjang bab ini, studi kasus dan contoh dunia nyata akan menjelaskan prinsip-prinsip yang dibahas, memberikan pemahaman teoritis dan praktis tentang setiap konsep. Contoh-contoh ini akan memberikan gambaran tentang bagaimana perusahaan-perusahaan papan atas memanfaatkan kekuatan data yang dikumpulkan dengan baik dan dipersiapkan dengan sempurna untuk mendorong strategi pemasaran mereka.

Saat Anda mempelajari bagian-bagian berikut, ingatlah bahwa wawasan yang diperoleh hanya sebaik data yang menjadi dasarnya. Sama seperti seorang ahli perhiasan yang menekankan pentingnya potongan dan kejernihan berlian, sebagai ilmuwan data pemasaran, fokus kita harus pada pengumpulan dan penyempurnaan aset kita yang paling berharga: data.

2.2 SUMBER DATA TRADISIONAL

Sebelum menyelami pengumpulan data, penting untuk memahami berbagai sumber data yang dapat dimanfaatkan oleh ilmuwan data pemasaran. Evolusi sumber data selama bertahun-tahun telah mengubah lanskap pemasaran, dengan munculnya big data dan teknik pembelajaran canggih yang membuka peluang baru bagi pemasar. Di bagian ini, kita akan membahas sumber data tradisional dan modern serta dampak big data pada pemasaran.

Dalam ranah ilmu data pemasaran, memahami dan memanfaatkan sumber data tradisional sangat penting. Sumber-sumber ini, yang dikategorikan sebagai internal, eksternal, dan pelacakan media, secara historis telah memainkan peran penting dalam membentuk strategi pemasaran. Mari kita jelajahi masing-masing secara detail:

- **Data internal.** Secara historis, organisasi sangat bergantung pada sumber data internal, seperti sistem manajemen hubungan pelanggan (CRM), data penjualan, dan umpan balik pelanggan, untuk menginformasikan strategi pemasaran mereka. Sumber-sumber ini memberikan wawasan berharga tentang perilaku pelanggan, preferensi, dan pola pembelian.
- **Data eksternal.** Sumber data eksternal, seperti laporan riset pasar, publikasi industri, dan informasi pesaing, juga memainkan peran penting dalam membentuk strategi pemasaran. Sumber-sumber ini menawarkan wawasan tentang tren pasar, lanskap persaingan, dan demografi pelanggan.
- **Pelacakan media.** Sebelum munculnya era digital, media tradisional adalah saluran utama untuk komunikasi massa. Ini termasuk publikasi cetak, siaran televisi, dan program radio. Organisasi sangat menghargai pelacakan penyebutan dan liputan di media-media ini. Dengan memantau sumber-sumber ini, perusahaan dapat mengukur kehadiran merek mereka, memahami persepsi publik, dan mengevaluasi dampak kampanye iklan mereka. Hal ini biasanya dilakukan dengan bantuan layanan pemantauan media dan agensi kliping, yang akan menyediakan kompilasi semua penyebutan atau iklan di berbagai platform media tradisional kepada bisnis. Mengpembahasan data ini membantu dalam mengukur efektivitas kampanye, manajemen reputasi, dan memahami jangkauan pesan merek mereka di pasar yang lebih luas. Pelacakan penyebutan media juga berfungsi sebagai indikator awal potensi masalah hubungan masyarakat atau tren yang muncul di industri. Dalam banyak hal, pelacakan media merupakan cikal bakal alat mendengarkan media sosial saat ini, yang menekankan pentingnya pemahaman tentang bagaimana suatu merek atau produk dipersepsikan di dunia yang lebih luas.

Munculnya Sumber Data Modern

Dengan pesatnya pertumbuhan internet, media sosial, dan teknologi seluler, sumber data baru telah muncul, mengubah lanskap data pemasaran (lihat Tabel 2.1):

- **Data media sosial.** Platform media sosial, seperti X (sebelumnya Twitter), Facebook, dan LinkedIn, menghasilkan sejumlah besar konten buatan pengguna, suka, berbagi, dan komentar. Data ini dapat digunakan untuk mengpembahasan sentimen pelanggan, preferensi, dan persepsi merek.
- **Data pembelajaran web.** Alat pembelajaran web, seperti Google Analytics atau Adobe Analytics, melacak perilaku pengguna, tampilan halaman, rasio pentalan, dan metrik kinerja situs web lainnya. Data ini memberikan wawasan tentang keterlibatan pelanggan, pengalaman pengguna, dan efektivitas upaya pemasaran online.
- **Data IoT.** Perangkat yang terhubung ke internet, seperti jam tangan pintar, sensor, dan beacon, menghasilkan data waktu nyata tentang perilaku dan preferensi pelanggan. Data ini dapat digunakan untuk mempersonalisasi upaya pemasaran, mengoptimalkan strategi penetapan harga, dan meningkatkan pengembangan produk.
- **Data aplikasi seluler.** Di era ponsel pintar, aplikasi seluler telah menjadi bagian integral dari kehidupan sehari-hari konsumen. Mulai dari jejaring sosial dan belanja online hingga pelacakan kebugaran dan hiburan, aplikasi memenuhi berbagai kebutuhan pengguna. Dengan lonjakan penggunaan aplikasi seluler ini, data yang dihasilkan dari interaksi pengguna dengan aplikasi telah menjadi tambang emas bagi pemasar. Setiap ketukan, gesekan, dan tindakan pada aplikasi memberikan wawasan tentang preferensi pengguna, pola perilaku, dan tingkat keterlibatan.

Tabel 2.1 Fitur, Keunggulan, dan Keterbatasan Sumber Data Tradisional Versus Modern.

Aspek	Sumber Data Tradisional	Sumber Data Modern
Fitur	Survei Sensus Wawancara tatap muka	Pembahasan web Sistem CRM Pembahasan media sosial
Keunggulan	Teruji dan terbukti Wawasan kualitatif Umpan balik langsung	Wawasan waktu nyata Cakupan luas Data kuantitatif
Keterbatasan	Memakan waktu Cakupan terbatas Tidak real-time	Membutuhkan infrastruktur teknologi Kekhawatiran privasi Kurang memberikan wawasan kualitatif

Big Data dan Dampaknya pada Pemasaran

Big data mengacu pada kumpulan data yang masif dan kompleks yang tidak dapat ditangani secara efisien oleh teknik pemrosesan data tradisional. Hal ini sering kali dicirikan oleh volumenya (jumlah data), variasinya (berbagai jenis data), dan kecepatannya (kecepatan pembuatan dan pemrosesan data). Munculnya big data telah secara signifikan memengaruhi lanskap pemasaran dalam beberapa hal:

- Peningkatan pemahaman pelanggan. Big data memungkinkan organisasi untuk mengumpulkan dan mengpembahasakan beragam sumber data, memberikan pemahaman yang lebih komprehensif dan bernuansa tentang pelanggan. Hal ini membantu pemasar mengembangkan strategi pemasaran yang lebih terarah dan personal.
- Pengambilan keputusan secara real-time. Dengan kemampuan untuk memproses dan mengpembahasakan data secara real-time, pemasar dapat membuat keputusan berbasis data dengan lebih cepat, memungkinkan mereka untuk beradaptasi dan mengoptimalkan strategi mereka sebagai respons terhadap perubahan perilaku pelanggan atau kondisi pasar.
- Teknik pembelajaran tingkat lanjut. Pertumbuhan big data telah menyebabkan pengembangan teknik pembelajaran canggih, seperti pembelajaran mesin, pemrosesan bahasa alami, dan pembelajaran prediktif, yang membantu pemasar mengungkap wawasan yang lebih dalam, membuat prediksi, dan mengoptimalkan strategi mereka untuk dampak maksimal.

Meskipun big data tidak dapat disangkal telah membuka jalan dan kemampuan baru bagi pemasar, big data juga telah menghadirkan beberapa tantangan yang harus diatasi oleh organisasi:

- **Masalah privasi data.** Dengan akumulasi data dalam jumlah besar, terutama data pribadi pengguna, muncul tanggung jawab yang lebih tinggi untuk melindungi data tersebut. Pelanggaran data dapat merugikan konsumen dan perusahaan, menyebabkan kerugian finansial, kerusakan reputasi, dan konsekuensi hukum. Selain itu, peraturan seperti Peraturan Perlindungan Data Umum (GDPR) di Uni Eropa dan Undang-Undang Privasi Konsumen California (CCPA) di AS memberlakukan pedoman ketat tentang pengumpulan, penyimpanan, dan penggunaan data, yang mengharuskan bisnis untuk memastikan bahwa mereka mematuhi peraturan tersebut.
- **Kebutuhan akan keterampilan khusus.** Kompleksitas dan volume big data yang sangat besar berarti bahwa metode dan alat pembahasan data tradisional seringkali tidak memadai. Hal ini telah menciptakan permintaan akan keterampilan khusus, seperti keahlian dalam pembelajaran tingkat lanjut, pembelajaran mesin, dan teknologi big data. Merekrut atau melatih anggota staf dengan kompetensi ini dapat memakan banyak sumber daya.
- **Kekhawatiran akan akurasi dan keandalan data.** Luasnya sumber big data meningkatkan kemungkinan menemukan data yang tidak akurat, usang, atau menyesatkan. Mengandalkan data tersebut dapat menyebabkan wawasan yang salah dan strategi pemasaran yang keliru. Organisasi membutuhkan proses validasi dan pembersihan data yang kuat untuk memastikan integritas data mereka. Lebih lanjut, karena sifat pengumpulan data yang terdesentralisasi dalam lingkungan big data, dapat terjadi inkonsistensi dan redundansi dalam data, yang menimbulkan tantangan dalam mencapai sumber kebenaran tunggal.

Dalam menavigasi lanskap big data, organisasi harus mencapai keseimbangan. Meskipun memanfaatkan kekuatan big data dapat menawarkan wawasan yang tak tertandingi dan keunggulan kompetitif, sangat penting untuk mengatasi tantangan ini secara langsung, memastikan bahwa strategi pemasaran berbasis data efektif dan bertanggung jawab. Perlu dicatat bahwa kita akan membahas lebih dalam seluk-beluk, aplikasi, dan tantangan big data di Bab 11, menawarkan pemahaman komprehensif bagi mereka yang ingin menguasai dampaknya pada pemasaran modern.

Dengan memahami evolusi sumber data dan dampak big data pada pemasaran, Anda akan lebih siap untuk mengidentifikasi sumber data yang paling relevan untuk proyek ilmu data pemasaran Anda dan memanfaatkannya secara efektif untuk mendorong strategi pemasaran berbasis data.

2.3 METODE PENGUMPULAN DATA

Ada berbagai metode untuk mengumpulkan data pemasaran, masing-masing dengan keunggulan dan tantangan uniknya. Memilih metode yang tepat bergantung pada persyaratan data spesifik proyek ilmu data pemasaran Anda, sumber daya yang tersedia, dan tingkat kualitas dan granularitas data yang diinginkan. Di bagian ini, kita akan membahas metode pengumpulan data yang paling umum secara lebih rinci.

Berakar secara historis dalam riset pemasaran, survei dan kuesioner terus menjadi sangat penting. Dilaksanakan melalui berbagai saluran—dari metode tradisional seperti jalur telepon, surat, atau tatap muka hingga alat survei digital modern seperti SurveyMonkey atau Google Forms—alat-alat tersebut menawarkan wawasan terstruktur tentang proses berpikir konsumen. Meskipun hemat biaya dan mudah diskalakan, alat-alat ini berpotensi dirusak oleh bias, seperti pemilihan sendiri dan nonrespons, yang seringkali tidak memberikan gambaran tentang kecenderungan konsumen yang tersirat. Namun, keunggulan survei meliputi biaya yang relatif rendah, skalabilitas, dan kemampuan untuk mengumpulkan data terstruktur.

Web Scraping

Web scraping melibatkan pengambilan data dari situs web dan platform online menggunakan alat dan skrip otomatis. Metode ini dapat berguna untuk mengumpulkan data tentang daftar produk, ulasan pelanggan, harga pesaing, dan informasi publik lainnya. Alat web scraping seperti BeautifulSoup atau Scrapy di Python adalah pilihan populer untuk tujuan ini. Meskipun web scraping dapat menghasilkan volume data yang besar, hal ini mungkin memerlukan keahlian teknis, dan kualitas serta struktur data dapat sangat bervariasi di berbagai situs web. Mungkin ada kekhawatiran etis dan/atau hukum ketika melakukan scraping situs web tertentu tanpa izin, yang juga perlu diperhatikan.

Antarmuka Pemrograman Aplikasi (API)

API adalah cara yang lebih terstruktur dan andal untuk mengakses data dari platform online. Contoh API yang banyak digunakan termasuk API Twitter, yang memungkinkan akses ke tweet, profil pengguna, dan konten publik lainnya; API Google Maps, yang memungkinkan penyematan Google Maps di halaman web dengan lapisan dan penanda yang disesuaikan; dan

API Data YouTube, yang memungkinkan pengembang untuk mengambil konten YouTube untuk diintegrasikan ke dalam aplikasi mereka sendiri.

API ini, antara lain, melayani platform seperti situs media sosial, mesin pencari, atau situs web e-commerce. API memungkinkan akses terprogram ke data, memungkinkan pengembang untuk menanyakan dan mengambil poin data spesifik langsung dari sumbernya. Meskipun API seringkali memiliki batasan penggunaan dan mungkin memerlukan otentikasi, API biasanya memberikan data yang lebih akurat, terkini, dan terstruktur dibandingkan dengan web scraping.

Pembelian Data

Membeli data dari penyedia pihak ketiga, seperti perusahaan riset pasar, broker data, atau penyedia data khusus industri, dapat menawarkan wawasan eksternal yang berharga untuk melengkapi data internal. Penyedia ini seringkali memiliki akses ke kumpulan data besar dan berkualitas tinggi yang mungkin sulit atau memakan waktu untuk dikumpulkan secara independen. Namun, membeli data bisa mahal, dan relevansi serta kualitas data harus dinilai dengan cermat sebelum mengambil keputusan.

Data Pengembangan teknologi masa depan

Dalam beberapa kasus, pengumpulan data mungkin melibatkan pengamatan perilaku pelanggan, interaksi, atau aktivitas lainnya, baik secara langsung maupun melalui saluran digital. Contoh data pengembangan teknologi masa depan meliputi perilaku pelanggan di dalam toko, interaksi pengguna situs web, atau keterlibatan media sosial. Data pengembangan teknologi masa depan dapat memberikan wawasan berharga tentang preferensi dan perilaku pelanggan, tetapi pengumpulannya mungkin memakan waktu dan mungkin memerlukan alat atau keahlian khusus untuk mengembahasannya.

STUDI KASUS: STUDI PENGEMBANGAN TEKNOLOGI MASA DEPAN TOKO APPLE

Untuk lebih memahami pengalaman pelanggan dan meningkatkan desain serta layanan di dalam toko, Apple melakukan studi pengembangan teknologi masa depan di toko-toko ritelnya. Para pengamat mencatat bagaimana pelanggan berinteraksi dengan produk yang dipajang, alur lalu lintas di dalam toko, titik-titik di mana pelanggan tampak bingung atau membutuhkan bantuan, dan area di mana mereka paling banyak berkumpul.

Dari pengamatan ini, Apple mengidentifikasi bahwa pelanggan sering mencari bantuan tepat di tempat mereka berdiri, daripada mendekati konter yang telah ditentukan. Wawasan ini mengarah pada pengenalan "Genius Bar" yang ditempatkan di tengah toko. Lebih lanjut, mereka menyadari bahwa pelanggan menghargai kemampuan untuk bebas menyentuh dan mencoba perangkat, yang memperkuat konsep desain meja terbuka tanpa penutup untuk produk.

Studi pengembangan teknologi masa depan ini secara langsung memengaruhi tata letak toko dan model layanan Apple, mengoptimalkan pengalaman pelanggan dan efisiensi toko.

Dengan memahami kekuatan dan keterbatasan setiap metode pengumpulan data, Anda dapat memilih pendekatan yang paling tepat untuk proyek ilmu data pemasaran Anda, memastikan bahwa Anda memiliki akses ke data yang akurat, relevan, dan berkualitas tinggi untuk mendukung pembahasan dan pengambilan keputusan Anda (lihat Tabel 2.2).

Tabel 2.2 Metode Pengumpulan Data beserta Kelebihan, Kekurangan, Kasus Penggunaan, dan Contoh.

Metode	Keunggulan	Kelemahan	Kasus Penggunaan	Contoh
Wawancara	Wawasan mendalam Fleksibilitas	Memakan waktu Risiko bias	Umpan balik produk Pengalaman pengguna	Wawancara keluar kerja
Kelompok fokus	Dinamika kelompok Berbagai sudut pandang	Sulit dikelola, Mahal	Ide produk baru Persepsi merek	Diskusi tentang kampanye iklan baru
Survei daring	Jangkauan luas Hemat biaya	Tingkat respons rendah Bias online	Riset pasar Kepuasan pelanggan	Formulir umpan balik di situs web
Survei luring	Tidak bergantung pada internet Kehadiran fisik	Tantangan logistik Memakan waktu	Studi pedesaan Umpan balik langsung	Survei di acara komunitas lokal
Studi pengembangan teknologi masa depan	Perilaku alami Konteks nyata	Efek pengamat Risiko subjektivitas	Perilaku pengguna Studi sosiologis	Mengamati pembeli di toko
Pengikisan web	Data skala besar Otomatisasi	Masalah hukum Ketidakkonsistenan	Pembahasan pesaing Penambangan data	Mengekstraksi data dari situs web pesaing
Sensor/IoT	Data waktu nyata, Metrik spesifik	Biaya pemasangan mahal, Membutuhkan perawatan	Rumah pintar Pelacakan kesehatan	Pelacak kebugaran yang dapat dikenakan
Kueri basis data	Data terstruktur Konsistensi	Data mungkin sudah usang Cakupan terbatas	Laporan bisnis Metrik kinerja	Permintaan penjualan bulanan di CRM
Pembelajaran arsip	Konteks historis Komprehensif	Tidak selalu didigitalisasi Bisa terfragmentasi	Studi historis Riset kebijakan	Pembelajaran surat kabar lama

2.4 PERSIAPAN DATA

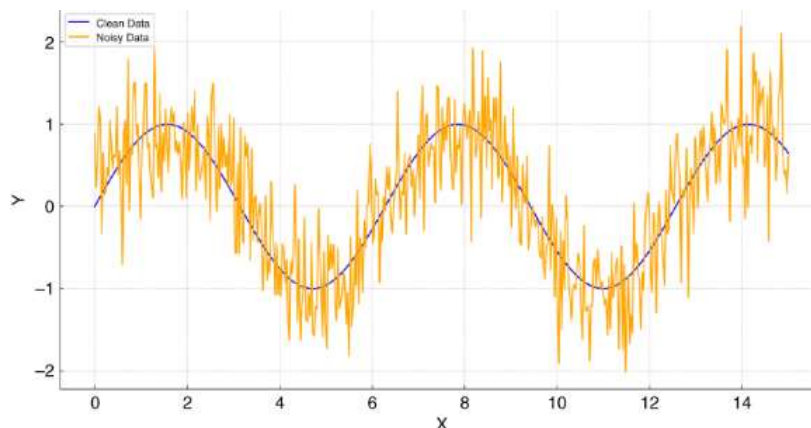
Setelah data dikumpulkan, sangat penting untuk mempersiapkannya untuk pembahasan guna memastikan kualitas, akurasi, dan relevansinya. Persiapan data merupakan langkah penting dalam proses ilmu data pemasaran, karena secara langsung memengaruhi efektivitas dan keandalan wawasan yang diperoleh dari pembahasan data. Ini adalah proses di mana data mentah, bising, dan seringkali tersebar diubah menjadi format yang terstruktur, bersih, dan dapat digunakan, siap untuk pembahasan atau pelatihan model. Kualitas data yang dimasukkan ke dalam algoritma pembelajaran mesin sangat memengaruhi kinerja model. Dengan demikian, menginvestasikan waktu dalam persiapan data yang menyeluruh seringkali menghasilkan hasil yang lebih akurat dan berwawasan. Pada bagian ini, kita akan mengeksplorasi langkah-langkah kunci yang terlibat dalam persiapan data secara lebih detail.

Pada Gambar 2.2, data bersih diwakili oleh garis abu-abu gelap, yang mengikuti pola gelombang sinus, dan data bising ditunjukkan dalam warna abu-abu terang, termasuk variasi acak untuk mewakili kebisingan yang ditambahkan ke sinyal bersih.

Pembersihan Data

Pembersihan data adalah proses mengidentifikasi dan mengatasi masalah dalam data, seperti nilai yang hilang, catatan duplikat, atau entri data yang salah. Teknik pembersihan data meliputi hal-hal berikut:

- **Imputasi.** Mengisi celah data sangat penting untuk memastikan kelengkapan suatu dataset. Meskipun teknik dasar seperti penggantian dengan rata-rata atau median umum digunakan, algoritma pembelajaran mesin tingkat lanjut dapat memprediksi nilai yang hilang dengan akurasi yang lebih tinggi.



Gambar 2.2 Data Berisik Versus Data yang Dibersihkan.

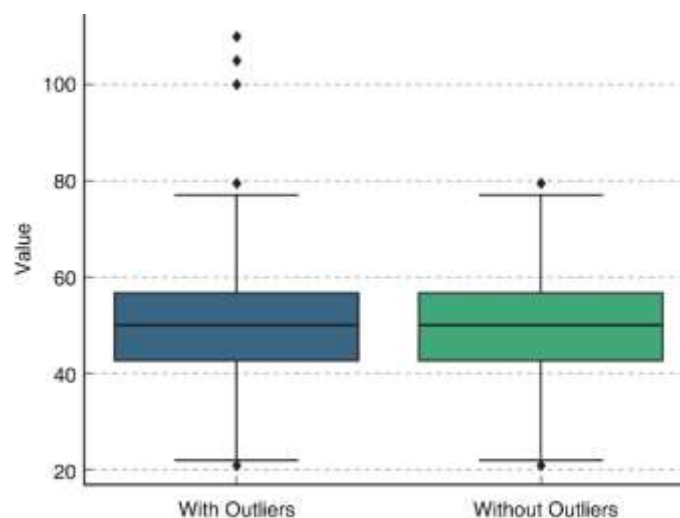
- **Penautan catatan.** Teknik ini memastikan bahwa entri yang merujuk pada entitas dunia nyata yang sama direpresentasikan sebagai satu kesatuan, memastikan konsistensi data dan mengurangi redundansi.
- **Deteksi outlier.** Outlier dapat memengaruhi hasil dan interpretasi. Meskipun beberapa outlier merupakan anomali yang sebenarnya, yang lain mungkin merupakan kesalahan data. Tekniknya berkisar dari ukuran statistik sederhana hingga algoritma pengelompokan tingkat lanjut.

Saat kita menjelajahi teknik yang digunakan dalam pembersihan data, penting untuk memvisualisasikan dampaknya. Gambar 2.3 menyajikan plot kotak yang menggambarkan data sebelum dan sesudah penghapusan outlier. Representasi visual ini merupakan kunci untuk memahami bagaimana teknik deteksi outlier dapat secara signifikan mengubah struktur dan interpretasi suatu dataset, yang menunjukkan efek praktis dari proses pembersihan ini.

Integrasi Data

Integrasi data melibatkan penggabungan data dari berbagai sumber, memastikan konsistensi, akurasi, dan kelengkapan di seluruh dataset. Teknik untuk integrasi data meliputi hal-hal berikut:

- **Pemetaan data.** Sebagai langkah dasar untuk memastikan dataset menggunakan bahasa yang sama, pemetaan data menyelaraskan bidang umum di seluruh sumber.
- **Transformasi data.** Untuk memastikan keseragaman di seluruh dataset, transformasi dapat melibatkan standardisasi unit, format, atau skala. Langkah ini menjamin konsistensi dan keterbandingan data.



Gambar 2.3 Data Sebelum dan Setelah Penghapusan Outlier.

- **Deduplikasi data.** Redundansi dapat mengaburkan pembahasan. Deduplikasi memastikan setiap entri data berbeda dan tidak ambigu, mengoptimalkan struktur data untuk pembahasan.

Transformasi Data

Transformasi data melibatkan konversi data mentah ke dalam format yang sesuai untuk pembahasan. Langkah ini dapat mencakup berbagai operasi, seperti yang dijelaskan di bagian selanjutnya.

Normalisasi

Normalisasi, dalam konteks persiapan data, mengacu pada proses penyesuaian variabel numerik sehingga dapat dibandingkan pada skala umum. Ini sangat penting untuk beberapa algoritma pembelajaran mesin, karena algoritma tersebut dapat sensitif terhadap variabel yang diukur pada skala yang berbeda. Berbagai metode normalisasi memenuhi berbagai kebutuhan analitis dan distribusi data:

- **Penskaliasian min-max.** Ini mungkin bentuk normalisasi yang paling intuitif. Untuk setiap titik data, metode ini mengurangi nilai minimum dari fiturnya dan kemudian membagi dengan rentang fitur tersebut. Akibatnya, seluruh dataset dibatasi dalam. Rumusnya adalah

$$\text{Normalized}(x) = \frac{x - \min(x)}{\max(x) - \min(x)}$$

di mana adalah nilai tertentu, $\min(x)$ adalah nilai terkecil dalam fitur tersebut, dan $\max(x)$ adalah nilai terbesar.

- **Penskaliasian desimal.** Di sini, data digeser dengan desimal. Setiap titik data dibagi dengan pangkat 10 tertinggi untuk nilai absolut maksimum dalam dataset. Transformasi ini membatasi data antara -1 dan 1.
- **Penskaliasian robust.** Penskalasian robust berguna untuk data yang mengandung banyak outlier dan menskalakan data berdasarkan median (bukan rata-rata) dan rentang interkuartil (bukan deviasi standar).

$$\text{Normalized}(x) = \frac{x - \text{median}(x)}{IQR(x)}$$

di mana IQR adalah rentang interkuartil, yaitu selisih antara persentil ke-75 (Q3) dan persentil ke-25 (Q1).

Setiap teknik normalisasi memiliki kelebihan masing-masing, dan pilihannya bergantung pada sifat dataset dan tujuan pembahasan selanjutnya. Misalnya, penskalaan min-max mungkin cocok untuk tugas pemrosesan gambar, sedangkan penskalaan desimal dapat bermanfaat ketika Anda ingin mengurangi orde besaran titik data, sehingga dataset menjadi lebih mudah dikelola.

Standardisasi

Standardisasi adalah proses penskalaan fitur sehingga memiliki sifat distribusi normal standar dengan rata-rata 0 dan deviasi standar 1. Ini sangat berguna ketika bekerja dengan algoritma yang sensitif terhadap skala fitur, seperti pengelompokan k-means, mesin vektor pendukung, atau algoritma apa pun yang bergantung pada perhitungan jarak atau optimasi penurunan gradien.

PEMBAHASAN PENGELUARAN PEMASARAN DAN RASIO KLIK (CTR)

Bayangkan sebuah departemen pemasaran di perusahaan ritel sedang menilai efisiensi saluran periklanannya. Mereka telah mengumpulkan data tentang dua metrik utama untuk setiap saluran: total pengeluaran pemasaran dan rasio klik (CTR) yang dihasilkan. Pengeluaran pemasaran dalam ribuan dolar (berkisar dari Rp10.000.000 hingga Rp1.000.000.000) dan CTR adalah persentase (berkisar dari 0% hingga 3%).

Untuk memahami saluran mana yang memberikan pengembalian investasi terbaik, sebuah model pembelajaran mesin sedang dirancang untuk memprediksi CTR

berdasarkan pengeluaran pemasaran. Namun, mengingat perbedaan besar dalam skala kedua metrik ini, penerapan langsung model tersebut mungkin menyebabkan model tersebut terlalu dipengaruhi oleh angka pengeluaran pemasaran karena besarnya angka tersebut.

Dengan menerapkan penskalaan min-max, baik pengeluaran pemasaran maupun CTR dinormalisasi antara 0 dan 1. Hal ini memastikan bahwa algoritma pembelajaran mesin memperlakukan kedua fitur tersebut secara adil, sehingga memungkinkan penilaian yang lebih akurat tentang hubungan antara pengeluaran pemasaran dan CTR.

Hasil pembahasan ini, yang didukung oleh dataset yang dinormalisasi dengan benar, mungkin mengungkapkan bahwa saluran-saluran tertentu, meskipun lebih mahal, memberikan CTR yang jauh lebih baik, sehingga dapat memandu keputusan alokasi anggaran tim pemasaran di masa mendatang.

- **Normalisasi skor-Z.** Ini adalah bentuk standardisasi yang paling umum. Untuk setiap fitur, nilai rata-rata dikurangi dari setiap titik data, dan hasilnya dibagi dengan simpangan baku. Skor-Z yang dihasilkan menunjukkan berapa banyak simpangan baku jarak suatu titik data dari nilai rata-rata.

$$Z = \frac{X - \mu}{\sigma}$$

di mana

- Z adalah nilai yang distandarisasi.
 - X adalah nilai fitur asli.
 - μ adalah rata-rata fitur.
 - σ adalah simpangan baku fitur.
- **Penskalasian robust.** Dengan adanya outlier, normalisasi z-score standar mungkin tidak berfungsi sebagaimana mestinya karena mean dan deviasi standar sensitif terhadap nilai ekstrem. Penskalasian robust mengatasi hal ini dengan menggunakan median dan rentang interkuartil (IQR) sebagai gantinya.

$$X_{scaled} = \frac{X - X_{median}}{IQR}$$

di mana

- IQR adalah rentang interkuartil, yaitu selisih antara persentil ke-75 dan persentil ke-25.
- **Penskalaan vektor satuan.** Pendekatan ini menskalakan komponen vektor fitur sedemikian rupa sehingga vektor lengkap memiliki panjang satu. Ini sering digunakan dalam klasifikasi teks atau pengelompokan untuk data yang jarang.

$$X_{unit} = \frac{X}{\|X\|},$$

di mana

- $\|X\|$ adalah norma Euclidean dari vektor fitur.

Saat memutuskan untuk menstandarisasi data, sangat penting untuk menyesuaikan skala hanya pada data pelatihan dan kemudian menerapkan transformasi yang sama pada dataset pelatihan dan pengujian. Ini mencegah kebocoran data, di mana informasi dari set pengujian dapat memengaruhi model selama pelatihan, berpotensi menyebabkan perkiraan kinerja yang terlalu optimis.

Singkatnya, standardisasi adalah langkah pra-pemrosesan penting yang dapat sangat meningkatkan kinerja dan stabilitas banyak algoritma pembelajaran mesin dengan memastikan bahwa fitur berada pada skala yang konsisten.

STUDI KASUS: MODEL PENILAIAN KREDIT DI PERBANKAN

Pertimbangkan sebuah bank yang ingin mengembangkan model pembelajaran mesin untuk memprediksi kelayakan kredit calon peminjam. Dataset berisi berbagai fitur seperti pendapatan tahunan, tingkat pemanfaatan kredit, dan usia:

- *Pendapatan tahunan para pemohon berkisar antara Rp100.000.000 hingga Rp10.000.000.000.*
- *Tingkat pemanfaatan kredit, yang merupakan fraksi dari kredit yang tersedia yang saat ini digunakan oleh individu, berkisar antara 0% hingga 100%.*
- *Usia para pemohon berkisar antara 18 hingga 80 tahun.*

Ketika mencoba mengelompokkan pemohon yang serupa menggunakan algoritma pengelompokan k-means, yang sensitif terhadap skala data, perbedaan besar dalam rentang fitur-fitur ini dapat memengaruhi hasil. Secara khusus, algoritma mungkin memprioritaskan pendapatan tahunan daripada fitur lain hanya karena nilai numeriknya yang lebih tinggi.

Dengan menstandarisasi fitur-fitur ini menggunakan normalisasi skor-Z, pendapatan tahunan, tingkat pemanfaatan kredit, dan usia disesuaikan agar memiliki rata-rata 0 dan deviasi standar 1. Hal ini memastikan bahwa setiap fitur berkontribusi secara setara terhadap metrik jarak yang digunakan dalam k-means.

Setelah standardisasi, bank menemukan bahwa algoritma pengelompokan mereka sekarang mengidentifikasi kelompok pelamar yang memiliki perilaku keuangan serupa, sehingga lebih mudah untuk menetapkan skor kredit atau suku bunga yang sesuai.

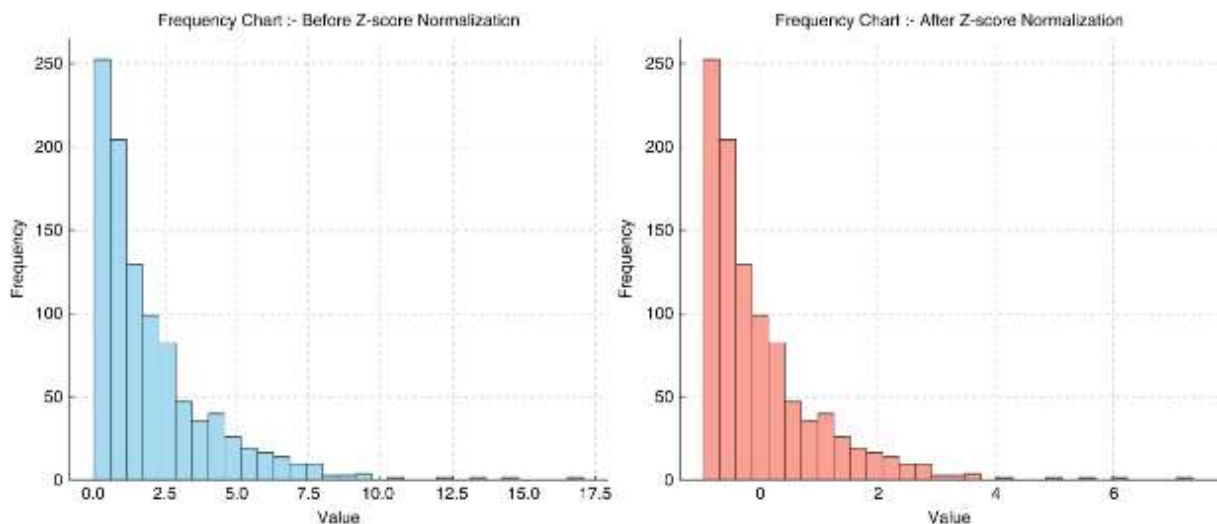
Selain itu, ketika bank menggunakan model seperti support vector machines untuk memprediksi gagal bayar pinjaman, data yang distandarisasi memastikan bahwa batas keputusan tidak dipengaruhi secara tidak proporsional oleh skala fitur tunggal apa pun, sehingga menghasilkan prediksi yang lebih akurat.

Gambar 2.4 menampilkan dua histogram yang mewakili distribusi frekuensi dari suatu dataset sebelum dan sesudah normalisasi z-score:

- Histogram di sebelah kiri menunjukkan distribusi frekuensi dari data asli yang miring ke kanan, dengan konsentrasi nilai di sisi kiri dan ekor panjang di sebelah kanan.
- Histogram di sebelah kanan menunjukkan distribusi frekuensi setelah normalisasi z-score, di mana data telah diskalakan agar memiliki rata-rata 0 dan deviasi standar 1. Transformasi ini tidak mengubah bentuk distribusi tetapi memusatkan data pada rata-rata dan mengekspresikan nilai dalam hal deviasi standar dari rata-rata.

Pengkodean Variabel Kategorikal

Sebelum membahas teknik pengkodean variabel kategorikal, penting untuk memahami mengapa langkah ini sangat diperlukan dalam pra-pemrosesan data. Variabel kategorikal tersebar luas dalam dataset, sering kali mewakili atribut kualitatif seperti jenis kelamin, kebangsaan, jenis produk, dan banyak lagi. Meskipun kategori-kategori ini membawa informasi yang signifikan, tantangannya adalah sebagian besar model pembelajaran mesin bersifat aljabar. Artinya, mereka mengharapkan input numerik dan tidak dapat memproses string atau data kategorikal dalam bentuk mentah.



Gambar 2.4 Data Sebelum dan Setelah Normalisasi Skor Z.

Untuk menjembatani kesenjangan ini, kita beralih ke strategi pengkodean, yang mengubah variabel kategorikal menjadi format numerik yang dapat digunakan oleh algoritma. Proses ini tidak hanya mempertahankan karakteristik informatif data tetapi juga memastikan kompatibilitas dengan berbagai teknik pembelajaran mesin. Pengkodean mempertahankan struktur penting data kategorikal, memungkinkan model untuk membedakan pola dan membuat prediksi yang bergantung pada fitur kategorikal ini.

Mari kita jelajahi beberapa metode utama untuk mengkodekan variabel kategorikal:

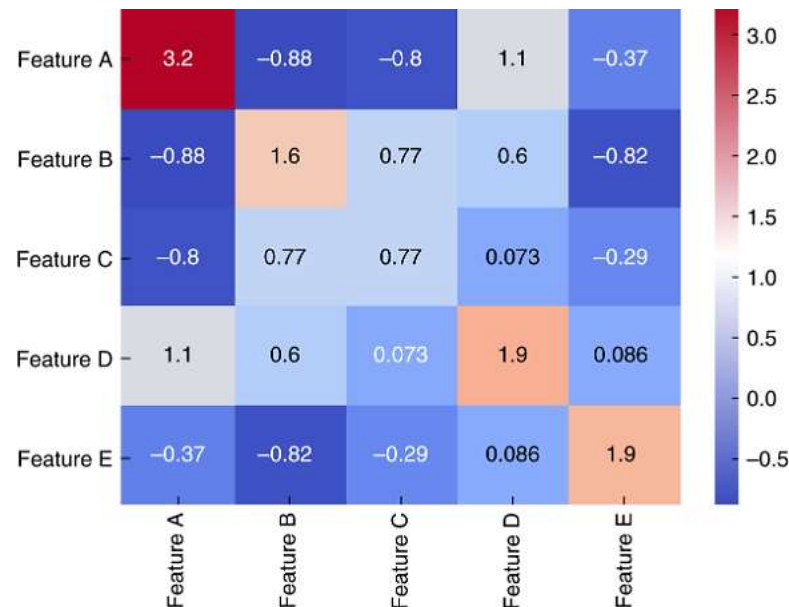
- **Pengkodean one-hot (variabel dummy).** Metode ini melibatkan konversi setiap kategori dari variabel kategorikal nominal menjadi kolom biner baru (0 atau 1). Misalnya, jika variabel warna memiliki tiga kategori (Merah, Hijau, Biru), pengkodean one-hot akan membuat tiga variabel baru—satu untuk setiap warna, di mana setiap

pengembangan teknologi masa depan direpresentasikan sebagai nilai biner di seluruh kolom baru ini.

- **Pengkodean ordinal.** Ketika variabel kategorikal memiliki peringkat atau urutan yang jelas (seperti rendah, sedang, tinggi), variabel tersebut dapat dikodekan dengan nilai ordinal. Misalnya, “rendah” dapat dikodekan sebagai 1, “sedang” sebagai 2, dan “tinggi” sebagai 3. Ini mempertahankan urutan inheren dalam kategori.
- **Pengkodean label.** Pengkodean label mirip dengan pengkodean ordinal tetapi digunakan untuk variabel nominal tanpa urutan inheren. Setiap kategori unik diberi nilai bilangan bulat. Meskipun sederhana, metode ini mungkin secara tidak sengaja memperkenalkan urutan dalam data yang tidak ada, berpotensi membingungkan beberapa algoritma.
- **Pengkodean biner.** Metode ini pertama-tama menetapkan label bilangan bulat ke kategori (seperti pengkodean label) dan kemudian mengubah bilangan bulat ini menjadi kode biner. Oleh karena itu, setiap kategori direpresentasikan sebagai bilangan biner. Ini lebih efisien daripada pengkodean one-hot untuk variabel dengan banyak kategori.
- **Pengkodean frekuensi (atau hitungan).** Kategori dikodekan berdasarkan frekuensi atau hitungannya dalam dataset. Metode ini dapat berguna ketika ada korelasi antara frekuensi fitur dan variabel target. Namun, ini mungkin memperkenalkan kolinearitas.

Gambar 2.5 menyajikan peta panas korelasi fitur, yang secara visual merepresentasikan korelasi antara fitur-fitur yang berbeda setelah menerapkan metode pengkodean ini. Visualisasi ini sangat berguna untuk mengamati potensi masalah kolinearitas, seperti yang mungkin muncul dari pengkodean frekuensi, sehingga menawarkan wawasan tentang pemilihan teknik pengkodean yang paling tepat untuk dataset tertentu.

- **Pengkodean rata-rata (atau target).** Ini melibatkan pengkodean kategori berdasarkan rata-rata variabel target untuk kategori tersebut. Misalnya, dalam masalah klasifikasi biner, kategori dapat dikodekan dengan rata-rata variabel target (0 atau 1) untuk setiap kategori. Meskipun ampuh, hal ini dapat menyebabkan overfitting, jadi harus berhati-hati, dan penting untuk menggunakan teknik seperti regularisasi atau menambahkan noise ke nilai yang dikodekan.



Gambar 2.5 Peta Panas Korelasi Fitur.

- **Lapisan penyematan.** Model pembelajaran mendalam, terutama jaringan saraf, memungkinkan penggunaan lapisan penyematan untuk menangani variabel kategorikal. Lapisan penyematan mempelajari representasi multidimensi dari setiap kategori selama proses pelatihan, yang dapat sangat berguna untuk menangani variabel kategorikal dengan kardinalitas tinggi.

Saat mengkodekan variabel kategorikal, penting untuk mempertimbangkan sifat algoritma dan data. Beberapa algoritma dapat menangani data kategorikal secara langsung, dan yang lain mungkin sensitif terhadap nilai numerik yang ditetapkan melalui pengkodean, yang berpotensi menyebabkan salah tafsir. Pengkodean variabel kategorikal yang tepat sangat penting untuk membangun model prediktif yang secara akurat menangkap pola dan hubungan yang mendasari dalam data. Dengan menerapkan teknik pengkodean ini, kita memfasilitasi transisi yang lancar dari data non-numerik ke format yang dapat dengan mudah diproses oleh model statistik, membuka jalan bagi pembahasan data yang lebih akurat dan mendalam.

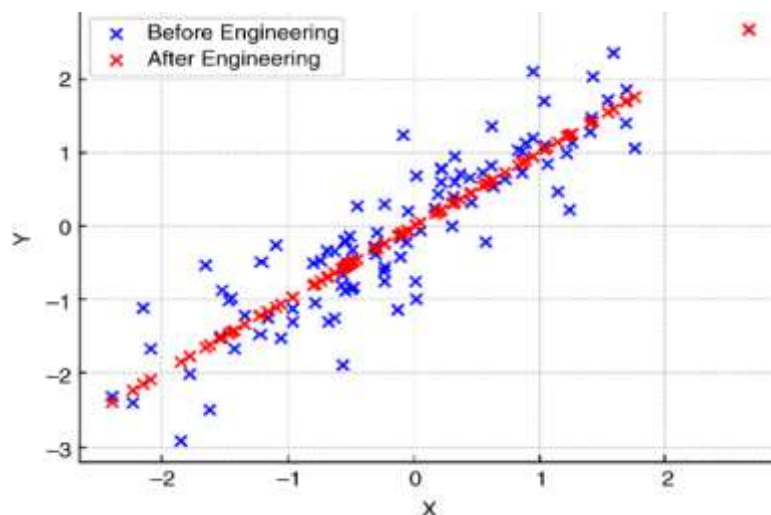
Rekayasa Fitur

Rekayasa fitur adalah seni menciptakan fitur baru dari fitur yang sudah ada, meningkatkan daya prediksi model pembelajaran mesin. Dengan mengubah data mentah menjadi bentuk yang lebih sesuai atau representatif, algoritma seringkali dapat menemukan pola dengan lebih efektif. Proses ini membutuhkan kreativitas, intuisi, dan keahlian domain.

- **Fitur polinomial.** Memperkenalkan suku polinomial (kuadrat, kubik, dan lain-lain.) atau suku interaksi antara dua atau lebih fitur dapat menangkap hubungan nonlinier antara fitur dan variabel target. Misalnya, jika memprediksi harga rumah, suku interaksi antara 'jumlah kamar' dan 'lokasi' mungkin dapat menangkap efek gabungan dengan lebih baik daripada memperlakukannya secara independen.
- **Pengelompokan/Baseting.** Fitur kontinu dapat diubah menjadi fitur kategorikal dengan mengelompokkannya ke dalam interval atau 'bin' yang berbeda. Misalnya, usia dapat dikelompokkan menjadi 'anak', 'remaja', 'dewasa', dan seterusnya. Ini sangat berguna ketika ada hubungan nonlinier di berbagai rentang variabel kontinu.

- **Fitur temporal.** Untuk data deret waktu, membuat fitur yang menangkap musiman (hari dalam seminggu, bulan, kuartal, dan lain-lain.), tren, atau waktu sejak peristiwa tertentu dapat sangat berharga. Misalnya, untuk situs web e-commerce, 'hari sejak kunjungan terakhir' atau 'pembelian dalam sebulan terakhir' mungkin merupakan prediktor yang signifikan.
- **Fitur teks.** Data teks dapat diubah menjadi fitur numerik menggunakan teknik seperti TF-IDF (frekuensi istilah-frekuensi dokumen terbalik), penyematan kata (misalnya Word2Vec), atau bag-of-words. Metode ini menangkap semantik dan konteks teks dalam bentuk yang dapat diolah oleh algoritma.
- **Fitur geospasial.** Jika berurusan dengan data lokasi, fitur baru seperti 'jarak dari landmark tertentu', 'kepadatan fasilitas tertentu di sekitarnya', atau 'kedekatan dengan rute transportasi umum' dapat memberikan wawasan yang berharga.
- **Agregasi.** Membuat fitur ringkasan dari suatu himpunan, seperti 'nilai transaksi rata-rata selama enam bulan terakhir', 'denyut jantung maksimum dalam 10 menit terakhir', atau 'frekuensi suatu peristiwa tertentu' dapat menambah daya prediksi yang signifikan.
- **Dekomposisi.** Teknik seperti PCA (pembahasan komponen utama) atau T-SNE dapat digunakan untuk mengurangi dimensi data, menciptakan fitur baru yang mempertahankan variabilitas maksimum dari data asli.
- **Fitur spesifik domain.** Tergantung pada masalah atau industri tertentu, fitur khusus tertentu dapat dibuat. Misalnya, di bidang keuangan, fitur seperti rasio P/E atau rata-rata bergerak dapat digunakan. Di bidang perawatan kesehatan, BMI (indeks massa tubuh) dapat dihitung dari tinggi dan berat badan.

Menggabungkan rekayasa fitur dapat secara substansial meningkatkan kinerja model (lihat Gambar 2.6). Namun, penting untuk memastikan bahwa fitur tersebut tidak terlalu kompleks, yang dapat menyebabkan overfitting. Konsultasi rutin dengan pakar bidang terkait dan pengujian model secara iteratif dapat membantu dalam menyempurnakan dan memvalidasi efektivitas fitur yang direkayasa.



Gambar 2.6 Data Sebelum dan Setelah Rekayasa Fitur.

Pengurangan Data

Pengurangan data melibatkan pengurangan kompleksitas dan ukuran dataset sambil mempertahankan informasi kuncinya. Teknik untuk pengurangan data meliputi hal-hal berikut.

Pengurangan Dimensi

Pengurangan dimensi mengacu pada proses transformasi data berdimensi tinggi menjadi bentuk berdimensi lebih rendah sambil mempertahankan sebanyak mungkin informasi yang relevan. Pengurangan ini dapat menghasilkan model yang lebih sederhana, peningkatan kinerja, dan interpretasi yang lebih baik. Berikut adalah uraian yang lebih detail:

Pembahasan Komponen Utama (PCA)

PCA adalah teknik linier yang mengidentifikasi "komponen utama" dari data, yang ortogonal (tegak lurus) satu sama lain dan menangkap varians maksimum dalam data.

- PCA memproyeksikan ulang data ke komponen utama ini, seringkali mengurangi dimensi secara signifikan.
- Meskipun beberapa komponen utama pertama biasanya menangkap sebagian besar varians dalam dataset, komponen tersebut mungkin tidak selalu dapat diinterpretasikan dalam konteks fitur aslinya.
- PCA sangat berguna untuk variabel kontinu dan ketika ada kebutuhan untuk menghilangkan multikolinearitas.

t-Distributed Stochastic Neighbor Embedding (t-SNE)

- t-SNE adalah teknik nonlinier yang sangat berguna untuk memvisualisasikan data berdimensi tinggi dalam dua atau tiga dimensi.
- Teknik ini menekankan pada mempertahankan struktur lokal data, memastikan bahwa instance yang berdekatan di ruang berdimensi tinggi tetap berdekatan di ruang berdimensi rendah.
- Meskipun ampuh untuk visualisasi, t-SNE biasanya tidak digunakan untuk pengurangan fitur dalam pemodelan.

Pembahasan Diskriminan Linier (LDA)

- LDA adalah pengklasifikasi dan teknik pengurangan dimensi. Tujuannya adalah untuk menemukan kombinasi linier fitur yang paling baik memisahkan dua atau lebih kelas dalam suatu dataset.
- Teknik ini sangat efektif ketika kelas-kelas tersebut terpisah dengan baik, dan tujuannya adalah untuk mengurangi dimensi dan mempertahankan keterpisahan kelas.

Autoencoder

- Autoencoder adalah jaringan saraf yang digunakan untuk pengurangan dimensi. Autoencoder dirancang untuk mempelajari pengkodean atau kompresi data input yang efisien, kemudian mendekode atau merekonstruksi data dari pengkodean ini.
- Lapisan tengah mewakili bentuk terkompresi dari data input. Autoencoder serbaguna dan dapat menangkap pola linier dan nonlinier dalam data.

Pemilihan Fitur

- Meskipun bukan teknik transformasi seperti yang lain, pemilihan fitur melibatkan mempertahankan subset fitur asli berdasarkan kepentingan, varians, atau relevansinya terhadap variabel target.
- Tekniknya meliputi eliminasi mundur, pemilihan maju, eliminasi fitur rekursif, dan menggunakan model berbasis pohon (seperti pohon keputusan atau hutan acak) untuk memberi peringkat kepentingan fitur.
- Perlu dicatat bahwa pemilihan fitur mempertahankan fitur asli yang dapat diinterpretasikan, tidak seperti metode seperti PCA.

Pilihan teknik pengurangan dimensi seringkali bergantung pada sifat masalah, karakteristik data, dan tujuan pembahasan.

Pengambilan Sampel

Pengambilan sampel adalah teknik memilih subset dari dataset yang lebih besar untuk membuat pembahasan data lebih mudah dikelola atau hemat biaya. Metode pengambilan sampel yang berbeda melayani berbagai kebutuhan:

- **Pengambilan sampel acak sederhana.** Setiap titik data memiliki peluang yang sama untuk dipilih, yang berguna ketika tidak perlu fokus pada subkelompok tertentu. Contoh: Bayangkan sebuah mangkuk berisi 1.000 kelereng berwarna (masing-masing mewakili satu titik data). Tutup mata Anda dan ambil 100 kelereng. Ini adalah sampel acak Anda.
- **Pengambilan sampel bertingkat.** Kumpulan data dibagi menjadi subkelompok, dan sampel diambil dari masing-masing subkelompok. Ini bagus untuk memastikan semua subkelompok terwakili, seperti kategori usia atau jenis kelamin. Contoh: Misalkan Anda memiliki sekolah dengan 1.000 mahasiswa, 500 laki-laki dan 500 perempuan. Jika Anda ingin mengambil sampel 100 mahasiswa, Anda dapat mengambil 50 dari setiap kelompok jenis kelamin untuk memastikan kedua jenis kelamin terwakili secara setara.
- **Pengambilan sampel klaster.** Data dibagi menjadi klaster, dan beberapa klaster dipilih secara acak untuk dipelajari. Ini berguna ketika data mencakup area yang luas, seperti mempelajari pembeli di kota-kota tertentu. Contoh: Jika Anda mempelajari kebiasaan pembelian ritel di suatu negara dengan 50 kota, Anda mungkin secara acak memilih 5 kota (kluster) dan kemudian mensurvei semua pelanggan atau sampel acak pelanggan di setiap kota yang dipilih tersebut.
- **Pengambilan sampel sistematis.** Pilih setiap item ke- n dari sebuah daftar. Ini berguna untuk interval reguler, seperti memeriksa setiap produk ke-10 dari jalur perakitan. Contoh: Anda sedang memeriksa kualitas barang di jalur produksi yang menghasilkan 1.000 barang per hari. Anda memutuskan untuk memeriksa setiap item ke-10, jadi Anda akan memeriksa item ke-10, ke-20, ke-30, dan seterusnya.
- **Pengambilan sampel berdasarkan kemudahan.** Metode ini membantu memilih data yang paling mudah didapatkan. Metode ini cepat tetapi dapat bias. Metode ini sering digunakan untuk studi pendahuluan.
- **Contoh:** Sebuah perusahaan minuman ringan mendirikan stan pencicipan di pusat perbelanjaan dan meminta orang yang lewat untuk mencicipi dan memberikan umpan

balik. Di sini, mereka mengambil sampel siapa pun yang datang ke stan, yang memang mudah tetapi belum tentu mewakili pasar yang lebih luas.

- **Pengambilan sampel berdasarkan kuota.** Memilih sampel berdasarkan kriteria atau kuota tertentu. Metode ini digunakan untuk memastikan kategori tertentu, seperti kelompok usia, tercakup dalam sampel. Contoh: Sebuah jaringan TV menginginkan umpan balik tentang acara baru. Mereka memutuskan membutuhkan 100 pemirsa dari setiap kelompok usia: 18–30, 31–50, dan 51+. Kemudian mereka mengambil sampel pemirsa hingga memenuhi kuota ini untuk setiap kelompok.

Saat mengambil sampel, sangat penting untuk memilih metode dan ukuran yang tepat untuk memastikan sampel Anda mewakili kumpulan data yang lebih besar secara akurat.

Agregasi

Agregasi adalah proses mengubah data mentah menjadi bentuk ringkasan atau tingkat yang lebih tinggi agar lebih mudah dipahami dan mengurangi ukurannya untuk pembahasan. Ini sangat berguna dalam kasus di mana data terperinci lebih granular daripada yang dibutuhkan untuk pembahasan spesifik. Berikut penjelasan lebih lanjut:

Tujuan Agregasi

- **Penyederhanaan.** Data yang teragregasi lebih mudah dikelola dan dipembahasan.
- **Pengurangan.** Dengan meringkas data, Anda mengurangi volumenya, sehingga pemrosesan dan visualisasi menjadi lebih cepat.
- **Peningkatan privasi.** Mengagregasi data dapat membantu dalam menganonimkan detail pribadi, melindungi privasi individu.

Teknik Agregasi Umum

- **Jumlah.** Total dari sekumpulan angka tertentu.
- **Rata-rata.** Nilai rata-rata dari suatu kumpulan.
- **Maks/min.** Nilai tertinggi dan terendah dalam suatu kumpulan.
- **Hitungan.** Jumlah kejadian atau item dalam suatu kumpulan.
- **Median.** Nilai tengah dalam daftar yang diurutkan.
- **Modus.** Nilai yang paling sering muncul dalam suatu kumpulan.

Agregasi Temporal

- Ini melibatkan meringkas data dari waktu ke waktu. Misalnya, angka penjualan harian dapat diagregasikan untuk memberikan total bulanan atau tahunan.
- Ini berguna untuk mengidentifikasi tren dan pola jangka panjang.

Agregasi Spasial

- Ini berarti meringkas data berdasarkan wilayah spasial. Misalnya, data tingkat kota dapat diagregasikan untuk memberikan gambaran di tingkat negara.
- Ini membantu dalam mengpembahasan tren dan pola geografis.

Agregasi Kategorikal

- Ini mengelompokkan data berdasarkan kategori, misalnya, mengagregasikan data penjualan berdasarkan jenis produk.
- Ini berguna untuk memahami bagaimana kinerja berbagai kategori relatif satu sama lain.

Tantangan

- **Kehilangan detail.** Agregasi dapat menyebabkan hilangnya detail yang lebih halus. Meskipun ini seringkali menjadi tujuannya, penting untuk memastikan granularitasnya masih sesuai untuk pembahasan.
- **Risiko salah tafsir.** Data yang diagregasikan terkadang dapat menutupi outlier atau anomali yang mungkin signifikan.
- **Keseimbangan.** Sangat penting untuk mencapai keseimbangan antara agregasi berlebihan, yang dapat menyembunyikan wawasan yang berguna, dan agregasi kurang, yang dapat membingungkan dengan terlalu banyak detail.

Agregasi adalah konsep fundamental dalam ilmu data, terutama dalam pemasaran di mana seringkali ada kebutuhan untuk menyajikan data dalam format yang mudah dicerna dan dapat ditindaklanjuti. Menentukan tingkat agregasi membutuhkan pemahaman yang mendalam tentang data dan pertanyaan bisnis yang sedang ditangani.

Dengan mengikuti praktik terbaik untuk persiapan data ini, Anda dapat memastikan bahwa proyek ilmu data pemasaran Anda dibangun di atas fondasi yang kokoh berupa data yang akurat, relevan, dan berkualitas tinggi, memungkinkan Anda untuk memperoleh wawasan yang andal dan membuat keputusan berbasis data yang mendorong nilai bisnis yang signifikan.

Persiapan data dapat menjadi proses yang memakan waktu dan kompleks, tetapi menginvestasikan upaya dan sumber daya yang diperlukan pada tahap ini akan memberikan keuntungan dalam jangka panjang dengan meningkatkan akurasi, relevansi, dan efektivitas inisiatif ilmu data pemasaran Anda. Dengan menggunakan alat, teknik, dan praktik terbaik yang tepat, Anda dapat mengubah data mentah dan tidak terstruktur menjadi informasi berharga yang memberdayakan organisasi Anda untuk membuat keputusan yang lebih baik dan mencapai tujuan pemasarannya.

Singkatnya, proses persiapan data melibatkan pembersihan, pengintegrasian, transformasi, dan reduksi data untuk memastikan data tersebut akurat, relevan, dan siap untuk dibahas. Dengan mempersiapkan data Anda secara cermat, Anda akan meletakkan dasar untuk proyek ilmu data pemasaran yang sukses dan memastikan bahwa wawasan yang Anda peroleh dari data Anda dapat diandalkan, dapat ditindaklanjuti, dan berdampak.

2.5 KOLEKSI & OLAH DATA CHURN PELANGGAN

Sebuah perusahaan telekomunikasi ingin meng pembahasan churn pelanggan dengan mempelajari pola perilaku pelanggan dan data penggunaan. Saat kita mempelajari proses persiapan data yang teliti untuk pembahasan churn pelanggan, akan sangat membantu jika kita melihat cuplikan data mentah yang dihadapi perusahaan telekomunikasi tersebut. Tabel 2.3 menyajikan contoh data mentah tersebut, yang menunjukkan sebagian catatan yang mencakup potensi masalah yang perlu ditangani dalam langkah-langkah persiapan data. Tabel ini memberikan perspektif dunia nyata tentang jenis tantangan yang umum dihadapi, seperti nilai yang hilang, duplikat, dan entri yang salah, dan menyiapkan pemahaman tentang

bagaimana masalah ini diidentifikasi dan diperbaiki. Untuk mengumpulkan dan mempersiapkan data yang diperlukan, mereka mengikuti langkah-langkah berikut:

1. **Pengumpulan data.** Perusahaan mengumpulkan data pelanggan dari sistem CRM mereka, catatan penagihan, catatan detail panggilan, dan survei umpan balik pelanggan. Selain itu, mereka mengumpulkan data tentang harga dan promosi pesaing melalui web scraping dan penyedia pihak ketiga.
2. **Integrasi data.** Mereka menggabungkan data yang terkumpul ke dalam satu set data terpadu, memastikan konsistensi dan akurasi di semua sumber.
3. **Pembersihan data.** Perusahaan mengidentifikasi dan mengatasi masalah seperti nilai yang hilang, duplikat, dan entri data yang salah.
4. **Transformasi data.** Mereka mengubah data mentah menjadi format yang sesuai untuk pembahasan, menormalisasi variabel numerik, mengkodekan variabel kategorikal, dan merekayasa fitur baru, seperti durasi panggilan rata-rata, pengeluaran bulanan, dan jumlah interaksi layanan pelanggan.
5. **Pengurangan data.** Perusahaan mengurangi kompleksitas dan ukuran set data dengan memilih fitur yang relevan, menggunakan teknik pengurangan dimensi, dan mengagregasi data jika sesuai.

Tabel 2.3 Contoh Data Mentah untuk Pembahasan Pengurangan Pelanggan, Menunjukkan Sebagian Catatan yang Mencakup Potensi Masalah yang Harus Diatasi dalam Langkah-Langkah Persiapan Data.

ID Pelanggan	Usia	Durasi Langganan (hari)	Rata-rata Penggunaan Bulanan	Login Terakhir	Skor Umpan Balik	Berhenti Berlangganan
1	62.0	100	78	2022-01-31	4.0	False
2	18.0	118	73	2022-02-28	1.0	False
3	21.0	344	10	2022-03-31	2.0	True
4	NaN	223	21	2022-04-30	4.0	False
5	57.0	69	81	2022-05-31	NaN	True
6	27.0	117	70	2022-06-30	4.0	True
7	37.0	204	80	2022-07-31	4.0	False
8	39.0	118	48	2022-08-31	1.0	False
9	54.0	195	65	2022-09-30	2.0	True
10	41.0	55	83	2022-10-31	2.0	False

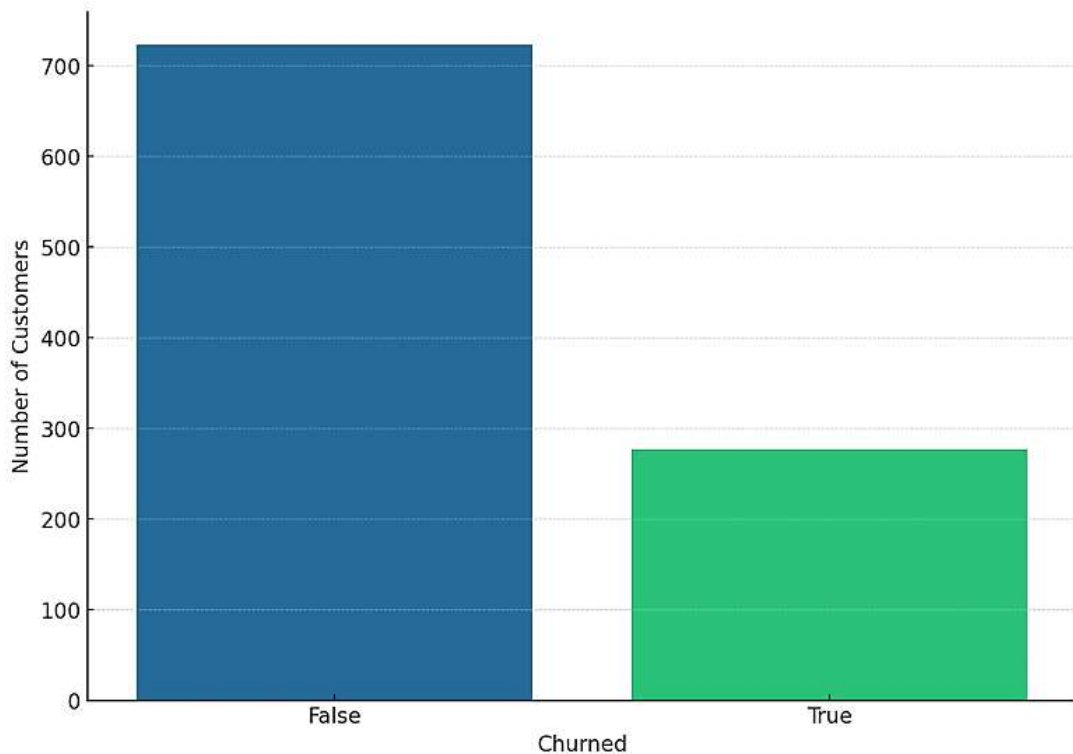
Gambar 2.7 menunjukkan distribusi pelanggan yang berhenti berlangganan versus pelanggan yang bertahan dalam dataset yang telah dibersihkan dan disiapkan.

- Batang "Benar" mewakili pelanggan yang berhenti berlangganan.
- Batang "Salah" mewakili pelanggan yang bertahan.

Dari visualisasi tersebut, kita dapat menyimpulkan bahwa proporsi pelanggan yang bertahan lebih besar, dan proporsi yang berhenti berlangganan lebih kecil.

Dengan mengumpulkan dan mempersiapkan data secara efektif, perusahaan telekomunikasi dapat memastikan bahwa tim ilmu data pemasaran mereka memiliki akses ke data yang akurat, relevan, dan berkualitas tinggi untuk pembahasan pelanggan yang berhenti

berlangganan. Hal ini, pada gilirannya, akan memungkinkan mereka untuk mengembangkan strategi berbasis data untuk mengurangi pelanggan yang berhenti berlangganan dan meningkatkan retensi pelanggan.



Gambar 2.7 Distribusi Pelanggan yang Berhenti Berlangganan versus Pelanggan yang Bertahan dalam Dataset yang Telah Dibersihkan dan Disiapkan.

2.6 KESIMPULAN

Kesimpulannya, proses pengumpulan dan persiapan data yang rumit membentuk tulang punggung ilmu data pemasaran yang efektif. Bab-bab awal buku ini telah menyajikan peta jalan melalui medan sumber data yang beragam, memberikan kejelasan tentang bagaimana munculnya big data telah merevolusi lanskap pemasaran. Saat kami membedah berbagai metodologi untuk akuisisi data—dari survei tradisional hingga API mutakhir dan seterusnya—kami menjelaskan kekuatan dan potensi jebakan yang melekat pada setiap pendekatan.

Bagian akhir bab ini berfungsi sebagai pendalaman mendalam tentang seni persiapan data yang teliti. Fase penting ini, yang sering diremehkan, adalah tempat data mentah dimurnikan menjadi aset strategis. Melalui pembersihan, integrasi, transformasi, dan pengurangan, data berhenti menjadi sekadar produk sampingan dari operasi bisnis dan menjadi sumber kehidupan pengambilan keputusan strategis.

Di sinilah, di tengah-tengah pengolahan data, fondasi untuk pembelajaran canggih diletakkan. Dengan menguasai langkah-langkah awal ini, ilmuwan data pemasaran mengubah kebisingan menjadi simfoni wawasan, membuka jalan bagi strategi yang dapat ditindaklanjuti yang beresonansi dengan presisi dan efektivitas.

Saat kita beralih ke bab-bab selanjutnya, kita membawa serta pemahaman bahwa persiapan data yang menyeluruh bukanlah sekadar langkah awal tetapi proses berkelanjutan yang sejalan dengan semua aktivitas ilmu data pemasaran. Ketelitian dan pandangan jauh ke depan yang diterapkan di sini bergema di seluruh siklus hidup pembahasan data, memengaruhi akurasi prediksi, relevansi wawasan, dan potensi strategi pemasaran.

Oleh karena itu, biarkan bab ini berfungsi sebagai bukti dan panduan bagi kekuatan transformatif dari data yang dikumpulkan dengan baik dan dikurasi dengan cermat. Saat kita menjelajahi lebih dalam ranah pembelajaran tingkat lanjut, pembelajaran mesin, dan seterusnya, pelajaran yang dipelajari di sini akan menjadi bintang penuntun yang memastikan integritas dan keberhasilan upaya pemasaran berbasis data Anda.

LATIHAN 2.1

PEMBERSIHAN DAN TRANSFORMASI DATA

Tujuan:

Membersihkan dan mentransformasi dataset untuk mempersiapkannya untuk pembahasan.

Tugas:

1. Menangani nilai yang hilang (NaN) dalam dataset "data_cleaning_transformation.csv".
2. Mengonversi 'LastLoginDate' dari string ke objek datetime.
3. Membuat fitur baru, 'MonthlySpendperDay', dengan membagi 'Pengeluaran Bulanan' dengan 'Lama Berlangganan'.

Langkah-langkah:

1. **Mengimpor Pustaka yang Diperlukan:**

```
import pandas as pd
```

- **Pandas** digunakan untuk manipulasi dan pembahasan data.

2. **Memuat Data:**

```
data_exercise_1 = pd.read_csv('path_to_csv_file')
```

Baris kode ini membaca file CSV yang berisi data ke dalam Pandas DataFrame, sehingga memungkinkan kita untuk mengolah data tersebut di Python.

3. **Menangani Nilai yang Hilang:**

- **Mengisi Nilai 'Age' yang Hilang:**

```
mean_age = data_exercise_1['Age'].mean()  
data_exercise_1['Age'].fillna(mean_age, inplace=True)
```

Di sini, kita menghitung rata-rata kolom 'Age' dan mengisi nilai yang hilang (NaN) di kolom 'Age' dengan rata-rata ini. Pendekatan ini dipilih karena data usia biasanya mengikuti distribusi normal, sehingga rata-rata menjadi perkiraan yang baik untuk nilai yang hilang.

- **Mengisi Nilai 'Pengeluaran Bulanan' yang Hilang:**

```
median_monthly_spend=data_exercise_1['MonthlySpend($)'].median()  
data_exercise_1['MonthlySpend($)'].fillna(median_monthly_spend,  
inplace=True)
```

Kami mengisi nilai yang hilang pada 'MonthlySpend(\$)' dengan median, karena data keuangan seringkali memiliki nilai pencilan, dan median kurang sensitif terhadap nilai pencilan tersebut dibandingkan dengan rata-rata.

▪ **Mengisi Nilai 'Feedback Score' yang Hilang:**

```
mode_feedback = data_exercise_1['Feedback Score'].mode()[0]
data_exercise_1['FeedbackScore'].fillna(mode_feedback,inplace=True)
```

Untuk 'Feedback Score', kami menggunakan modus (nilai yang paling sering muncul) untuk mengisi nilai yang hilang, karena skor ini kemungkinan mewakili data kategorikal atau ordinal.

4. Mengonversi 'LastLoginDate' ke DateTime:

```
data_exercise_1['LastLoginDate'].pd.to_datetime(data_exercise_1['Last Login Date'])
```

'LastLoginDate' awalnya dibaca sebagai string. Baris ini mengubahnya menjadi objek Pandas DateTime, sehingga lebih mudah untuk melakukan operasi terkait tanggal di kemudian hari.

5. Membuat Fitur Baru: 'MonthlySpendperDay':

```
data_exercise_1['MonthlySpendperDay']=data_exercise_1['MonthlySpend ($)'] / data_exercise_1['Subscription Length (days)']
```

Kita membuat kolom baru, 'MonthlySpendperDay' dengan membagi 'MonthlySpend(\$)' dengan 'Subscription Length (days)'. Fitur baru ini memberikan wawasan tambahan tentang kebiasaan pengeluaran pelanggan per hari.

Langkah-langkah ini secara komprehensif mencakup proses pembersihan data dan rekayasa fitur yang diperlukan untuk mempersiapkan dataset untuk pembahasan lebih lanjut.

LATIHAN 2.2

AGREGASI DAN REDUKSI DATA

Tujuan

Melakukan agregasi data dan reduksi dimensi pada dataset pemasaran.

Tugas:

1. Menggabungkan data "**data_aggregation_reduction.csv**" berdasarkan 'Wilayah' dan menghitung rata-rata 'Pengeluaran Bulanan' dan total 'Frekuensi Pembelian' per wilayah.
2. Melakukan pembahasan komponen utama (PCA) untuk mengurangi dimensi data sambil mempertahankan informasi kunci.

Langkah-langkah:

1. Mengimpor Pustaka yang Diperlukan:

```
1. import pandas as pd
2. from sklearn.decomposition import PCA
3. import matplotlib.pyplot as plt
```

- Pandas digunakan untuk manipulasi dan pembahasan data.
- PCA dari sklearn.decomposition adalah teknik pengurangan dimensi.
- Matplotlib digunakan untuk visualisasi.

2. Memuat Data:

```
4. data_exercise_2 = pd.read_csv('path_to_csv_file')
```

Baris ini membaca file CSV yang berisi data untuk Latihan 2.2 ke dalam Pandas DataFrame, memungkinkan manipulasi dan pembahasan data.

5. Agregasi Data:

Agregasi berdasarkan 'Region':

```
5. region_aggregated_data = data_exercise_2.groupby('Region').agg(
    Average_Monthly_Spend=pd.NamedAgg(column='MonthlySpend($)', aggfunc=
    'mean'), Total_Purchase_Frequency=pd.NamedAgg(column='PurchaseFrequency',
    aggfunc='sum') ).reset_index()
```

Di sini, kita mengelompokkan data berdasarkan kolom 'Region' dan menghitung dua metrik agregat: rata-rata 'MonthlySpend(\$)' dan total 'PurchaseFrequency' untuk setiap wilayah. Fungsi **groupby** dan **agg** di Pandas digunakan untuk mencapai hal ini, memberikan ringkasan perilaku pengeluaran dan pembelian berdasarkan wilayah.

6. Pengurangan Data menggunakan Pembahasan Komponen Utama (PCA):

▪ Mempersiapkan Data untuk PCA:

```
6. pca_data = data_exercise_2[['Monthly Spend ($)', 'Purchase
Frequency']]
```

Kami hanya memilih kolom numerik 'Monthly Spend (\$)' dan 'Purchase Frequency' untuk PCA.

▪ Standardisasi Data:

```
7. pca_data_standardized=(pca_data-pca_data.mean())/pca_data.std()
```

PCA sensitif terhadap skala data, jadi kami menstandarisasi fitur agar memiliki rata-rata 0 dan deviasi standar 1.

▪ Melakukan PCA:

```
8. pca = PCA(n_components=2)
```

```
9. principal_components=pca.fit_transform(pca_data_standardized)
```

Kami membuat objek PCA untuk mengurangi data menjadi dua dimensi. Metode **'fit_transform'** menghitung komponen utama dan mentransformasikan data sesuai dengan itu.

▪ Membuat DataFrame dengan Komponen Utama:

```
10. principal_df=pd.DataFrame(data=principal_components,columns=['Principal Component 1', 'Principal Component 2'])
```

Komponen utama yang dihasilkan disimpan dalam DataFrame baru. Komponen-komponen ini adalah titik data yang telah ditransformasikan dalam ruang dua dimensi yang baru. Bagian pertama kode tersebut memberikan tampilan agregat data berdasarkan wilayah, yang berguna untuk pembahasan regional. Bagian kedua, PCA, mengurangi dimensi data, memungkinkan visualisasi yang lebih mudah dan berpotensi mengungkapkan pola atau hubungan intrinsik yang tidak terlihat dalam ruang berdimensi tinggi.

BAB 3

NAVIGASI DATA

3.1 PENDAHULUAN

Pembahasan deskriptif berfungsi sebagai landasan proses ilmu data pemasaran. Ini melibatkan pemeriksaan cermat terhadap data historis untuk membedakan dan menggambarkan pola, tren, dan hubungan utama. Teknik pembelajaran dasar ini sangat penting bagi pemasar yang ingin memperoleh pemahaman menyeluruh tentang perilaku pelanggan, kondisi pasar, dan efisiensi inisiatif pemasaran. Dengan memanfaatkan pembahasan deskriptif, pemasar dapat menggali wawasan yang mendorong pengambilan keputusan yang tepat dan menyempurnakan strategi pemasaran mereka untuk dampak yang lebih besar.

Bab ini menguraikan berbagai aspek pembahasan deskriptif dalam ranah pemasaran, memadukan teori dengan aplikasi praktis. Kita akan membedah konsep-konsep kunci, mempelajari teknik-teknik yang kuat, dan menampilkan aplikasi penting pembahasan deskriptif dalam pemasaran. Setiap bagian diperkuat dengan contoh-contoh praktis dan alat bantu visual ilustratif, yang menghidupkan prinsip dan teknik yang dibahas.

Keunggulan strategis yang diberikan oleh pembelajaran deskriptif tidak dapat dilebih-lebihkan. Ini adalah landasan analitis yang menjadi dasar bagi pendekatan yang lebih canggih, seperti pembelajaran prediktif dan preskriptif. Tahapan pembelajaran tingkat lanjut ini, yang mengantisipasi peristiwa di masa depan dan merekomendasikan tindakan pemasaran yang optimal, dibangun di atas wawasan yang diperoleh dari pembahasan deskriptif. Penguasaan pembelajaran deskriptif bukan hanya tentang memahami masa lalu—tetapi juga tentang membentuk masa depan. Ini memberdayakan pemasar dengan ketajaman untuk mengidentifikasi area yang siap untuk ditingkatkan, memanfaatkan tren yang berkembang, dan memperkuat efektivitas upaya pemasaran.

Perjalanan kita melalui bab ini akan mencakup eksplorasi statistik deskriptif, yang mengungkapkan kecenderungan sentral, penyebaran, dan bentuk distribusi data Anda. Kita juga akan menjelajahi ranah visualisasi data, mengilustrasikan kekuatan grafik, bagan, dan platform interaktif untuk mengkomunikasikan informasi kompleks secara jelas dan efektif. Selanjutnya, kita akan mengpembahasan secara kritis kinerja kampanye pemasaran, mempelajari cara mengukur dan menafsirkan keberhasilannya secara akurat.

Kita akan mendalami penerapan konsep-konsep ini di dunia nyata, mempelajari bukan hanya apa yang harus dipelajari tetapi juga bagaimana menerapkan teknik-teknik ini dalam proyek ilmu data pemasaran. Dengan memahami berbagai aspek pembelajaran deskriptif, Anda akan diperlengkapi untuk mengubah data mentah menjadi wawasan yang dapat ditindaklanjuti yang mendorong strategi pemasaran Anda ke depan. Bab ini berfungsi sebagai panduan komprehensif melalui lanskap pembelajaran deskriptif yang dinamis, memberikan tidak hanya pemahaman dasar tetapi juga wawasan praktis tentang bagaimana teknik-teknik ini dapat diterapkan secara efektif dalam proyek ilmu data pemasaran. Dengan mempelajari

peran pembelajaran deskriptif dalam pemasaran, Anda akan belajar mengubah data mentah menjadi wawasan yang dapat ditindaklanjuti, sehingga meletakkan dasar yang kuat untuk topik pembelajaran yang lebih lanjut yang akan dieksplorasi dalam bab-bab selanjutnya dari buku ini.

3.2 PEMAHAMAN KOMPREHENSIF TENTANG DATA HISTORIS

Peran Pembelajaran Deskriptif dalam Pemasaran

Pembelajaran deskriptif memainkan peran penting dalam proses ilmu data pemasaran dengan memberikan pemahaman komprehensif tentang data historis. Hal ini memungkinkan pemasar untuk meringkas, memvisualisasikan, dan menafsirkan pola, tren, dan hubungan utama dalam data mereka, meletakkan dasar untuk teknik pembelajaran yang lebih canggih, seperti pembelajaran prediktif dan preskriptif. Pembelajaran deskriptif sangat penting karena alasan berikut:

- Memahami perilaku dan preferensi pelanggan
- Mengidentifikasi tren pasar dan dinamika persaingan
- Mengevaluasi efektivitas kampanye dan strategi pemasaran
- Memberikan informasi untuk pengambilan keputusan berbasis data dalam pemasaran

Dengan memberikan gambaran yang jelas tentang data historis, pembelajaran deskriptif memungkinkan pemasar untuk mendapatkan wawasan tentang faktor-faktor yang mendorong keterlibatan pelanggan, loyalitas, dan penjualan, sehingga mereka dapat membuat keputusan yang lebih tepat dan mengoptimalkan upaya pemasaran mereka. Saat kita mengeksplorasi dampak pembelajaran deskriptif pada pemasaran, akan sangat bermanfaat untuk membandingkannya dengan jenis pembelajaran lain—prediktif dan preskriptif—yang juga sangat penting dalam domain ini. Tabel 3.1 menawarkan perbandingan komprehensif dari ketiga jenis pembelajaran ini: deskriptif, prediktif, dan preskriptif. Tabel ini menguraikan fitur-fitur utama, penggunaan dalam pemasaran, dan peran berbeda yang mereka mainkan.

Tabel ini tidak hanya memperjelas definisi dan tujuan dari setiap jenis, tetapi juga merinci metode dan teknik yang terlibat, beserta hasil dan kasus penggunaan spesifik dalam pemasaran. Perbandingan seperti ini memberikan kerangka kerja yang jelas untuk memahami di mana pembelajaran deskriptif berada dalam spektrum pembelajaran yang lebih luas dan bagaimana ia melengkapi pendekatan pembelajaran lainnya dalam mendorong strategi pemasaran yang efektif.

Tabel 3.1 Berbagai Jenis Pembelajaran—Deskriptif, Prediktif, dan Preskriptif—dengan Fitur Utama dan Penggunaannya dalam Pemasaran.

Dimensi	Pembahasan Deskriptif	Pembahasan Prediktif	Pembahasan Preskriptif
Definisi	Mengpembahasan data historis untuk memahami apa yang telah terjadi	Menggunakan data historis untuk memprediksi hasil di masa depan	Merekomendasikan tindakan berdasarkan pembahasan untuk mencapai hasil yang diinginkan

Tujuan	Untuk meringkas dan memahami perilaku masa lalu	Untuk meramalkan peristiwa atau perilaku di masa mendatang	Memberikan saran tentang potensi tindakan di masa mendatang
Metode/teknik	Statistik Agregasi data Visualisasi	Pembelajaran Mesin Regresi Pembahasan Deret Waktu	Simulasi Optimasi Pohon keputusan
Hasil	Laporan Dasbor Grafik	Skor Probabilitas Prediksi	Rekomendasi Model keputusan
Studi kasus dalam pemasaran	Pembahasan tren pasar Segmentasi pelanggan Pembahasan kinerja penjualan	Peramalan penjualan Prediksi kehilangan pelanggan Penilaian prospek	Optimasi harga Penargetan kampanye Manajemen inventaris

Teknik-Teknik Utama dalam Pembahasan Deskriptif

Pembahasan deskriptif menggunakan berbagai teknik untuk meng pembahasan dan meringkas data pemasaran:

- **Statistik deskriptif.** Ini adalah ukuran matematis yang digunakan untuk meringkas dan menggambarkan karakteristik utama dari suatu kumpulan data, seperti ukuran kecenderungan sentral (rata-rata, median, modus), ukuran dispersi (rentang, varians, deviasi standar), dan ukuran asosiasi (korelasi, kovarians).
- **Visualisasi data.** Ini melibatkan representasi grafis data menggunakan grafik, plot, dan peta untuk membantu pemasar dengan cepat mengidentifikasi pola, tren, dan hubungan dalam data mereka. Teknik visualisasi data umum meliputi grafik batang, grafik garis, scatterplot, dan peta panas.
- **Agregasi data.** Ini adalah proses meringkas data pada tingkat granularitas yang lebih tinggi, seperti mengagregasi data penjualan harian ke tingkat bulanan atau data tingkat pelanggan ke tingkat segmen. Agregasi data membantu pemasar menyederhanakan data mereka dan fokus pada tren dan pola yang paling penting.

Pentingnya Visualisasi Data

Visualisasi data merupakan komponen penting dari pembelajaran deskriptif, karena memungkinkan pemasar untuk mengkomunikasikan data kompleks dengan cepat dan efektif dalam format yang menarik secara visual dan mudah dicerna. Dengan menggunakan teknik visualisasi data, pemasar dapat mencapai hal-hal berikut:

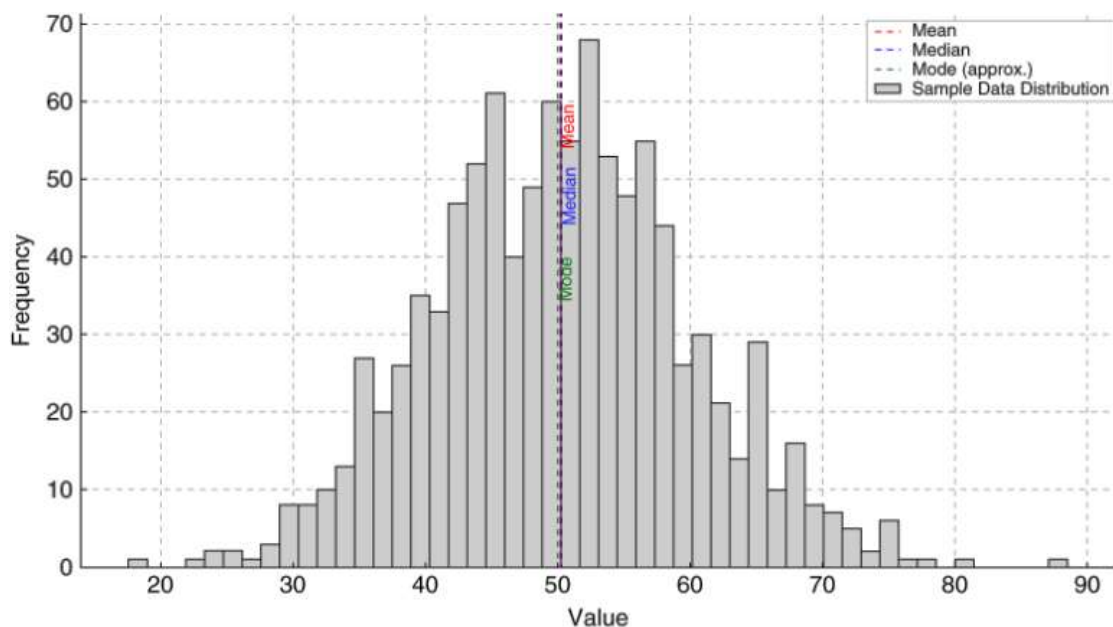
- Mengidentifikasi pola dan tren dalam data mereka dengan lebih mudah dan intuitif.
- Membandingkan berbagai titik data, variabel, atau segmen dengan lebih efektif.
- Mengkomunikasikan temuan dan wawasan mereka kepada pemangku kepentingan dengan cara yang menarik.
- Memfasilitasi pengambilan keputusan berbasis data dengan membuat data kompleks lebih mudah diakses dan ditindaklanjuti.

Pada bagian selanjutnya, kita akan membahas lebih dalam teknik dan aplikasi spesifik pembelajaran deskriptif dalam pemasaran, termasuk statistik deskriptif, visualisasi data, dan

pembahasan kinerja kampanye pemasaran. Melalui contoh praktis dan ilustrasi yang relevan, kita akan menunjukkan nilai nyata pembelajaran deskriptif dalam mendorong strategi dan keputusan pemasaran berbasis data.

3.3 PARAMETER DESKRIPTIF METRIK PEMASARAN

Statistik deskriptif memberikan ringkasan karakteristik utama dari suatu dataset, membantu pemasar memahami kecenderungan sentral data, dispersi, dan asosiasi antar variabel. Ukuran-ukuran ini berfungsi sebagai titik awal untuk pembahasan data yang lebih lanjut dan sangat penting untuk membuat keputusan yang tepat dalam pemasaran.



Gambar 3.1 Berbagai Ukuran Kecenderungan Sentral (Rata-rata, Median, Modus) pada Distribusi Data Sampel.

Ukuran Kecenderungan Sentral

Ukuran kecenderungan sentral menggambarkan nilai sentral atau tipikal dalam suatu dataset. Ukuran kecenderungan sentral yang paling umum adalah mean, median, dan modus (lihat Gambar 3.1):

- **Mean.** Rata-rata aritmatika dari semua titik data
- **Median.** Nilai tengah dalam suatu dataset ketika diurutkan dari terkecil hingga terbesar
- **Modus.** Nilai yang paling sering muncul dalam suatu dataset

Memahami kecenderungan sentral data pemasaran, seperti demografi pelanggan atau metrik penjualan, membantu pemasar mengidentifikasi pelanggan atau kinerja penjualan yang "khas" dan membuat keputusan yang lebih tepat.

Ukuran Dispersi

Ukuran dispersi menggambarkan penyebaran atau variabilitas titik data dalam suatu dataset. Ukuran dispersi yang paling umum adalah rentang, varians, dan deviasi standar:

- **Rentang.** Perbedaan antara nilai tertinggi dan terendah dalam suatu dataset

- **Varians.** Rata-rata selisih kuadrat antara setiap titik data dan rata-rata
- **Deviasi standar.** Akar kuadrat dari varians, yang mewakili jarak rata-rata setiap titik data dari rata-rata

Memahami dispersi data pemasaran membantu pemasar mengevaluasi variabilitas dan risiko dalam kampanye dan strategi pemasaran mereka.

Ukuran Asosiasi

Ukuran asosiasi menggambarkan hubungan antara dua atau lebih variabel dalam suatu dataset. Ukuran asosiasi yang paling umum adalah korelasi dan kovarians:

- **Korelasi.** Ukuran standar dari hubungan linier antara dua variabel, berkisar dari -1 (hubungan negatif sempurna) hingga $+1$ (hubungan positif sempurna)
- **Kovarians.** Ukuran variabilitas gabungan dari dua variabel, yang menunjukkan arah hubungan mereka

Memahami asosiasi antara variabel pemasaran, seperti perilaku pelanggan dan kinerja penjualan, membantu pemasar mengidentifikasi pendorong utama keterlibatan pelanggan, loyalitas, dan penjualan, yang menginformasikan strategi dan taktik pemasaran mereka.

Tabel 3.2 menyajikan statistik deskriptif utama untuk kumpulan data pemasaran, yang menggambarkan korelasi yang kuat, hampir satu banding satu, antara pengeluaran iklan dan klik, dengan keduanya menunjukkan variabilitas yang signifikan. Sebaliknya, klik menunjukkan korelasi positif sedang dengan penjualan, yang lebih konsisten dengan variasi yang jauh lebih kecil, menunjukkan faktor lain dapat memengaruhi konversi klik menjadi penjualan. Tabel 3.2 menggarisbawahi pentingnya memahami distribusi dan hubungan antara berbagai metrik pemasaran.

Singkatnya, statistik deskriptif memainkan peran penting dalam memahami dan menafsirkan data pemasaran, memberikan dasar untuk pembahasan data dan pengambilan keputusan yang lebih maju. Dengan menguasai ukuran-ukuran ini, pemasar dapat memperoleh wawasan berharga tentang data mereka dan membuat keputusan yang lebih tepat dalam upaya pemasaran mereka.

Tabel 3.2 Contoh Kumpulan Data Pemasaran dengan Perhitungan untuk Ukuran Kecenderungan Sentral, Dispersi, dan Asosiasi.

Variabel	Rata-rata (Mean)	Median	Modus	Rentang (Range)	Varians	Standar Deviasi	Korelasi dengan Variabel Berikutnya
Biaya Iklan (Ad Spend)	2880,72	2856,57	1022,09	3925,46	1415999,00	1189,96	0,98
Klik (Clicks)	145,35	147,50	188,00	236,00	3817,06	61,78	0,55
Penjualan (Sales)	2,85	2,00	2,00	11,00	4,82	2,19	0,53

Simetri dan Kemiringan

Memahami distribusi data Anda sangat penting dalam pembahasan data. Dua aspek penting dari hal ini adalah simetri dan kemiringan, yang memberikan wawasan tentang sifat dan bentuk distribusi.

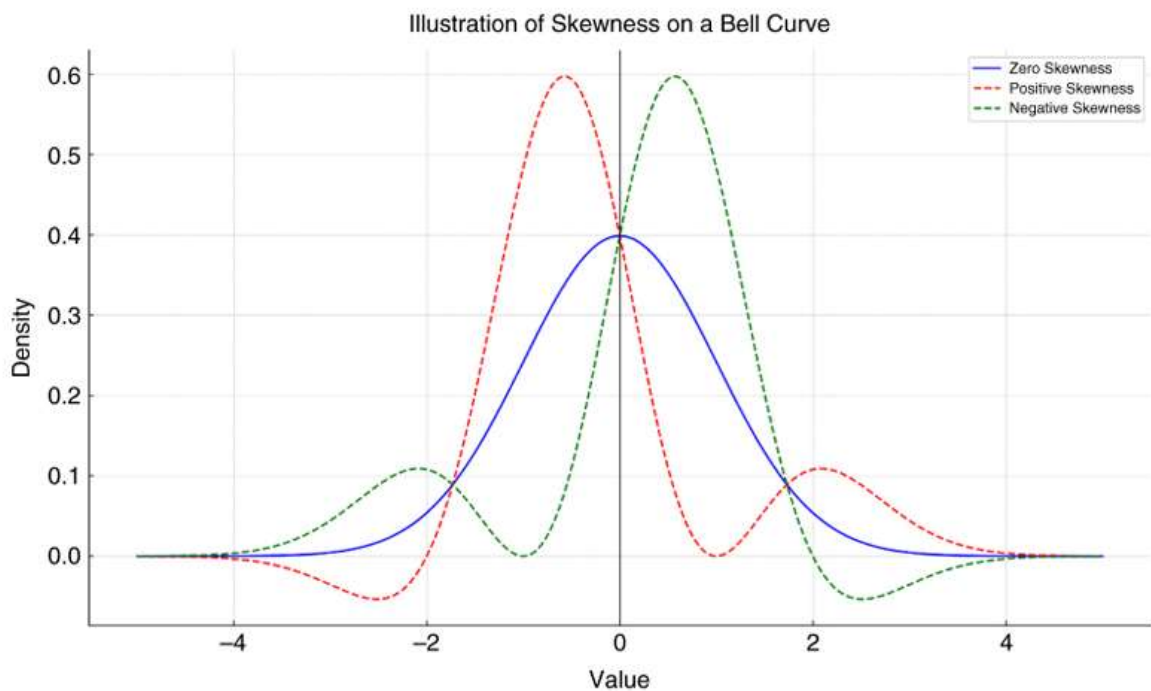
Simetri

Simetri mengacu pada situasi di mana setengah dari data merupakan bayangan cermin dari setengah lainnya. Misalnya, pertimbangkan tinggi badan orang dewasa dalam populasi besar. Jika sebagian besar orang memiliki tinggi badan rata-rata (dengan jumlah yang sama lebih tinggi atau lebih pendek dari rata-rata), distribusi tinggi badan akan simetris. Distribusi yang simetris sempurna akan memiliki nilai rata-rata, median, dan modus yang sama.

Kemiringan

Kemiringan mengukur derajat asimetri dalam distribusi data. Suatu distribusi dapat miring negatif (miring ke kiri), miring positif (miring ke kanan), atau tidak memiliki kemiringan (atau nol) (simetris) (lihat Gambar 3.2).

- **Kemiringan positif.** Kemiringan positif terjadi ketika ekor di sisi kanan (ujung nilai yang lebih tinggi) lebih panjang daripada ekor kiri, menunjukkan bahwa data memiliki lebih banyak outlier di sebelah kanan. Pertimbangkan distribusi kekayaan dalam suatu masyarakat. Beberapa individu mungkin memiliki kekayaan bersih yang sangat tinggi, menarik rata-rata ke kanan, tetapi mayoritas akan terkumpul di sekitar nilai rata-rata yang lebih rendah.
- **Kemiringan negatif.** Kemiringan negatif terjadi ketika ekor di sisi kiri (ujung nilai yang lebih rendah) lebih panjang daripada ekor kanan, menunjukkan bahwa data memiliki lebih banyak outlier di sebelah kiri. Pikirkan tentang waktu yang dibutuhkan untuk lari maraton. Beberapa atlet profesional mungkin menyelesaikan perlombaan dengan sangat cepat, tetapi sebagian besar peserta akan menyelesaikannya dalam waktu rata-rata yang lebih lama, yang menyebabkan distribusi miring ke kiri.
- **Kemiringan nol.** Kemiringan nol menunjukkan distribusi yang simetris sempurna.



Gambar 3.2 Kemiringan Positif, Negatif, dan Nol.

Implikasi

Kemiringan dalam distribusi data memiliki implikasi yang luas terhadap akurasi analitis dan interpretasi data pemasaran.

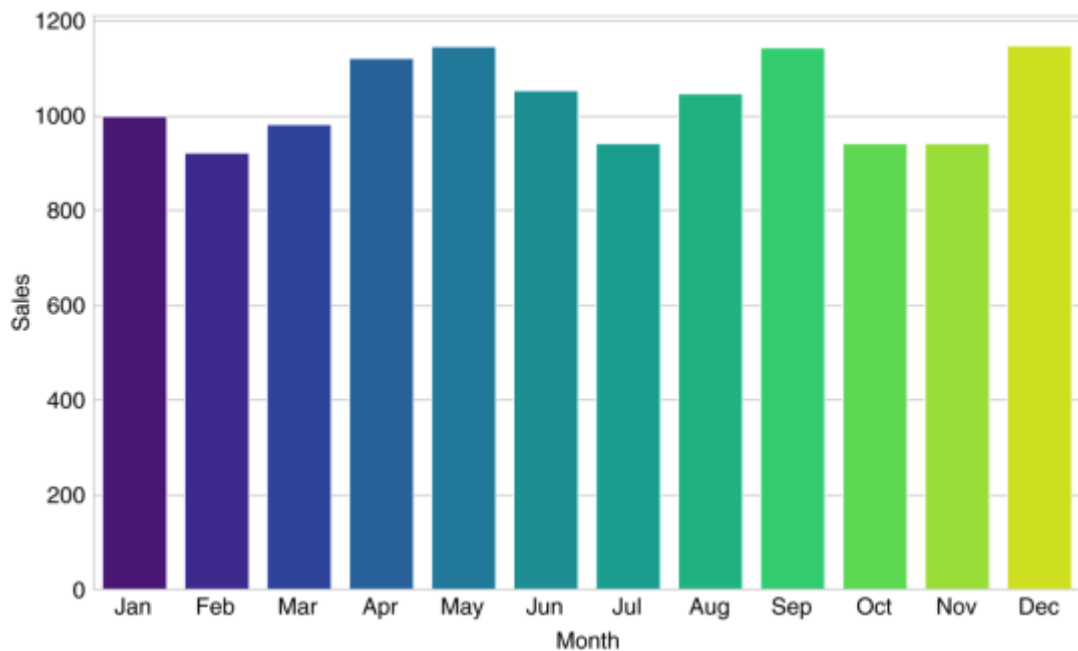
- **Pembahasan statistik.** Tingkat dan arah kemiringan dapat memengaruhi pilihan dan hasil uji statistik. Beberapa uji mengasumsikan data terdistribusi normal, dan kemiringan yang signifikan dapat melanggar asumsi ini.
- **Transformasi data.** Dalam situasi di mana kemiringan mungkin bermasalah, berbagai teknik transformasi data, seperti transformasi logaritmik atau akar kuadrat, dapat diterapkan untuk membuat distribusi lebih simetris.
- **Statistik deskriptif.** Tergantung pada kemiringan, rata-rata mungkin bukan ukuran kecenderungan sentral yang paling representatif. Dalam distribusi miring, median mungkin menawarkan nilai sentral yang lebih baik.

Memahami simetri dan kemiringan suatu dataset menawarkan wawasan tentang strukturnya, memandu pendekatan analitis yang tepat dan membantu dalam menafsirkan hasil secara efektif.

3.4 TEKNIK VISUALISASI DATA

Visualisasi data memainkan peran penting dalam pembelajaran deskriptif, memungkinkan pemasar untuk dengan cepat dan efektif mengkomunikasikan data kompleks dalam format yang menarik secara visual dan mudah dicerna. Dengan menggunakan teknik visualisasi data, pemasar dapat mengidentifikasi pola dan tren dalam data mereka dengan lebih mudah dan intuitif, membandingkan berbagai titik data atau segmen dengan lebih efektif, dan mengkomunikasikan temuan mereka kepada pemangku kepentingan dengan cara yang

menarik. Pada bagian ini, kita akan membahas beberapa teknik visualisasi data yang paling umum digunakan dalam pembelajaran pemasaran.



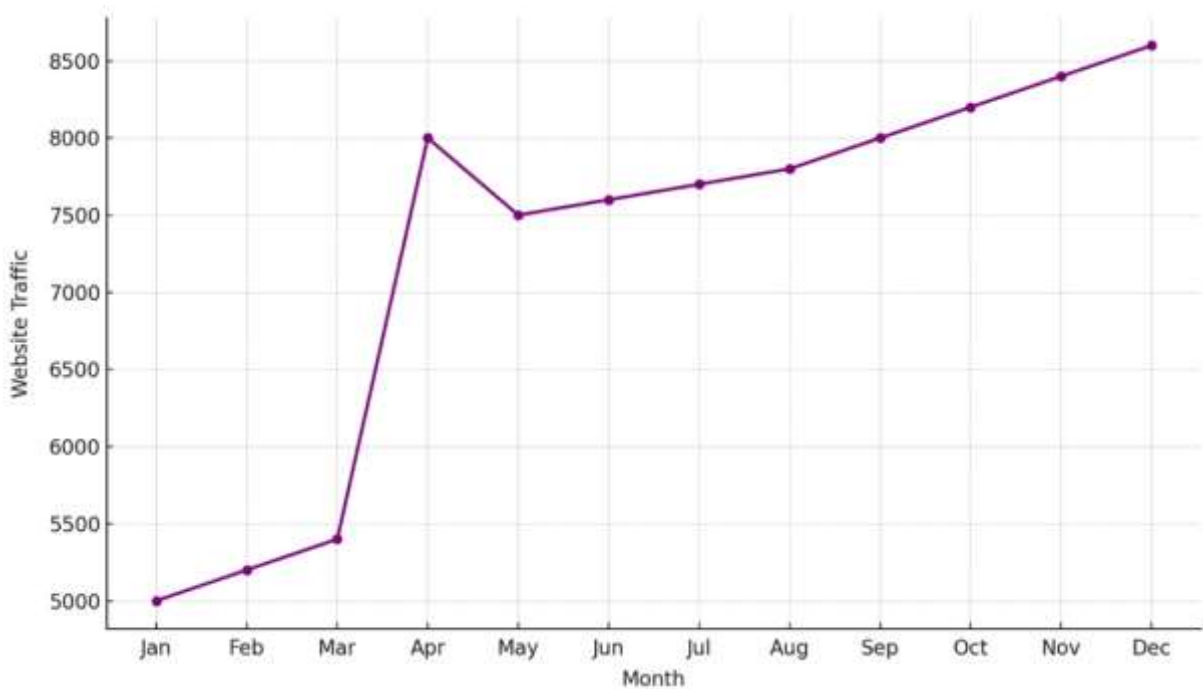
Gambar 3.3 Distribusi Frekuensi Penjualan Bulanan Suatu Produk.

Bagan Batang dan Histogram

Bagan batang dan histogram banyak digunakan dalam pembelajaran pemasaran untuk menampilkan distribusi variabel kategorikal atau frekuensi kejadian dalam interval variabel kontinu, masing-masing. Bagan batang merepresentasikan data menggunakan batang persegi panjang, dengan tinggi atau panjang setiap batang proporsional dengan nilai yang diwakilinya. Namun, histogram menggunakan batang yang berdekatan untuk merepresentasikan frekuensi titik data dalam interval tertentu. Visualisasi ini sangat berguna untuk membandingkan berbagai kategori atau segmen dalam data pemasaran, seperti demografi pelanggan, kategori produk, atau wilayah geografis (lihat Gambar 3.3).

Grafik Garis dan Plot Deret Waktu

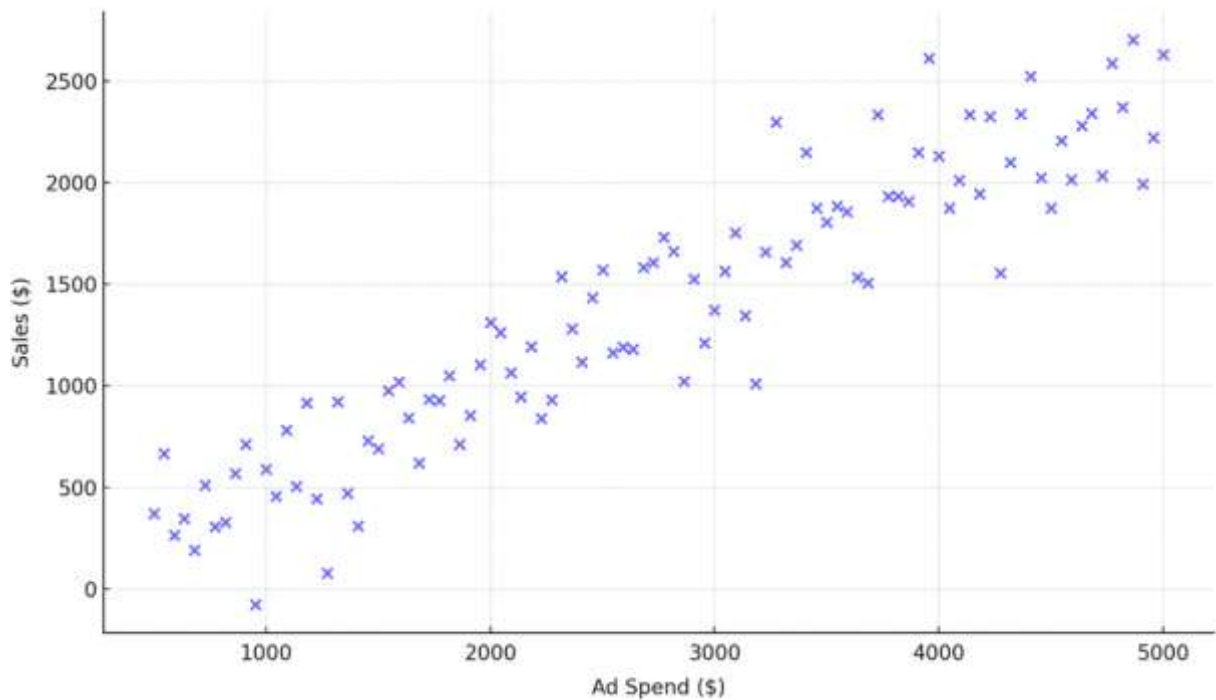
Grafik garis dan plot deret waktu digunakan untuk menampilkan tren data dari waktu ke waktu, membantu pemasar mengidentifikasi pola, musiman, dan fluktuasi dalam data mereka. Dengan menghubungkan titik data dengan garis, grafik ini memungkinkan pemasar untuk memvisualisasikan perubahan variabel dari waktu ke waktu, sehingga lebih mudah untuk melihat tren dan membuat perkiraan. Plot deret waktu sangat berguna dalam pemasaran untuk melacak indikator kinerja utama (KPI), seperti penjualan, lalu lintas situs web, atau akuisisi pelanggan, dari waktu ke waktu (lihat Gambar 3.4).



Gambar 3.4 Lalu Lintas Situs Web Bulanan Selama Satu Tahun.

Diagram Sebar dan Bagan Gelembung

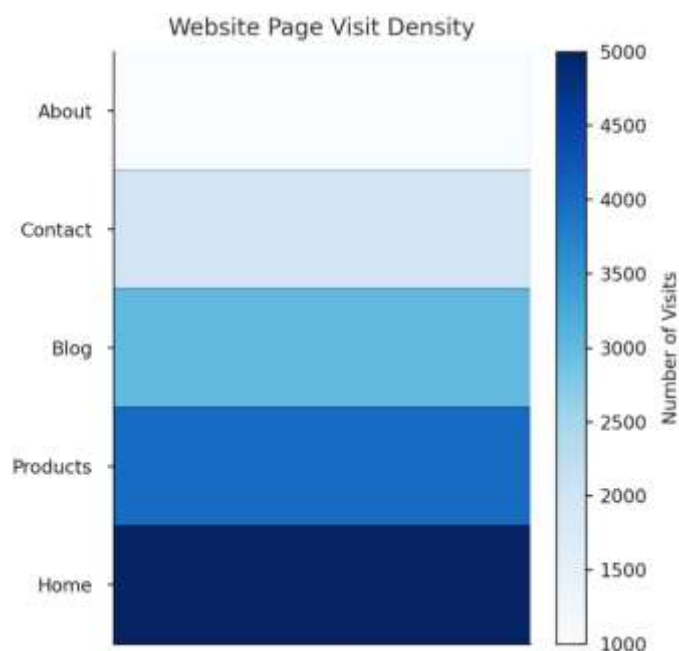
Diagram sebar dan bagan gelembung digunakan untuk menampilkan hubungan antara dua atau lebih variabel dalam suatu dataset. Diagram sebar merepresentasikan titik data sebagai titik-titik individual pada bidang Kartesius, dengan posisi setiap titik ditentukan oleh nilai dari dua variabel. Bagan gelembung memperluas diagram sebar dengan menambahkan dimensi ketiga, yang direpresentasikan oleh ukuran titik data. Visualisasi ini berguna untuk mengeksplorasi korelasi, tren, dan kluster dalam data pemasaran, seperti hubungan antara demografi pelanggan dan kinerja penjualan atau hubungan antara pengeluaran pemasaran dan pendapatan (lihat Gambar 3.5).



Gambar 3.5 Korelasi Antara Pengeluaran Iklan dan Penjualan untuk Suatu Produk.

Peta Panas dan Peta Geografis

Peta panas dan peta geografis digunakan untuk memvisualisasikan data dalam konteks spasial, membantu pemasar memahami pola dan tren geografis dalam data mereka. Peta panas menggunakan gradien warna untuk merepresentasikan kepadatan atau intensitas titik data dalam area tertentu, sedangkan peta geografis menampilkan data pada peta, seringkali menggunakan kode warna, simbol, atau simbol proporsional untuk merepresentasikan nilai data. Visualisasi ini sangat berguna untuk meng pembahasan distribusi spasial pelanggan, kinerja penjualan, atau efektivitas kampanye pemasaran di berbagai wilayah atau pasar (lihat Gambar 3.6).



Gambar 3.6 Kepadatan Kunjungan Halaman Situs Web.

Praktik Terbaik Visualisasi

Visualisasi data adalah seni sekaligus sains. Jika dieksekusi secara efektif, visualisasi dapat mengubah kumpulan data yang kompleks menjadi visual yang intuitif dan berwawasan yang mengkomunikasikan pesan yang jelas. Namun, praktik visualisasi yang buruk dapat menyebabkan kebingungan atau bahkan salah tafsir. Berikut adalah beberapa praktik terbaik untuk memastikan visualisasi Anda menyampaikan pesan yang dimaksud secara akurat dan efektif.

Kesederhanaan Adalah Kunci

Selalu mulai dengan tujuan yang jelas. Cerita apa yang ingin Anda sampaikan melalui data Anda? Setiap visualisasi harus menyampaikan satu pesan atau wawasan utama. Hilangkan kekacauan; hapus elemen yang tidak perlu seperti garis kisi, label, atau dekorasi yang berlebihan. Bagan yang lebih bersih seringkali merupakan bagan yang lebih efektif. Prinsip Tufte tentang memaksimalkan “rasio data-tinta” menekankan pengurangan tinta non-data.

Pilih Jenis Bagan yang Tepat

Jenis bagan yang berbeda menekankan jenis data dan hubungan yang berbeda. Misalnya, bagan batang efektif untuk membandingkan kuantitas diskrit, dan bagan garis ideal untuk menunjukkan tren dari waktu ke waktu. Bagan pai, misalnya, dapat menjadi masalah ketika membandingkan banyak kategori, dan bagan 3D dapat mendistorsi persepsi nilai. Pastikan pilihan Anda tidak secara tidak sengaja menyesatkan atau membingungkan audiens Anda.

Gunakan Warna dengan Bijaksana

Warna harus memiliki tujuan. Gunakan warna untuk membedakan item, menyoroti titik data tertentu, atau menunjukkan kategori. Hindari menggunakan terlalu banyak warna, karena dapat secara visual membingungkan. Sekitar 8% pria dan 0,5% wanita keturunan Eropa Utara terpengaruh oleh kekurangan penglihatan warna. Alat seperti Color Oracle dapat membantu memastikan visualisasi Anda dapat dipahami oleh semua orang.

Pertahankan Konsistensi

Saat membandingkan beberapa grafik, pastikan skala konsisten sehingga perbandingan valid. Pastikan simbol, warna, dan terminologi tetap konsisten di semua visualisasi.

Prioritaskan Integritas Data

Pastikan representasi visual titik data proporsional dengan nilai data aktual. Memotong sumbu y dapat melebih-lebihkan perbedaan dan dapat menyesatkan. Jika Anda harus memotong, selalu beri label sumbu dengan jelas untuk menunjukkan bahwa sumbu tersebut tidak dimulai dari nol.

Visualisasi Interaktif

Alat seperti SAS Visual Analytics atau D3.js memungkinkan pembuatan visualisasi interaktif, membantu pengguna untuk menyelami lebih dalam bagian-bagian spesifik dari data, memberikan gambaran umum yang luas dan wawasan yang detail. Saat menggunakan

alat interaktif, pandu pengguna melalui cerita data dengan memberikan petunjuk atau jalur yang terfokus.

Uji dan Ulangi

Bagikan visualisasi Anda dengan kelompok yang beragam, kumpulkan umpan balik, dan bersiaplah untuk menyesuaikan. Visualisasi yang tampak jelas bagi Anda mungkin membingungkan bagi orang lain. Seperti semua keterampilan, kemahiran Anda dalam visualisasi data akan meningkat dengan latihan dan pembelajaran berkelanjutan. Tetaplah mengikuti perkembangan teknik, alat, dan praktik terbaik terbaru.

Kesimpulannya, visualisasi yang efektif adalah alat yang ampuh bagi pemasar, memungkinkan mereka untuk menyampaikan data kompleks dalam format yang mudah dicerna. Dengan mengikuti praktik terbaik ini, pemasar dapat memastikan visualisasi mereka tidak hanya terlihat bagus tetapi juga mengkomunikasikan pesan mereka secara efektif, yang mengarah pada pengambilan keputusan yang lebih baik dan presentasi yang lebih berdampak.

3.5 PEMBAHASAN DATA EKSPLORASI DALAM PEMASARAN

Pembahasan data eksplorasi (EDA) adalah langkah fundamental dalam proses ilmu data pemasaran, bertindak sebagai jembatan antara data mentah dan teknik pembelajaran yang lebih rumit. Ini melibatkan pemeriksaan komprehensif dan sistematis terhadap kumpulan data untuk menemukan pola, menemukan anomali, menguji hipotesis, dan memahami struktur data—semuanya dengan bantuan metode visual dan statistik deskriptif. Bagi pemasar, EDA sangat penting karena menawarkan pandangan awal terhadap data, membantu mengungkap wawasan, menentukan alat pembelajaran yang tepat untuk digunakan, dan membentuk strategi pemodelan data.

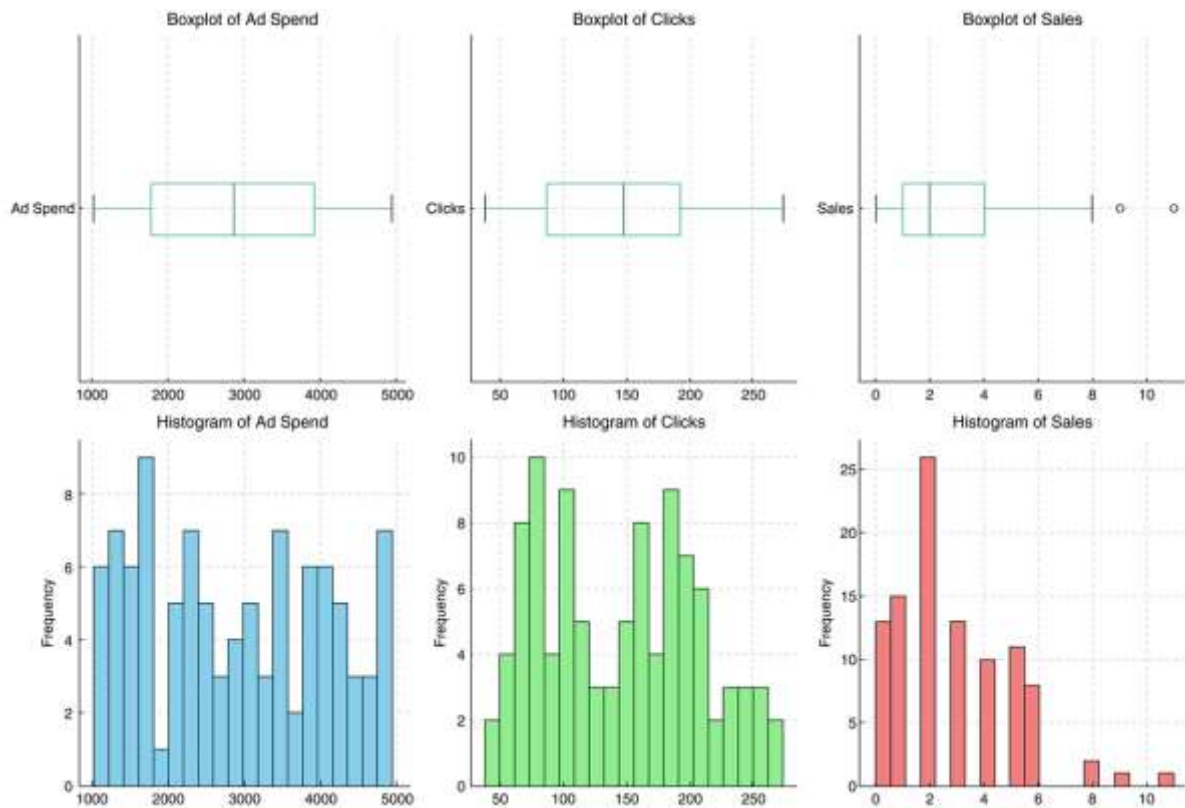
Pembahasan Distribusi Data

Sebelum mempelajari metode pembelajaran tingkat lanjut, sangat penting bagi pemasar untuk terlebih dahulu memahami distribusi data mereka.

- **Distribusi univariat.** Ini melihat satu variabel pada satu waktu. Alat seperti histogram, box plot, dan density plot dapat membantu memvisualisasikan distribusi, kecenderungan sentral, dan penyebaran data. Misalnya, histogram dapat mengungkapkan bagaimana usia pelanggan didistribusikan di berbagai kelompok usia, memberikan gambaran tentang demografi basis pelanggan.
- **Distribusi multivariat.** Ketika pemasar ingin mengamati interaksi antara beberapa variabel, teknik seperti scatterplot berperan. Misalnya, scatterplot dapat menggambarkan bagaimana lalu lintas situs web (satu variabel) berhubungan dengan penjualan (variabel lain).

Memahami distribusi data adalah langkah mendasar dalam setiap proses analitis, dan alat visual memainkan peran kunci dalam pemahaman ini. Gambar 3.7 menampilkan box plot dan histogram dari contoh dataset pemasaran, yang menggambarkan bagaimana alat-alat ini dapat secara jelas menggambarkan distribusi data. Box plot memberikan wawasan tentang kecenderungan sentral dan penyebaran data, dan histogram secara efektif menunjukkan distribusi frekuensi suatu variabel, seperti kelompok usia pelanggan. Dengan memeriksa

representasi visual ini, pemasar dapat memperoleh pemahaman yang lebih jelas tentang sifat data mereka, mengidentifikasi outlier, dan menilai kebutuhan akan teknik transformasi data apa pun. Memahami distribusi data membantu dalam menentukan sifat data, menemukan outlier, dan memutuskan apakah teknik transformasi data (seperti normalisasi atau standarisasi) diperlukan.

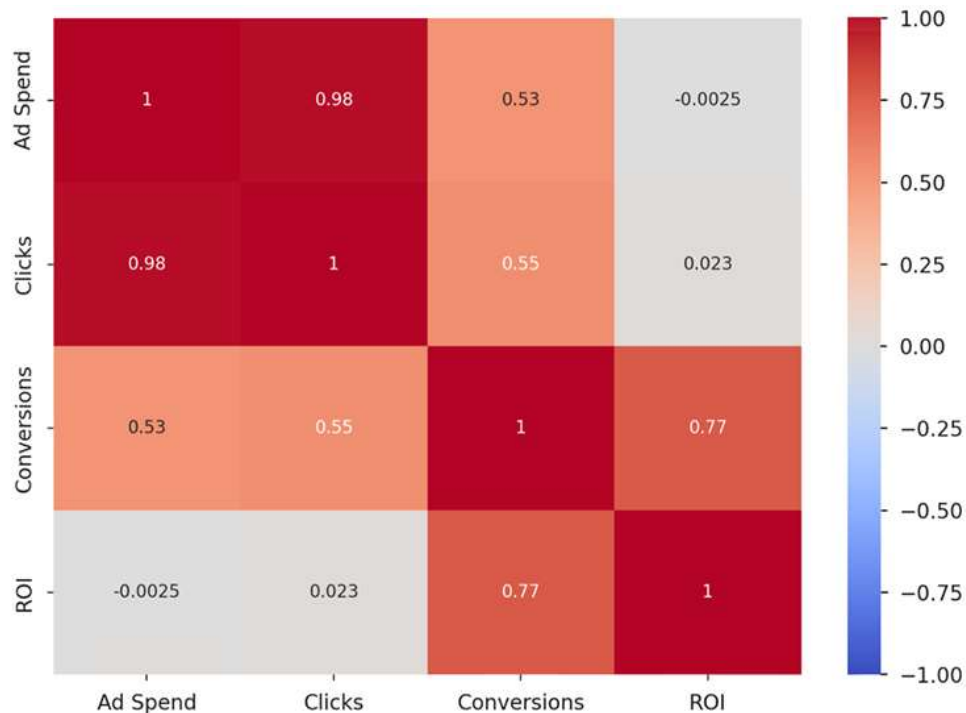


Gambar 3.7 Contoh Dataset Pemasaran yang Menunjukkan Distribusi Data.

Korelasi dan Kovarians

Untuk memahami hubungan antar variabel, pemasar menggunakan ukuran asosiasi seperti korelasi dan kovarians:

- **Korelasi.** Ini adalah ukuran standar, yang memberikan kekuatan dan arah hubungan linier antara dua variabel. Korelasi 1 menunjukkan hubungan linier positif sempurna, -1 hubungan negatif sempurna, dan 0 tidak ada hubungan linier. Misalnya, mungkin ada korelasi positif yang kuat antara pengeluaran iklan dan penjualan, yang menunjukkan bahwa ketika satu meningkat, yang lain cenderung meningkat juga.
- **Kovarians.** Meskipun juga menunjukkan arah suatu hubungan, kovarians tidak terstandarisasi. Dengan demikian, meskipun dapat menunjukkan arah hubungan linier (positif atau negatif), kovarians sendiri tidak dapat mengukur kekuatan hubungan tersebut.



Gambar 3.8 Hubungan Antar Berbagai Variabel Pemasaran.

Untuk lebih menjelaskan konsep korelasi dan kovarians dalam ranah data pemasaran, Gambar 3.8 menyajikan matriks korelasi yang secara visual menggambarkan hubungan antar berbagai variabel pemasaran. Matriks ini tidak hanya menunjukkan arah tetapi juga kekuatan hubungan linier antara pasangan variabel, menawarkan pandangan yang jelas dan komprehensif yang dapat lebih informatif daripada nilai numerik saja. Visualisasi semacam ini sangat berguna bagi pemasar dalam mengidentifikasi potensi hubungan, membantu dalam pemilihan fitur, mendeteksi multikolinearitas, dan menyempurnakan model prediktif. Dengan memberikan gambaran umum tentang bagaimana berbagai variabel pemasaran saling berhubungan, gambar ini mencontohkan penerapan praktis korelasi dan kovarians dalam strategi pemasaran berbasis data.

Kedua ukuran ini sangat penting bagi pemasar karena membantu dalam pemilihan fitur, deteksi multikolinearitas, dan penyempurnaan model dalam pembahasan selanjutnya.

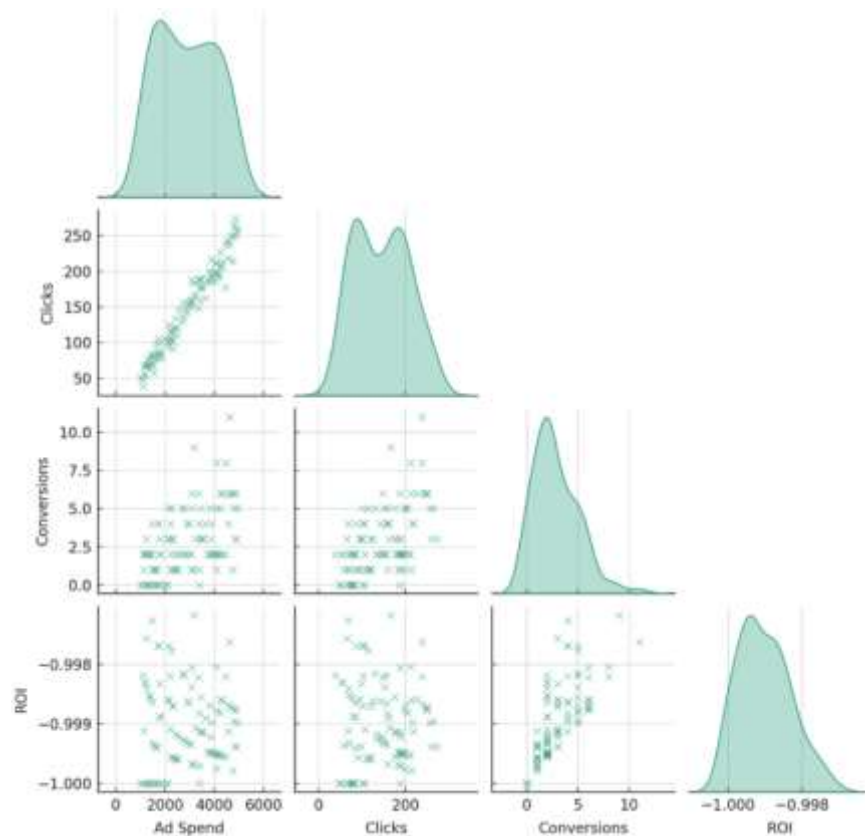
Teknik Eksplorasi Visual

Visualisasi adalah inti dari EDA. Ini menerjemahkan kumpulan data yang rumit menjadi visual yang mudah dipahami, membantu dalam interpretasi data dan pengambilan keputusan (lihat Gambar 3.9).

- **Histogram dan box plot.** Meskipun histogram menggambarkan distribusi frekuensi, box plot memberikan gambaran sekilas tentang kuartil suatu dataset, menunjukkan sebaran data dan outlier.
- **Scatterplot.** Ideal untuk melihat tren, scatterplot secara visual menggambarkan hubungan antara dua variabel. Scatterplot sangat membantu dalam memeriksa potensi kausalitas atau melihat kluster.

- **Heat map.** Ini sangat bagus untuk memahami korelasi antara beberapa variabel sekaligus. Misalnya, seorang pemasar dapat menggunakan heat map untuk melihat bagaimana berbagai metrik pemasaran digital berkorelasi satu sama lain.
- **Pair plot.** Teknik ini menawarkan pandangan holistik dengan memplot hubungan berpasangan dalam suatu dataset. Dengan menggunakan pair plot, pemasar dapat langsung mengamati distribusi dan hubungan antara beberapa variabel.

Dalam konteks pemasaran, teknik eksplorasi visual dapat memberikan wawasan tentang pertanyaan-pertanyaan seperti: Saluran pemasaran mana yang paling kuat berkorelasi dengan nilai seumur hidup pelanggan yang tinggi? Apakah ada pola yang tidak terduga dalam distribusi frekuensi pembelian pelanggan?



Gambar 3.9 Pola Hubungan Antar Beberapa Variabel Pemasaran.

Dalam pemasaran, di mana pemahaman tentang perilaku pelanggan, dinamika pasar, dan hasil kampanye sangat penting, EDA muncul sebagai alat yang sangat berharga. Dengan menggabungkan teknik statistik dan visual, EDA membekali pemasar dengan pandangan holistik tentang data mereka, membuka jalan bagi pembahasan yang lebih canggih dan, pada akhirnya, pengambilan keputusan yang lebih tepat. Baik itu untuk mengidentifikasi segmen pelanggan utama, menentukan saluran pemasaran yang paling efektif, atau memprediksi penjualan di masa mendatang, perjalanan sering kali dimulai dengan pembahasan data eksploratif.

3.6 PARAMETER KEBERHASILAN KAMPANYE BERBASIS DATA

Mengevaluasi kinerja kampanye pemasaran sangat penting bagi pemasar untuk memahami efektivitas strategi, taktik, dan saluran mereka, dan untuk mengoptimalkan upaya pemasaran mereka berdasarkan wawasan berbasis data. Pada bagian ini, kita akan membahas metrik, teknik, dan pendekatan utama yang digunakan untuk meng pembahasan kinerja kampanye pemasaran, serta pentingnya menggunakan pembelajaran deskriptif dalam proses ini.

Indikator Kinerja Utama

KPI adalah ukuran kuantitatif yang digunakan untuk mengevaluasi keberhasilan kampanye pemasaran dalam mencapai tujuannya. KPI dapat bervariasi tergantung pada saluran pemasaran, tujuan kampanye, dan target audiens, tetapi beberapa KPI umum yang digunakan dalam pembelajaran pemasaran meliputi:

- Pendapatan penjualan
- Pengembalian investasi (ROI)
- Biaya akuisisi pelanggan
- Tingkat konversi
- Nilai seumur hidup pelanggan
- Tingkat klik-tayang (CTR)

Memilih KPI yang tepat sangat penting untuk mengukur kinerja kampanye secara akurat dan membuat keputusan berbasis data untuk mengoptimalkan upaya pemasaran (lihat Tabel 3.3).

Tabel 3.3 KPI Umum dalam Pemasaran beserta Definisi dan Metode Perhitungannya.

KPI	Definisi	Metode Perhitungan
Pengembalian investasi	Mengukur profitabilitas suatu kampanye	$\frac{\text{Keuntungan Bersih}}{\text{Biaya Pemasaran}} \times 100$
Biaya per akuisisi	Biaya rata-rata untuk memperoleh pelanggan melalui suatu kampanye	$\frac{\text{Total Biaya Pemasaran}}{\text{Jumlah Akuisisi}}$
Tingkat konversi	Persentase pengunjung yang melakukan tindakan yang diinginkan	$\frac{\text{Jumlah Konversi}}{\text{Jumlah Pengunjung}} \times 100$
Nilai seumur hidup pelanggan	Perkiraan laba bersih dari pelanggan selama masa hidupnya	$\text{Nilai Pembelian Rata – rata} \times \text{Frekuensi Pembelian Rata – rata} \times \text{Rata – rata Masa Hidup Pelanggan}$
Tingkat klik-tayang	Persentase orang yang mengklik iklan setelah melihatnya	$\frac{\text{Jumlah Klik}}{\text{Jumlah Tayang}} \times 100$
Tingkat pentalan	Persentase pengunjung yang meninggalkan situs setelah hanya melihat satu halaman	Dihitung oleh alat pembahasan web seperti Google Analytics.

Teknik untuk Meng pembahasan Kinerja Kampanye Pemasaran

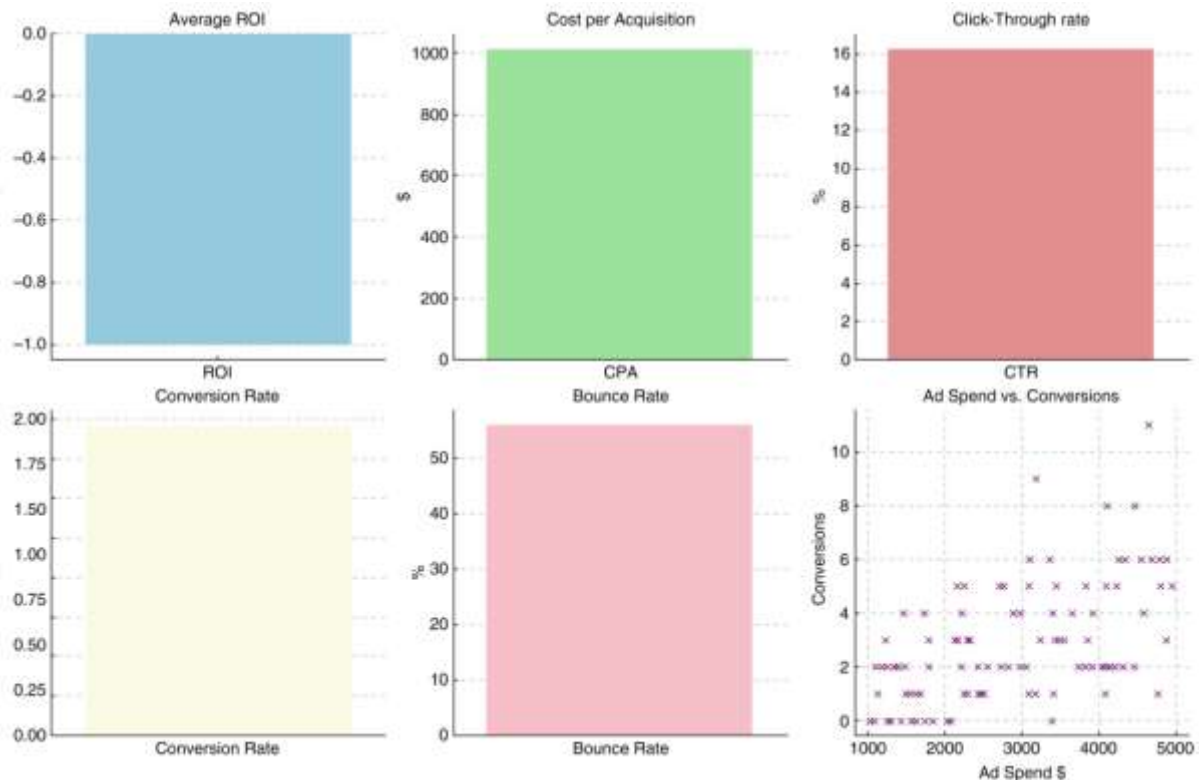
Beberapa teknik dapat digunakan untuk mengpembahasan kinerja kampanye pemasaran, termasuk pembahasan kohort, segmentasi, dan pembahasan peningkatan:

- **Pembahasan kohort.** Pendekatan ini melibatkan pembahasan kinerja berbagai kelompok pelanggan yang memiliki karakteristik umum (misalnya, saluran akuisisi, tanggal pendaftaran) dari waktu ke waktu. Pembahasan kohort membantu pemasar mengidentifikasi tren dan pola dalam perilaku pelanggan, serta efektivitas berbagai kampanye atau saluran pemasaran.
- **Segmentasi.** Segmentasi melibatkan pembagian pelanggan ke dalam kelompok berdasarkan karakteristik yang sama, seperti demografi, preferensi, atau perilaku. Dengan mengpembahasan kinerja berbagai segmen pelanggan, pemasar dapat mengidentifikasi target audiens, menyesuaikan strategi pemasaran mereka, dan mengalokasikan sumber daya secara lebih efektif.
- **Pembahasan Lift.** Pembahasan lift membandingkan kinerja kampanye pemasaran dengan garis dasar atau kelompok kontrol. Teknik ini membantu pemasar menentukan dampak tambahan dari sebuah kampanye pada metrik utama, seperti penjualan atau tingkat konversi, dan menghitung ROI dari upaya pemasaran mereka.

Saat kita menjelajahi berbagai teknik untuk mengpembahasan kinerja kampanye pemasaran, sangat penting untuk memiliki alat yang secara visual merangkum hasilnya dengan cara yang ringkas dan jelas. Gambar 3.10 menyediakan hal itu—visualisasi bergaya dasbor yang menyajikan beberapa KPI dari contoh kampanye pemasaran. Dengan menampilkan berbagai metrik dan wawasan dalam satu visual yang terkonsolidasi, gambar ini mencontohkan bagaimana data kompleks dapat disintesis menjadi informasi yang dapat ditindaklanjuti, membantu pemasar dalam membuat keputusan yang tepat tentang strategi dan investasi mereka.

Peran Pembahasan Deskriptif dalam Pembahasan Kinerja Kampanye Pemasaran

Pembahasan deskriptif memainkan peran penting dalam mengpembahasan kinerja kampanye pemasaran dengan memberikan pemahaman komprehensif tentang data historis, membantu pemasar mengidentifikasi tren, pola, dan hubungan dalam data mereka.



Gambar 3.10 Beberapa KPI dari Contoh Kampanye Pemasaran.

Dengan memanfaatkan statistik deskriptif, teknik visualisasi data, dan alat pembelajaran lainnya, pemasar dapat memperoleh wawasan berharga tentang kinerja kampanye, mengidentifikasi peluang untuk perbaikan, dan membuat keputusan berbasis data untuk mengoptimalkan strategi dan taktik pemasaran mereka.

Kesimpulannya, meng pembahasan kinerja kampanye pemasaran sangat penting bagi pemasar untuk mengoptimalkan upaya pemasaran mereka dan mencapai hasil yang lebih baik. Dengan menggunakan teknik pembelajaran deskriptif dan memilih KPI yang tepat, pemasar dapat memperoleh wawasan berharga tentang efektivitas kampanye mereka, membuat keputusan berdasarkan wawasan berbasis data, dan pada akhirnya meningkatkan pengembalian investasi pemasaran mereka.

3.7 EVALUASI DATA PEMASARAN DIGITAL PRODUK MINUMAN

Dalam contoh ini, kita akan mengeksplorasi bagaimana sebuah perusahaan minuman dapat memanfaatkan pembelajaran deskriptif untuk mengevaluasi kinerja kampanye pemasaran media sosialnya, mendapatkan wawasan tentang keterlibatan pelanggan, dan membuat keputusan berbasis data untuk mengoptimalkan upaya pemasarannya.

Pengumpulan dan Persiapan Data

Perusahaan minuman mengumpulkan data dari platform media sosialnya, seperti Facebook, X (sebelumnya Twitter), dan Instagram, untuk melacak kinerja kampanye pemasarannya. Data ini mencakup metrik seperti keterlibatan (suka, komentar, berbagi), jangkauan, tayangan, dan CTR. Perusahaan membersihkan dan mengkonsolidasikan data untuk memastikan keakuratan dan konsistensi sebelum pembahasan (lihat Tabel 3.4).

Tabel 3.4 Data Mentah Simulasi yang Dikumpulkan untuk Kampanye Pemasaran Media Sosial.

Tanggal	Tipe Postingan	Jangkauan	Klik	Suka	Dibagikan	Komentar
01-01-2023	Tautan	3047	133	113	23	66
02-01-2023	Gambar	4547	225	192	55	76
03-01-2023	Tautan	3747	187	128	40	75
04-01-2023	Tautan	1975	115	88	20	44
05-01-2023	Gambar	2806	146	126	24	61

Pemilihan Indikator Kinerja Utama (KPI)

Perusahaan minuman tersebut mengidentifikasi KPI yang relevan dan selaras dengan tujuan pemasarannya:

- Tingkat keterlibatan
- Tingkat pertumbuhan pengikut
- Tingkat konversi (klik ke situs web atau pembelian online)
- Biaya per keterlibatan.

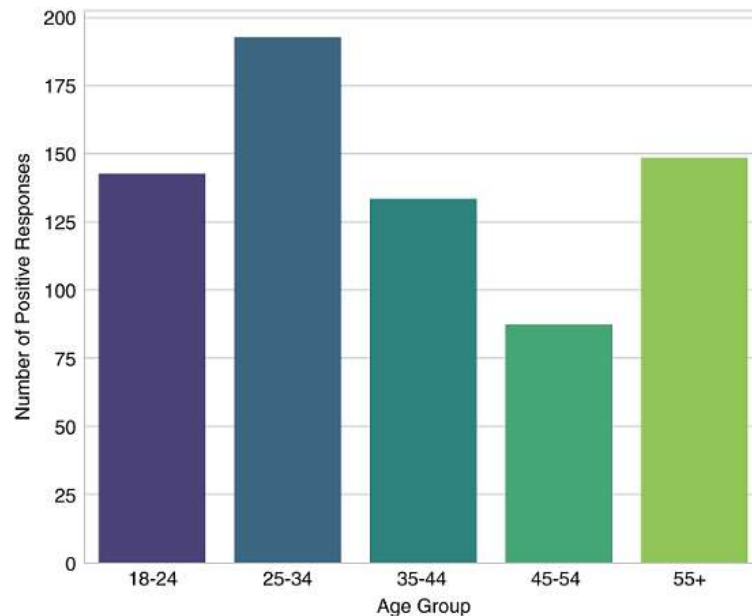
Pembahasan Deskriptif

Perusahaan tersebut menggunakan teknik pembahasan deskriptif, seperti statistik deskriptif dan visualisasi data, untuk mengpembahasan data pemasaran media sosialnya dan mengidentifikasi pola, tren, dan hubungan.

- **Statistik deskriptif.** Perusahaan menghitung rata-rata, median, dan deviasi standar dari tingkat keterlibatan, tingkat pertumbuhan pengikut, dan tingkat konversi untuk mengidentifikasi kecenderungan sentral dan dispersi kinerja media sosialnya.
- **Visualisasi data.** Perusahaan menggunakan grafik batang untuk membandingkan kinerja platform media sosialnya dan grafik garis untuk melacak kinerja KPI-nya dari waktu ke waktu. Misalnya, grafik garis dapat menunjukkan bagaimana tingkat keterlibatan bervariasi sepanjang kampanye, mengungkapkan lonjakan atau penurunan keterlibatan yang mungkin memerlukan penyelidikan lebih lanjut.

Pembahasan Segmentasi

Perusahaan minuman tersebut melakukan segmentasi audiensnya berdasarkan faktor-faktor seperti demografi, geografi, dan perilaku keterlibatan untuk lebih memahami preferensi pelanggan dan menyesuaikan upaya pemasarannya. Misalnya, perusahaan dapat mengpembahasan tingkat keterlibatan berdasarkan kelompok usia atau lokasi untuk mengidentifikasi audiens target yang paling responsif terhadap upaya pemasarannya (lihat Gambar 3.11).



Gambar 3.11 Hasil Pembahasan Segmentasi (Respons yang Disegmentasikan berdasarkan Kelompok Usia).

Wawasan dan Optimalisasi

Berdasarkan wawasan yang diperoleh dari pembelajaran deskriptif, perusahaan minuman dapat membuat keputusan berbasis data untuk mengoptimalkan kampanye pemasaran media sosialnya. Misalnya, jika pembahasan mengungkapkan bahwa Instagram memiliki tingkat keterlibatan tertinggi di antara audiens targetnya, perusahaan dapat mengalokasikan lebih banyak sumber daya ke platform ini. Selain itu, perusahaan dapat menyesuaikan strategi kontennya berdasarkan wawasan segmentasi audiens, menyesuaikan pesan dan elemen kreatifnya agar lebih sesuai dengan segmen pelanggan tertentu.

Kesimpulannya, memanfaatkan teknik pembelajaran deskriptif dapat membantu perusahaan minuman mengevaluasi kinerja kampanye pemasaran media sosialnya, memperoleh wawasan berharga tentang keterlibatan pelanggan, dan membuat keputusan yang tepat untuk mengoptimalkan upaya pemasarannya.

3.8 KESIMPULAN

Pembelajaran deskriptif merupakan landasan pemasaran berbasis data, bertindak sebagai dasar tempat semua metodologi pembelajaran selanjutnya dibangun. Dalam bab ini, kita memulai perjalanan yang mencerahkan melalui ranah pembelajaran deskriptif dalam pemasaran, menggarisbawahi peran pentingnya dalam menjelaskan narasi yang tersembunyi di dalam lautan data yang luas.

Melalui diskusi kita, menjadi jelas bahwa pada intinya, pembelajaran deskriptif bukan hanya tentang pengolahan data. Ini adalah seni dan ilmu menghidupkan data, mengubah angka mentah menjadi cerita bermakna yang menerangi kinerja masa lalu, perilaku pelanggan, dan dinamika pasar. Alat dan teknik yang kita bahas—dari ukuran kecenderungan sentral sederhana hingga metode visualisasi data yang canggih—memberikan pemasar pandangan

menyeluruh tentang data mereka, memberdayakan mereka untuk menguraikan pola yang kompleks, menentukan kekuatan dan kelemahan, dan membuat keputusan yang lebih tepat.

Pentingnya memahami data historis sangatlah penting karena menawarkan lensa yang memungkinkan pemasar untuk merefleksikan strategi masa lalu, mengukur efektivitasnya, dan mengambil pelajaran untuk inisiatif di masa mendatang. Hal ini dicontohkan dalam contoh praktis kita dengan perusahaan minuman. Dengan menggunakan pembelajaran deskriptif, perusahaan tidak hanya dapat mengevaluasi kinerja kampanyenya tetapi juga mengungkap wawasan yang dapat ditindaklanjuti yang secara langsung menginformasikan optimasi strategis. Selanjutnya, interaksi antara pembelajaran deskriptif dan tahapan selanjutnya dari proses ilmu data dalam pemasaran tidak dapat terlalu ditekankan.

Pembelajaran deskriptif menyiapkan landasan, mempersiapkan pemasar untuk fase pembelajaran prediktif dan preskriptif. Ini adalah dasar di mana prediksi tentang tren masa depan dibuat dan strategi preskriptif untuk hasil optimal dirancang.

Seiring lanskap pemasaran digital terus berkembang dan menjadi lebih kompleks, keharusan bagi pemasar untuk merangkul dan menguasai pembelajaran deskriptif semakin kuat. Di dunia yangibanjiri data, kemampuan untuk secara efektif meringkas, memvisualisasikan, dan menafsirkan data ini bukan lagi sekadar keuntungan tetapi suatu kebutuhan.

Sebagai penutup, seiring pemasar maju dalam perjalanan ilmu data mereka, mereka harus mengingat pepatah, "Anda tidak dapat mengetahui ke mana Anda akan pergi sampai Anda tahu dari mana Anda berasal." Pembelajaran deskriptif menawarkan pengetahuan itu, memastikan pemasar tidak hanya mengetahui dari mana mereka berasal tetapi juga memahami seluk-beluk perjalanan mereka, memungkinkan mereka untuk merencanakan jalur yang lebih terinformasi dan sukses ke depan.

LATIHAN SOAL 3.1

PEMBAHASAN DESKRIPTIF DATA PEMASARAN

Tujuan:

Memahami dan mendeskripsikan kecenderungan sentral, dispersi, dan asosiasi dalam data pemasaran.

Tugas:

1. **Hitung Statistik Deskriptif:** Hitung rata-rata, median, dan modus untuk variabel seperti 'ad spend', 'Click', dan 'Sell'.
2. **Visualisasi:** Buat grafik batang untuk metrik keterlibatan, grafik garis untuk pengeluaran iklan dari waktu ke waktu, dan diagram sebar untuk menunjukkan hubungan antara pengeluaran iklan dan konversi.
3. **Interpretasi:** Pembahasan hasilnya, diskusikan temuan atau pola yang menarik.

Langkah-langkah:

1. Impor Pustaka:

```
1. import pandas as pd
```

- Kita mengimpor pustaka Pandas, yang sangat penting untuk manipulasi dan pembahasan data.

2. Muat Dataset:

```
2. marketing_data = pd.read_csv('path_to_csv_file')
```

- Ganti **'path_to_csv_file'** dengan jalur sebenarnya dari file CSV marketing_data.
- Baris ini membaca file CSV ke dalam Pandas DataFrame, yang merupakan struktur data berlabel dua dimensi.

3. Hitung Statistik Deskriptif:

```
3. pca_data= marketing_data[['Feature1', 'Feature2', 'Feature3']]
```

- Fungsi **describe()** di Pandas memberikan ringkasan kecenderungan sentral, dispersi, dan bentuk distribusi dataset, tidak termasuk nilai NaN. Fungsi ini menghitung statistik seperti rata-rata, deviasi standar, nilai minimum, dan nilai maksimum untuk setiap kolom.

Visualisasi:

1. Impor Pustaka Visualisasi:

```
4. from sklearn.preprocessing import StandardScaler
```

- Kita mengimpor Matplotlib, sebuah pustaka yang banyak digunakan untuk membuat visualisasi statis, animasi, dan interaktif di Python.

2. Buat Bagan Batang untuk 'Metrik Keterlibatan Total':

```
5. marketing_data[['Likes', 'Shares', 'Comments']].sum().plot(kind='bar')
6. plt.title('Total Engagement Metrics')
7. plt.ylabel('Total Count')
8. plt.show()
```

- Cuplikan kode ini membuat grafik batang yang mewakili jumlah total 'Likes', 'Shares', dan 'Comments'.
- Fungsi **sum()** digunakan untuk menghitung total untuk setiap metrik keterlibatan.
- **plot(kind='bar')** menghasilkan grafik batang, dan **plt.title**, **plt.ylabel**, dan **plt.show()** digunakan untuk mengatur judul, memberi label pada sumbu y, dan menampilkan grafik, masing-masing.

3. Buat Grafik Garis untuk 'Ad Spend Over Time':

```
9. marketing_data.plot(x='Date', y='Ad Spend', kind='line')
10. plt.title('Ad Spend Over Time')
11. plt.ylabel('Ad Spend')
12. plt.xlabel('Date')
13. plt.show()
```

- Baris kode ini menghasilkan grafik garis yang menunjukkan bagaimana 'Ad Spend' bervariasi dari waktu ke waktu.
- Sumbu x mewakili tanggal, dan sumbu y menunjukkan 'Ad Spend' untuk setiap tanggal.

4. Scatterplot untuk Menunjukkan Hubungan Antara 'AdSpend' dan 'Conversions':

```
14. marketing_data.plot(x='AdSpend', y='Conversions', kind='scatter')
15. plt.title('Ad Spend vs Conversions')
16. plt.xlabel('Ad Spend')
```

```
17. plt.ylabel('Conversions')
18. plt.show()
```

- Kode ini membuat scatterplot untuk memvisualisasikan hubungan antara 'Ad Spend' dan 'Conversions'.
- Setiap titik pada plot mewakili sepasang nilai dari dataset.

LATIHAN SOAL 3.2

VISUALISASI DAN INTERPRETASI DATA

Tujuan:

Membuat dan menginterpretasikan berbagai visualisasi data untuk memahami tren pasar dan kinerja kampanye.

Tugas:

1. **Pembahasan Deret Waktu:** Gunakan grafik garis untuk meng pembahasan tren 'Click' dan 'Conversion' dari waktu ke waktu.
2. **Pembahasan Segmentasi:** Buat peta panas untuk memvisualisasikan metrik keterlibatan di berbagai segmen pelanggan.
3. **Pembahasan Kinerja:** Kembangkan visualisasi bergaya dasbor yang menyajikan beberapa KPI dan interpretasikan hasilnya untuk mengukur efektivitas kampanye pemasaran.

Langkah-langkah:

1. Buat Grafik Garis untuk 'Klik dan Konversi dari Waktu ke Waktu':

```
19. marketing_data.plot(x='Date', y=['Clicks', 'Conversions'],
    kind='line')
20. plt.title('Clicks and Conversions Over Time')
21. plt.ylabel('Count')
22. plt.xlabel('Date')
23. plt.legend(['Clicks', 'Conversions'])
24. plt.show()
```

- Mengikuti langkah-langkah yang dijalankan untuk Latihan 3.1, baris kode ini menghasilkan grafik garis yang menunjukkan bagaimana 'Clicks' dan 'Conversions' bervariasi dari waktu ke waktu.
- Kolom 'Click' dan 'Conversions' diplot pada sumbu y, dan kolom 'date' digunakan untuk sumbu x.
- Fungsi **legenda** digunakan untuk membedakan antara dua baris ('Click' dan 'Conversions').

2. Membuat Peta Panas untuk Metrik Keterlibatan di Berbagai Hari dalam Seminggu:

- Pertama, kita akan membuat kolom baru untuk mewakili 'Day of Week', dan kemudian mengelompokkan data sesuai dengan itu.

```
25. import seaborn as sns
26. # Creating a new column for the day of the week
27. marketing_data['Day of Week'] =
    marketing_data['Date'].dt.day_name()
28. # Grouping data by the day of the week and summing
```

```
29. # engagement metrics engagement_by_day = engagement_by_day =
    marketing_data.groupby('Day of Week')[['Likes', 'Shares',
    'Comments']].sum()
30. # Creating a heatmap
31. plt.figure(figsize=(10, 6))
32. sns.heatmap(engagement_by_day, annot=True, fmt="d",
    cmap='viridis')
33. plt.title('Engagement Metrics by Day of the Week')
34. plt.show()
```

- Fungsi **groupby** digunakan untuk mengelompokkan data berdasarkan 'Hari dalam Seminggu'.
- **sns.heatmap** dari pustaka Seaborn membuat peta panas untuk memvisualisasikan metrik keterlibatan untuk setiap hari dalam seminggu.

3. Membuat Visualisasi Bergaya Dasbor yang Menyajikan Beberapa KPI:

- Kita akan memilih beberapa KPI dan membuat visualisasi gabungan.

```
35. fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(15, 10))
36. # Plotting Ad Spend over time
37. marketing_data.plot(x='Date', y='Ad Spend', ax=axes[0,0])
38. axes[0,0].set_title('Ad Spend Over Time')
39. axes[0,0].set_ylabel('Ad Spend ($)')
40. # Plotting Impressions over time
41. marketing_data.plot(x='Date', y='Impressions', ax=axes[0,1],
    color='green')
42. axes[0,1].set_title('Impressions Over Time')
43. axes[0,1].set_ylabel('Impressions')
44. # Plotting Clicks over time marketing_data.plot(x='Date',
    y='Clicks', ax=axes[1,0], color='orange')
45. axes[1,0].set_title('Clicks Over Time')
46. axes[1,0].set_ylabel('Clicks')
47. # Plotting Conversions over time marketing_data.plot(x='Date',
    y='Conversions', ax=axes[1,1], color='red')
48. axes[1,1].set_title('Conversions Over Time')
49. axes[1,1].set_ylabel('Conversions')
50. plt.tight_layout()
51. plt.show()
```

- Kode ini membuat grid plot 2x2, masing-masing menampilkan KPI yang berbeda dari waktu ke waktu.
- Fungsi **subplots** digunakan untuk membuat tata letak grid, dan **plot** individual dibuat menggunakan metode plot dengan sumbu yang ditentukan.

Kode ini, ketika dijalankan menggunakan dataset, akan memberikan pandangan komprehensif tentang kinerja kampanye pemasaran, menyoroti tren dan pola yang penting untuk pengambilan keputusan strategis. Mari kita jalankan kode ini dan amati hasilnya.

BAB 4

MENGUKUR KEPASTIAN DAN MENGUJI HIPOTESIS PEMASARAN

4.1 PENDAHULUAN

Pembahasan inferensial dan pengujian hipotesis merupakan pilar utama ilmu data pemasaran, yang memungkinkan para profesional untuk melampaui sekadar pengamatan dan bergerak menuju pengambilan keputusan yang proaktif dan berbasis data. Karena bisnis dibanjiri dengan sejumlah besar data, pertanyaan mendesaknya adalah, Bagaimana data ini dapat diubah menjadi wawasan yang dapat ditindaklanjuti? Jawabannya terletak pada kemampuan untuk menyimpulkan tren yang lebih luas dari data sampel dan memvalidasi asumsi melalui pengujian hipotesis yang ketat.

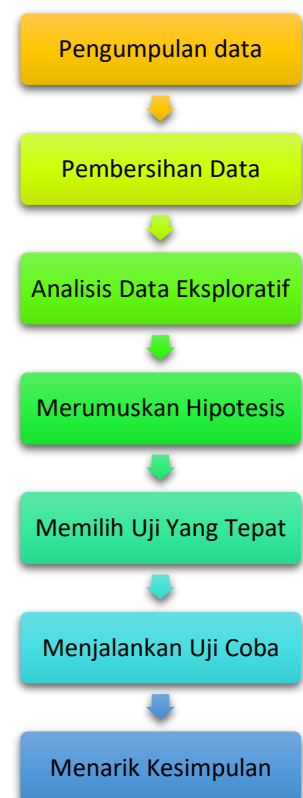
Bab ini membahas secara mendalam dunia pembahasan inferensial, mengungkapkan peran pentingnya dalam pemasaran. Dengan memeriksa teknik statistik yang memungkinkan pemasar untuk menggeneralisasi temuan dari sampel ke populasi yang lebih besar, kami bertujuan untuk menyoroti nilai luar biasa yang ditawarkan teknik-teknik ini.

Di luar sekadar teori, bab ini menyoroti aplikasi dunia nyata, menunjukkan bagaimana bisnis menggunakan alat-alat ini untuk mendorong hasil. Dari memahami perilaku, preferensi, dan tren pelanggan pada tingkat makroskopis hingga memverifikasi dampak intervensi pemasaran tertentu, pembahasan inferensial dan pengujian hipotesis muncul sebagai aset yang tak ternilai dalam persenjataan pemasar.

Melalui eksplorasi konsep-konsep kunci, teknik, dan contoh-contoh praktis, bab ini memberikan pemahaman komprehensif kepada pembaca tentang pembelajaran inferensial dan pengujian hipotesis dalam konteks pemasaran. Dengan bekal pengetahuan ini, para profesional pemasaran akan lebih siap untuk menavigasi lanskap data yang kompleks, membuat keputusan yang tidak hanya informatif tetapi juga berdampak.

Saat kita memulai eksplorasi mendetail tentang pembelajaran inferensial dalam domain pemasaran, penting untuk memvisualisasikan seluruh proses dari awal hingga akhir. Gambar 4.1 menawarkan bagan alur yang secara tepat mewakili perjalanan ini, dari tahap awal pengumpulan data hingga langkah-langkah akhir penarikan kesimpulan yang bermakna.

Dengan menyajikan langkah-langkah ini dengan cara yang jelas dan terorganisir, gambar tersebut membantu menghilangkan misteri proses, memberikan pembaca peta jalan tentang bagaimana pembelajaran inferensial diterapkan untuk mengubah data mentah menjadi wawasan yang dapat ditindaklanjuti. Ini menggarisbawahi pendekatan sistematis



Gambar 4.1 Proses Analisis Inferensial, dari Pengumpulan Data hingga Penarikan Kesimpulan.

yang diperlukan dalam pengujian hipotesis dan pembahasan inferensial, yang sangat penting dalam membuat keputusan pemasaran berbasis data.

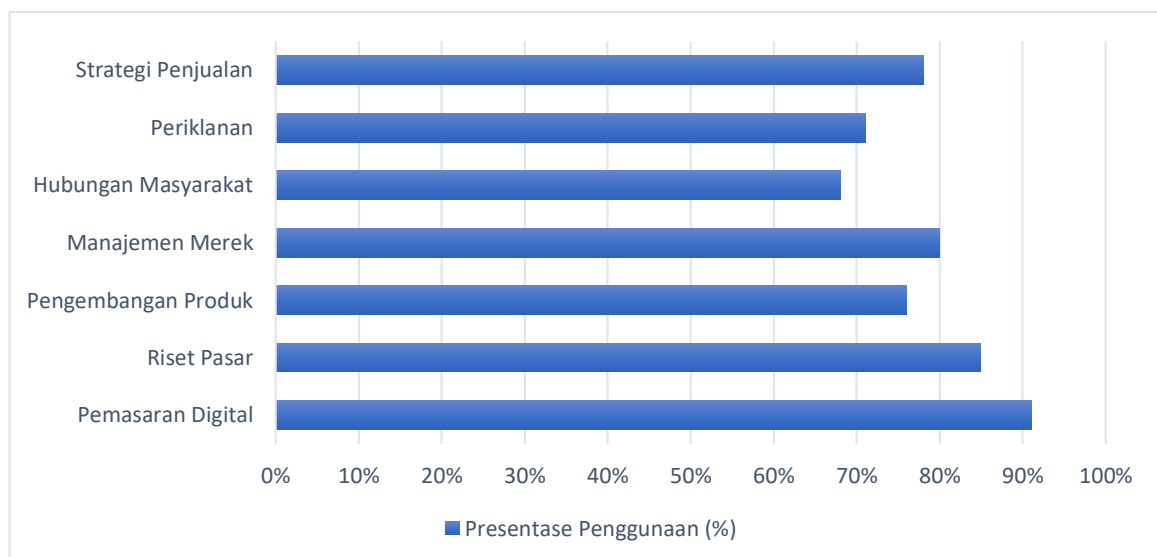
4.2 PROYEKSI POPULASI DAN VALIDASI ASUMSI PASAR

Gambaran Umum Pembelajaran Inferensial

Pembelajaran inferensial adalah cabang statistik yang berkaitan dengan pengambilan kesimpulan tentang suatu populasi berdasarkan sampel data yang lebih kecil. Dalam konteks pemasaran, pembelajaran inferensial memungkinkan organisasi untuk memahami perilaku, preferensi, dan tren pelanggan pada tingkat yang lebih luas, memberikan wawasan berharga untuk membuat keputusan yang tepat dan mengoptimalkan strategi pemasaran.

Tidak seperti pembelajaran deskriptif, yang berfokus pada meringkas dan memvisualisasikan data dari satu dataset, pembelajaran inferensial bertujuan untuk menggeneralisasi temuan dari sampel ke populasi yang lebih besar. Hal ini sangat berguna dalam pemasaran karena seringkali tidak praktis atau tidak mungkin untuk mengumpulkan data dari setiap pelanggan atau calon pelanggan. Dengan menggunakan pembelajaran inferensial, pemasar dapat memperoleh wawasan tentang karakteristik populasi yang lebih besar, seperti pengeluaran rata-rata, preferensi, dan pola pembelian, berdasarkan sampel yang lebih kecil dan lebih mudah dikelola.

Pembahasan inferensial melibatkan penggunaan teori probabilitas dan berbagai teknik statistik untuk memperkirakan parameter populasi, seperti rata-rata, proporsi, dan varians, berdasarkan data sampel. Teknik-teknik ini memungkinkan pemasar untuk mengukur ketidakpastian yang terkait dengan perkiraan mereka dan membuat prediksi dengan tingkat kepercayaan tertentu. Misalnya, seorang pemasar dapat menggunakan pembahasan inferensial untuk memperkirakan pendapatan rata-rata yang dihasilkan oleh segmen pelanggan tertentu, bersama dengan interval kepercayaan yang memberikan rentang di mana rata-rata populasi sebenarnya kemungkinan berada.



Gambar 4.2 Berbagai Sektor dalam Pemasaran yang Paling Sering Menggunakan Pembahasan Inferensial.

Sebagai penutup diskusi kita tentang kekuatan pembahasan inferensial dalam membentuk keputusan pemasaran berbasis data, sangat bermanfaat untuk meneliti di mana teknik-teknik ini paling sering diterapkan dalam industri pemasaran. Gambar 4.2 menyajikan bagan batang yang menunjukkan berbagai sektor dalam pemasaran yang paling sering menggunakan pembahasan inferensial. Visualisasi ini membantu mengkontekstualisasikan penerapan pembahasan inferensial di dunia nyata, menyoroti sektor-sektor di mana teknik-teknik ini memberikan dampak yang signifikan. Dari bagan ini, pembaca dapat menghargai kegunaan pembahasan inferensial yang luas di berbagai domain pemasaran, memberikan gambaran yang jelas tentang bagaimana metode-metode ini diintegrasikan ke dalam berbagai strategi pemasaran untuk mendorong wawasan pelanggan yang lebih dalam dan meningkatkan hasil pemasaran.

Singkatnya, pembahasan inferensial adalah alat yang ampuh yang memungkinkan pemasar untuk membuat keputusan berbasis data dan mengoptimalkan upaya pemasaran mereka berdasarkan wawasan yang diperoleh dari sampel data. Dengan memahami dan menerapkan teknik pembelajaran inferensial, para profesional pemasaran dapat memperoleh wawasan yang lebih dalam tentang perilaku, preferensi, dan tren pelanggan, yang mengarah pada hasil pemasaran yang lebih baik.

Dasar-Dasar Probabilitas

Probabilitas adalah fondasi yang menjadi dasar pembelajaran inferensial, menawarkan cara matematis untuk mengukur ketidakpastian. Pada intinya, probabilitas memberikan ukuran kemungkinan terjadinya suatu peristiwa tertentu, biasanya dinyatakan sebagai angka antara 0 dan 1, termasuk. Bagian ini akan memperkenalkan konsep-konsep inti dalam probabilitas yang mendasari banyak teknik pembelajaran yang dibahas kemudian dalam bab ini.

Definisi Probabilitas

Probabilitas (P) suatu peristiwa E didefinisikan sebagai:

$$P(E) = \frac{\text{Jumlah Hasil yang Menguntungkan Untuk } E}{\text{Jumlah Total Kemungkinan Hasil}},$$

di mana

Probabilitas suatu peristiwa yang mustahil adalah 0.

Probabilitas suatu peristiwa yang pasti adalah 1.

Untuk setiap peristiwa E , $0 \leq P(E) \leq 1$.

CONTOH SOAL:

PROBABILITAS MENKLIK IKLAN

Bayangkan Anda adalah seorang pemasar digital untuk sebuah perusahaan yang menjual peralatan olahraga. Anda telah mendesain dua iklan banner yang berbeda (Iklan A dan Iklan B) untuk produk yang sama, dan Anda ingin menentukan probabilitas bahwa pengunjung situs web akan mengklik setiap iklan.

Mari kita asumsikan bahwa situs web Anda memiliki arus trafik yang stabil, dan Anda memiliki data historis yang menunjukkan bahwa dari 1.000 pengunjung, 50 mengklik Iklan A, dan 75 mengklik Iklan B.

Menghitung Probabilitas

Untuk Iklan A:

- Jumlah hasil yang menguntungkan (klik) adalah 50.
- Jumlah total kemungkinan hasil (pengunjung) adalah 1000.

Menggunakan rumus probabilitas:

$$P(\text{Mengklik Iklan A}) = \frac{\text{Jumlah hasil yang Menguntungkan Untuk Iklan A}}{\text{Jumlah Total Kemungkinan Hasil}}$$
$$P(\text{Mengklik Iklan A}) = \frac{50}{1000} = 0.05$$

Jadi, probabilitas bahwa pengunjung mengklik Iklan A adalah 0,05, atau 5%.

Untuk Iklan B:

- Jumlah hasil yang menguntungkan (klik) adalah 75.
- Jumlah total kemungkinan hasil (pengunjung) adalah 1000.

$$P(\text{Mengklik Iklan B}) = \frac{75}{1000} = 0.075$$

Dengan demikian, probabilitas bahwa pengunjung mengklik Iklan B adalah 0,075, atau 7,5%.

Interpretasi Hasil

- Probabilitas suatu peristiwa tertentu, seperti pengguna mengunjungi situs web (dengan asumsi situs web tersebut dapat diakses dan online), adalah 1 (atau 100%).
- Probabilitas suatu peristiwa yang mustahil, seperti klik dari iklan yang tidak ada, adalah 0.
- Probabilitas yang dihitung untuk mengklik Iklan A dan Iklan B berada di antara 0 dan 1, yang sesuai dengan prinsip bahwa probabilitas peristiwa apa pun E harus memenuhi $0 \leq P(E) \leq 1$.

Dengan menggunakan informasi probabilitas ini, Anda dapat menyimpulkan bahwa Iklan B saat ini memiliki peluang lebih tinggi untuk diklik oleh pengunjung situs web Anda. Hal ini dapat membantu Anda dalam memutuskan iklan mana yang akan dialokasikan lebih banyak sumber daya atau apakah akan mendesain ulang Iklan A.

Ruang Sampel dan Kejadian

Ruang sampel, yang dilambangkan sebagai S , mewakili himpunan semua kemungkinan hasil dari suatu percobaan acak. Suatu kejadian adalah himpunan bagian dari ruang sampel ini. Misalnya, dalam pelemparan koin, ruang sampelnya adalah $S = \{H, T\}$, dan suatu kejadian mungkin adalah mendapatkan kepala.

Aturan Probabilitas Dasar

- **Kejadian komplementer.** Probabilitas suatu kejadian tidak terjadi, yang dilambangkan dengan $P(E')$, diberikan oleh:

$$P(E') = 1 - P(E)$$

- **Aturan penjumlahan.** Untuk setiap dua kejadian E dan F :

$$P(E \text{ or } F) = P(E) + P(F) - P(E \text{ dan } F)$$

Jika E dan F saling eksklusif (keduanya tidak dapat terjadi bersamaan), maka:

$$P(E \text{ atau } F) = P(E) + P(F)$$

- **Aturan perkalian.** Untuk setiap dua kejadian independen dan (terjadinya salah satu kejadian tidak memengaruhi probabilitas kejadian lainnya):

$$P(E \text{ dan } F) = P(E) \times P(F)$$

Peluang Bersyarat

Peluang bersyarat adalah peluang suatu peristiwa E terjadi dengan syarat peristiwa lain telah terjadi, F yang diwakili sebagai $P(E|F)$. Peluang ini dihitung sebagai:

$$P(E|F) = \frac{P(E \text{ dan } F)}{P(F)}$$

Teorema Bayes

Teorema Bayes adalah konsep fundamental dalam teori probabilitas dan statistik, yang menyediakan cara untuk menemukan probabilitas ketika probabilitas lain tertentu diketahui. Teorema ini direpresentasikan sebagai:

$$P(E|F) = \frac{P(F|E) \times P(E)}{P(F)},$$

di mana

$P(E|F)$ adalah probabilitas posterior.

$P(F|E)$ adalah kemungkinan.

$P(E)$ adalah probabilitas awal dari E .

$P(F)$ adalah probabilitas total dari F .

Teorema Bayes memiliki aplikasi luas di berbagai bidang seperti pengujian medis, pembelajaran mesin, dan bahkan pembelajaran pemasaran.

STUDI KASUS:

MEMPERSONALISASI KAMPANYE EMAIL UNTUK PERITEL ONLINE

Mari kita ambil contoh peritel online, ShopStream, yang menawarkan berbagai produk mulai dari elektronik hingga pakaian. ShopStream memiliki basis pelanggan yang luas, dan mereka mengirimkan email promosi mingguan. Mereka menyadari bahwa tingkat keterlibatan email promosi umum mereka menurun. Untuk mengatasi hal ini, mereka ingin mempersonalisasi kampanye email mereka berdasarkan kemungkinan minat pengguna.

Tujuan: Untuk memprediksi kemungkinan bahwa pelanggan tertarik pada kategori produk tertentu (misalnya, elektronik) berdasarkan perilaku mereka di masa lalu dan menggunakan informasi ini untuk mengirimkan promosi email yang ditargetkan.

Penerapan Teorema Bayes

1. Probabilitas Awal ($P(E)$):

- Ini mewakili probabilitas keseluruhan bahwa pelanggan acak tertarik pada elektronik, berdasarkan data historis. Katakanlah 20% pelanggan mereka telah menunjukkan minat pada elektronik di masa lalu.

2. Kemungkinan ($P(F|E)$):

- Ini adalah probabilitas bahwa pelanggan yang tertarik pada elektronik menunjukkan perilaku tertentu. Misalnya, dari kampanye sebelumnya, mereka telah memperhatikan bahwa 60% pengguna yang tertarik pada elektronik mengklik konten terkait elektronik di email mereka.

3. Probabilitas Total $P(F)$:

- Ini adalah probabilitas keseluruhan bahwa setiap pelanggan, terlepas dari minat mereka, mengklik konten terkait elektronik. Misalkan ini adalah 25%.

Dengan menggunakan teorema Bayes, mereka ingin menemukan probabilitas posterior, $P(E|F)$, yaitu probabilitas bahwa pelanggan tertarik pada elektronik setelah mereka mengklik konten yang berkaitan dengan elektronik.

Dengan memasukkan angka-angka:

$$P(E|F) = \frac{P(F|E) \times P(E)}{P(F)}$$

$$P(E|F) = \frac{(0.6 \times 0.2)}{(0.25)}$$

$$P(E|F) = 0.48 \text{ atau } 48\%$$

- **Interpretasi.** Jika pelanggan mengklik konten terkait elektronik, ada kemungkinan 48% mereka tertarik pada elektronik, yang secara signifikan lebih tinggi daripada probabilitas awal 20%.
- **Hasil.** Berdasarkan pemahaman yang lebih baik ini, ShopStream mulai mengirimkan lebih banyak promosi khusus elektronik kepada pengguna yang mengklik konten elektronik. Hal ini menghasilkan peningkatan 15% dalam rasio klik-tayang untuk segmen ini, membuktikan efektivitas pendekatan Bayesian dalam menyempurnakan strategi pemasaran mereka.
- **Kesimpulan.** Teorema Bayes memungkinkan ShopStream untuk memperbarui keyakinannya tentang pelanggan berdasarkan data baru, mengoptimalkan upaya pemasaran mereka. Dengan mendasarkan keputusan pada probabilitas Bayesian, mereka mencapai kampanye email yang lebih personal dan efektif. Ini menunjukkan kekuatan teorema Bayes dalam mengubah data menjadi wawasan pemasaran yang dapat ditindaklanjuti.

4.3 PENTINGNYA ILMU DATA PEMASARAN

Memahami dasar-dasar probabilitas membekali pemasar untuk menafsirkan data secara lebih intuitif. Misalnya, dengan menilai probabilitas pelanggan membeli produk setelah melihat iklan, pemasar dapat mengoptimalkan penempatan iklan. Atau, dengan menggunakan teorema Bayes, mereka dapat memperbarui keyakinan mereka tentang preferensi pelanggan berdasarkan data baru.

Kesimpulannya, probabilitas menawarkan kerangka kerja yang kuat untuk memahami ketidakpastian dan variabilitas, yang penting untuk membuat keputusan yang terinformasi dan dinilai risikonya di bidang pemasaran. Seiring kita melanjutkan bab ini, pentingnya konsep-konsep dasar ini akan menjadi lebih jelas, meletakkan dasar bagi teknik pembelajaran inferensial tingkat lanjut.

Tes Parametrik Versus Nonparametrik

Tes statistik berfungsi sebagai tulang punggung pembelajaran inferensial, membantu pemasar membuat keputusan berdasarkan data sampel.

Secara umum, tes ini dapat dikategorikan menjadi dua jenis utama: tes parametrik dan nonparametrik. Bagian ini akan membahas perbedaan antara kedua kategori ini dan mengeksplorasi implikasinya untuk ilmu data pemasaran.

Apa Itu Tes Parametrik?

Uji parametrik adalah uji statistik yang membuat asumsi spesifik tentang parameter distribusi populasi dari mana sampel diambil.

Karakteristik Uji Parametrik

- Mengasumsikan data mengikuti distribusi tertentu (misalnya, distribusi normal)
- Mengharapkan data diukur setidaknya pada skala interval atau rasio
- Mengasumsikan homogenitas varians saat membandingkan dua kelompok atau lebih

Sensitif terhadap outlier

- Uji Parametrik Umum
- Uji t (untuk membandingkan rata-rata)
- Pembahasan varians (ANOVA)
- Regresi linier

Apa Itu Uji Nonparametrik?

Uji nonparametrik, sering disebut uji bebas distribusi, tidak membuat asumsi ketat apa pun tentang parameter populasi.

Karakteristik Uji Nonparametrik

- Tidak mengasumsikan distribusi spesifik untuk data
- Cocok untuk data ordinal, nominal, interval, atau rasio
- Tangguh terhadap heterogenitas varians
- Kurang sensitif terhadap outlier

Uji Nonparametrik Umum

- Uji Mann-Whitney U
- Uji peringkat bertanda Wilcoxon
- Uji Kruskal-Wallis

- Korelasi peringkat Spearman

Memilih Antara Uji Parametrik dan Nonparametrik

Keputusan tentang jenis uji mana yang akan digunakan bergantung pada hal-hal berikut:

- **Distribusi data.** Jika data Anda terdistribusi normal, uji parametrik lebih disukai karena kekuatan statistiknya yang lebih besar; jika tidak, pertimbangkan uji nonparametrik.
- **Skala pengukuran.** Uji nonparametrik lebih serbaguna dan dapat menangani berbagai skala data, termasuk nominal dan ordinal.
- **Ukuran sampel.** Untuk ukuran sampel kecil, uji nonparametrik seringkali lebih tepat karena tidak bergantung pada asumsi distribusi.
- **Kehadiran outlier.** Jika data Anda memiliki outlier yang signifikan, uji nonparametrik mungkin lebih tepat karena kurang sensitif terhadap nilai ekstrem.

Tabel 4.1 memberikan perbandingan yang jelas dan ringkas antara uji parametrik dan nonparametrik, menyoroti perbedaan dasar, kelebihan, dan kekurangannya. Tabel ini berfungsi sebagai referensi yang sangat berharga bagi pemasar dan ilmuwan data, membantu mereka dalam membuat keputusan yang tepat berdasarkan sifat data mereka, seperti distribusi, skala pengukuran, ukuran sampel, dan keberadaan outlier. Memahami aspek-aspek ini sangat penting dalam memilih uji statistik yang paling tepat yang sesuai dengan karakteristik data dan tujuan pembahasan.

Tabel 4.1 Perbedaan, Kelebihan, dan Kekurangan Uji Parametrik Versus Nonparametrik.

	Tes Parametrik	Tes Nonparametrik
Perbedaan mendasar	Mengasumsikan data mengikuti distribusi tertentu, seringkali distribusi normal	Tidak mengasumsikan data mengikuti distribusi tertentu
Keuntungan	Ampuh, dengan sensitivitas yang lebih besar untuk mendeteksi hasil yang signifikan ketika asumsi terpenuhi	Lebih kuat, dapat digunakan ketika data tidak memenuhi asumsi parametrik
Kerugian	Sensitif terhadap outlier, memerlukan data untuk memenuhi asumsi tertentu (misalnya, normalitas, kesamaan varians)	Umumnya kurang ampuh, mungkin tidak mendeteksi perbedaan halus dalam data

Implikasi untuk Ilmu Data Pemasaran

Dalam ilmu data pemasaran, pilihan antara uji parametrik dan nonparametrik dapat sangat memengaruhi kesimpulan yang diambil dari data. Misalnya:

- Saat menilai efek iklan baru terhadap penjualan, jika data penjualan terdistribusi normal, uji parametrik dapat digunakan untuk menentukan apakah ada perbedaan signifikan dalam rata-rata sebelum dan sesudah kampanye iklan.
- Namun, jika pemasar meng pembahasan data ordinal, seperti peringkat kepuasan pelanggan dari 1 hingga 5, uji nonparametrik mungkin lebih sesuai.

Pada intinya, memahami asumsi dan karakteristik yang mendasari uji ini memungkinkan pemasar untuk memilih metode pembahasan yang paling tepat, memastikan kesimpulan yang kuat dan bermakna.

Baik uji parametrik maupun nonparametrik memiliki kelebihan dan keterbatasannya masing-masing. Kuncinya adalah memahami kapan harus menerapkan masing-masing, berdasarkan sifat data dan pertanyaan pembelajaran yang ada. Seiring berjalannya bab ini, kita akan mengeksplorasi berbagai teknik statistik secara detail, menekankan penerapannya dalam skenario pemasaran dunia nyata.

Konsep Kunci dalam Pembahasan Inferensial

Untuk menerapkan pembahasan inferensial secara efektif dalam pemasaran, penting untuk memahami beberapa konsep kunci yang mendasari pendekatan statistik ini. Pada bagian ini, kita akan membahas konsep-konsep tersebut, termasuk populasi, sampel, teknik pengambilan sampel, dan distribusi sampel, serta margin kesalahan, interval kepercayaan, dan kesalahan standar.

Populasi dan Sampel

Populasi mengacu pada seluruh kelompok individu atau entitas yang menjadi perhatian dalam studi atau pembahasan tertentu. Dalam pemasaran, ini mungkin termasuk semua pelanggan, calon pelanggan, atau pengguna produk atau layanan. Namun, sampel adalah subset yang lebih kecil dari populasi, yang dipilih untuk mewakili kelompok yang lebih luas.

Teknik Pengambilan Sampel

Teknik pengambilan sampel adalah metode yang digunakan untuk memilih sampel representatif dari suatu populasi. Teknik pengambilan sampel yang umum meliputi pengambilan sampel acak sederhana, pengambilan sampel bertingkat, dan pengambilan sampel kluster. Setiap teknik memiliki kelebihan dan kekurangannya masing-masing, dan pemilihan metode yang tepat bergantung pada tujuan pembelajaran dan sifat populasi.

Distribusi Sampel

Distribusi sampel adalah distribusi probabilitas dari statistik tertentu berdasarkan sampel acak yang diambil dari suatu populasi. Teorema limit pusat, hasil fundamental dalam teori probabilitas, menyatakan bahwa distribusi sampel dari rata-rata sampel mendekati distribusi normal seiring bertambahnya ukuran sampel, terlepas dari bentuk distribusi populasi.

Margin Kesalahan

Margin kesalahan mengukur ketidakpastian yang terkait dengan estimasi yang diperoleh dari sampel. Ini mewakili rentang di mana parameter populasi sebenarnya kemungkinan berada, mengingat statistik sampel yang diamati. Margin kesalahan biasanya dinyatakan sebagai persentase dan bergantung pada ukuran sampel, tingkat kepercayaan, dan variabilitas populasi.

Interval Kepercayaan

Interval kepercayaan adalah rentang nilai di mana parameter populasi sebenarnya kemungkinan berada, dengan tingkat kepercayaan tertentu. Misalnya, interval kepercayaan

95% berarti bahwa jika sampel berulang diambil dan interval kepercayaan dihitung untuk setiap sampel, 95% dari interval ini akan berisi parameter populasi sebenarnya.

Kesalahan Standar

Kesalahan standar adalah ukuran variabilitas statistik sampel, seperti rata-rata atau proporsi, di berbagai sampel yang diambil dari populasi yang sama. Kesalahan standar digunakan untuk menghitung interval kepercayaan dan berbanding terbalik dengan ukuran sampel—seiring bertambahnya ukuran sampel, kesalahan standar menurun, menghasilkan interval kepercayaan yang lebih sempit dan estimasi yang lebih tepat.

4.4 INTERVAL KEPERCAYAAN

Interval kepercayaan (CI) adalah konsep fundamental dalam statistik dan sangat penting untuk membuat keputusan yang tepat berdasarkan data sampel. Interval kepercayaan menawarkan rentang nilai yang kemungkinan besar mengandung parameter populasi yang diminati. Bagian ini memberikan penjelasan mendalam tentang interval kepercayaan, khususnya berfokus pada signifikansinya dalam memperkirakan parameter populasi.

Memperkirakan Rata-rata Populasi

Saat melakukan pembelajaran atau meng pembahasan data, kita sering menggunakan sampel untuk membuat kesimpulan tentang seluruh populasi. Rata-rata sampel adalah estimasi titik yang baik dari rata-rata populasi, tetapi akan lebih bermanfaat untuk memberikan rentang di mana kita percaya rata-rata populasi yang sebenarnya berada.

Rumus:

$$\bar{x} \pm z \left(\frac{\sigma}{\sqrt{n}} \right),$$

di mana

- $\bar{x}$ = rata-rata sampel
- z = nilai z , yang sesuai dengan tingkat kepercayaan yang diinginkan (misalnya, untuk tingkat kepercayaan 95%, kira-kira 1,96)
- σ = simpangan baku populasi
- n = ukuran sampel

Dalam banyak skenario dunia nyata, simpangan baku populasi (σ) tidak diketahui. Dalam kasus seperti itu, saat mencoba memperkirakan rata-rata populasi dari sampel, kita mengandalkan simpangan baku sampel (s) sebagai perkiraan untuk σ . Alih-alih distribusi z , yang mengasumsikan simpangan baku populasi diketahui, kita beralih ke distribusi t , yang lebih cocok untuk situasi ini.

Distribusi t mirip dengan distribusi z dalam bentuknya tetapi memiliki ekor yang lebih tebal. Ini membuatnya lebih mengakomodasi variabilitas yang diharapkan saat memperkirakan rata-rata populasi dan simpangan baku dari sampel. Rumus untuk memperkirakan rata-rata populasi menggunakan distribusi t adalah:

$$\bar{x} \pm t \left(\frac{s}{\sqrt{n}} \right),$$

di mana

- $\bar{x}$ = rata-rata sampel
- t = nilai t , yang sesuai dengan tingkat kepercayaan dan derajat kebebasan (df) yang diinginkan. Derajat kebebasan untuk pengujian ini adalah $n-1$. Misalnya, untuk tingkat kepercayaan 95% dan ukuran sampel 30, Anda akan merujuk ke tabel t untuk menemukan nilai t yang sesuai.
- s = simpangan baku sampel (digunakan sebagai estimasi untuk σ)
- n = ukuran sampel

Poin penting. Alasan menggunakan distribusi t daripada distribusi z ketika simpangan baku populasi tidak diketahui adalah untuk memberikan rentang (interval kepercayaan) yang lebih akurat untuk rata-rata populasi. Karena simpangan baku sampel mungkin bukan estimasi yang sempurna untuk simpangan baku populasi, distribusi t mengkompensasi ketidakpastian ini, terutama ketika ukuran sampel kecil. Seiring bertambahnya ukuran sampel, distribusi t mendekati bentuk distribusi z .

Margin Kesalahan dan Tingkat Kepercayaan

Margin kesalahan (MOE) memberikan jumlah di mana kita mengharapkan estimasi sampel kita bervariasi dari nilai populasi sebenarnya. Semakin besar MOE, semakin kurang akurat estimasi kita.

Rumus:

$$MOE = z \left(\frac{\sigma}{\sqrt{n}} \right),$$

di mana komponen-komponennya didefinisikan seperti sebelumnya.

Tingkat kepercayaan menentukan probabilitas bahwa metode konstruksi interval akan mencakup parameter populasi. Tingkat kepercayaan umum adalah 90%, 95%, dan 99%. Pilihan tingkat kepercayaan memengaruhi lebar CI: tingkat kepercayaan yang lebih tinggi menghasilkan interval yang lebih lebar.

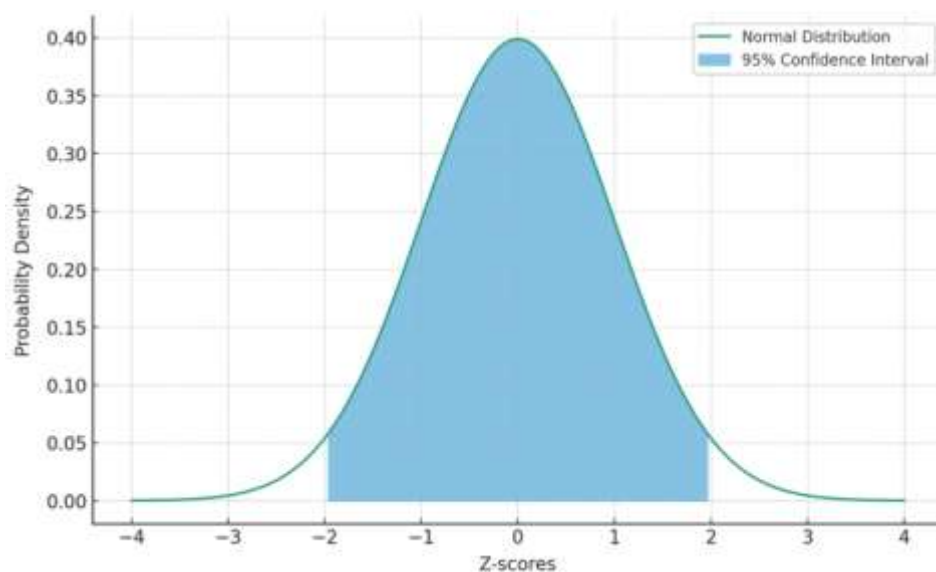
Menginterpretasikan Interval Kepercayaan

Memahami cara menginterpretasikan interval kepercayaan dengan benar sangat penting untuk membuat keputusan yang tepat.

- **Rentang interval.** Jika CI 95% untuk rata-rata populasi adalah (50, 60), itu berarti kita 95% yakin bahwa rata-rata populasi sebenarnya terletak di antara 50 dan 60.
- **Bukan pernyataan probabilitas tentang parameter.** Kesalahpahaman umum adalah berpikir bahwa ada peluang 95% bahwa rata-rata sebenarnya terletak di dalam interval. Ini salah. Rata-rata sebenarnya terletak di dalam interval atau tidak. Tingkat kepercayaan mengacu pada metode konstruksi interval. Jika kita mengulangi proses pengambilan sampel berkali-kali, kita memperkirakan sekitar 95% dari interval yang dibangun akan mengandung rata-rata populasi.

- **Interval yang lebih lebar versus interval yang lebih sempit.** Interval yang lebih lebar menunjukkan lebih banyak ketidakpastian tentang parameter populasi, sementara interval yang lebih sempit menunjukkan presisi yang lebih besar. Namun, untuk mendapatkan interval yang lebih sempit mungkin memerlukan peningkatan ukuran sampel.
- **Implikasi praktis.** Dalam konteks bisnis, lebar interval kepercayaan (CI) dapat memengaruhi pengambilan keputusan. Misalnya, jika ROI kampanye pemasaran memiliki CI 95% antara 5% dan 15%, ada lebih banyak ketidakpastian dibandingkan dengan CI antara 8% dan 10%.

Gambar 4.3 menyajikan kurva lonceng (distribusi normal) yang menunjukkan interval kepercayaan dan menyoroti wilayah di bawah kurva. Visual ini membantu dalam menggambarkan rentang interval, arti interval yang lebih lebar versus interval yang lebih sempit, dan implikasi praktis dari interval ini dalam pengambilan keputusan.



Gambar 4.3 Distribusi Normal Interval Kepercayaan dan Penyorotan Wilayah di Bawah Kurva.

Interval kepercayaan berfungsi sebagai jembatan antara statistik sampel dan parameter populasi, memberikan rentang di mana kita percaya nilai sebenarnya berada. Interval kepercayaan (CI) menggabungkan estimasi titik sampel, variabilitas data, dan tingkat kepercayaan yang kita inginkan. Dengan memahami dan menafsirkan CI dengan benar, para profesional di berbagai bidang, termasuk pemasaran, dapat membuat keputusan yang lebih tepat dan berbasis bukti.

Contoh Praktis: Interval Kepercayaan dalam Evaluasi Kampanye Pemasaran

Mari kita pertimbangkan kembali perusahaan e-commerce kita, ShopStream, yang berencana meluncurkan kampanye pemasaran baru. Sebelum meluncurkannya dalam skala besar, mereka memutuskan untuk mengujinya pada sampel 1.000 pelanggan. Tujuannya adalah untuk memperkirakan peningkatan rata-rata penjualan akibat kampanye dan untuk membuat proyeksi untuk basis pelanggan yang lebih luas.

Pengumpulan Data. Setelah kampanye, ShopStream mencatat peningkatan penjualan untuk pelanggan yang disampel. Mereka menemukan peningkatan rata-rata sampel ($\bar{x}$) sebesar Rp500.000 dan simpangan baku sampel (s) sebesar Rp150.000.

Membangun Interval Kepercayaan Menggunakan Distribusi T. Karena simpangan baku populasi tidak diketahui, ShopStream menggunakan distribusi t. Dengan 1.000 pelanggan, derajat kebebasannya adalah $df = 1.000 - 1 = 999$. Dengan merujuk pada tabel t untuk tingkat kepercayaan 95%, mereka menemukan nilai t mendekati 1,96 (sangat mirip dengan nilai z karena ukuran sampel yang besar).

Menggunakan rumus:

$$CI = \bar{x} \pm t \left(\frac{s}{\sqrt{n}} \right)$$

$$CI = 50 \pm 1.96 \left(\frac{15}{\sqrt{1000}} \right)$$

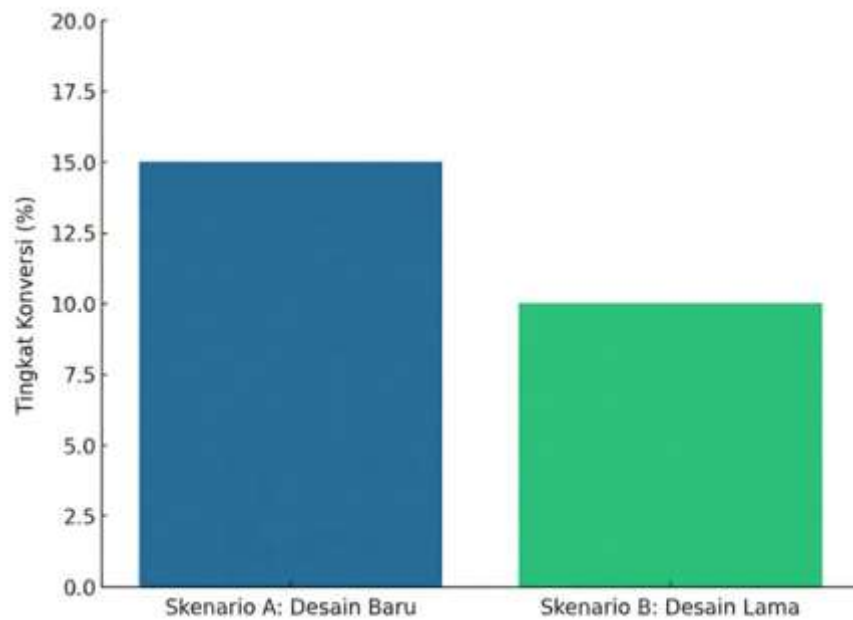
$$CI = (49.08.50.92)$$

Interpretasi. Interval kepercayaan 95% untuk peningkatan penjualan akibat kampanye pemasaran berada antara Rp490.800 dan Rp509.200. Ini berarti ShopStream 95% yakin bahwa peningkatan rata-rata penjualan untuk seluruh basis pelanggan, akibat kampanye tersebut, akan berada dalam kisaran ini.

Keputusan bisnis. Dengan interval yang sempit ini, tim pemasaran ShopStream dapat dengan yakin memprediksi dampak kampanye pada basis pelanggan yang lebih luas. Jika peningkatan yang diproyeksikan ini selaras dengan target pengembalian investasi (ROI) mereka, mereka dapat memutuskan untuk menerapkan kampanye tersebut untuk semua pelanggan. Contoh praktis ini menunjukkan penerapan interval kepercayaan dalam membuat keputusan bisnis yang terinformasi. Dengan menguji kampanye mereka pada sampel terlebih dahulu, ShopStream mampu mengukur potensi hasil tanpa berkomitmen penuh, sehingga mengoptimalkan sumber daya, dan memastikan efektivitas kampanye.

4.5 PENGUJIAN A/B DALAM PEMASARAN

Pengujian A/B, terkadang dikenal sebagai pengujian terpisah, telah menjadi alat utama dalam dunia pemasaran. Dengan semakin meningkatnya penekanan pada pengambilan keputusan berbasis data, pengujian ini menawarkan metode ilmiah untuk menguji dan mengoptimalkan berbagai upaya pemasaran. Pada bagian ini, kita akan membahas secara mendalam ranah pengujian A/B, khususnya yang disesuaikan dengan pemasaran. Gambar 4.4 menawarkan bantuan visual dalam hal ini, menyajikan grafik batang yang membandingkan hasil dari dua skenario pengujian A/B yang berbeda. Perbandingan ini tidak hanya menggambarkan efektivitas setiap skenario tetapi juga menggarisbawahi pentingnya pengujian yang dirancang dengan cermat dalam memperoleh wawasan yang bermakna.



Gambar 4.4 Membandingkan Hasil Dua Skenario Pengujian A/B yang Berbeda.

Dasar dan Pentingnya

Pengujian A/B adalah metode membandingkan dua versi (A dan B) dari halaman web, iklan, atau aset pemasaran lainnya untuk menentukan mana yang berkinerja lebih baik dalam mencapai tujuan tertentu, seperti meningkatkan rasio klik-tayang, penjualan, atau metrik konversi lainnya.

Signifikansi pengujian A/B dalam pemasaran modern tidak dapat dilebih-lebihkan, karena menawarkan pendekatan metodis untuk meningkatkan strategi pemasaran di berbagai dimensi. Aspek-aspek penting dari pengujian A/B meliputi:

- **Keputusan berbasis data.** Alih-alih mengandalkan intuisi, pemasar dapat menggunakan bukti empiris untuk memutuskan versi aset pemasaran mana yang lebih efektif.
- **Optimalkan kampanye.** Melalui pengujian berulang, pemasar dapat terus menyempurnakan dan meningkatkan kampanye mereka, yang mengarah pada peningkatan ROI.
- **Kurangi risiko.** Sebelum berkomitmen pada perubahan atau kampanye skala besar, pemasar dapat menguji variasi untuk memastikan mereka bergerak ke arah yang benar.

Desain Eksperimental untuk Uji A/B

- **Kontrol dan variasi.** Untuk uji A/B, satu versi bertindak sebagai kontrol, dan versi lainnya adalah variasi dengan perubahan yang diusulkan.
- **Penugasan acak.** Sangat penting untuk menugaskan pengguna secara acak ke kelompok kontrol atau variasi. Ini memastikan bahwa kelompok-kelompok tersebut sebanding, dan setiap perbedaan yang diamati dapat dikaitkan dengan perubahan yang dilakukan daripada faktor eksternal.

- **Ukuran sampel.** Ukuran sampel dapat memengaruhi keandalan hasil. Sampel yang lebih besar dapat memberikan hasil yang lebih akurat, tetapi juga penting untuk memastikan ukuran sampel minimum untuk mendeteksi perbedaan yang berarti.
- **Durasi.** Menjalankan uji untuk durasi yang sesuai memastikan bahwa hasil tidak bias karena efek hari dalam seminggu atau faktor jangka pendek lainnya.

Tabel 4.2 memberikan garis besar sederhana dari desain eksperimental untuk pengujian A/B, termasuk variabel yang terlibat, hasil yang diharapkan, dan metrik utama yang digunakan untuk pengukuran.

Tabel 4.2 Desain Eksperimental Sederhana untuk Pengujian A/B Termasuk Variabel, Hasil, dan Metrik.

	Variabel	Hasil	Metrik Utama
Tes 1	Desain situs web (lama vs. baru)	Peningkatan keterlibatan situs web	Waktu yang dihabiskan di situs web, rasio pantulan
Tes 2	Penempatan iklan (atas vs. bawah)	Tingkat klik iklan	Jumlah klik, tayangan
Tes 3	Teks ajakan bertindak (beli sekarang vs. tambahkan ke keranjang)	Peningkatan pembelian produk	Tingkat konversi, tingkat pengabaian keranjang belanja

Menyiapkan Tes A/B: Panduan Langkah demi Langkah

Menyiapkan tes A/B yang sukses membutuhkan perencanaan yang cermat dan perhatian terhadap detail. Berikut adalah panduan langkah demi langkah yang terstruktur untuk membantu Anda menavigasi prosesnya:

1. **Tentukan tujuan.** Sebelum melakukan hal lain, pahami dengan jelas apa yang ingin Anda capai dengan tes ini. Ini bisa berupa peningkatan rasio pembukaan email, peningkatan penjualan produk, atau peningkatan keterlibatan pengguna pada halaman web tertentu.
2. **Identifikasi variabel.** Tentukan elemen atau fitur spesifik yang ingin Anda uji. Ini bisa berupa warna tombol, teks situs web, dan gambar produk hingga baris subjek email.
3. **Kembangkan hipotesis.** Rumuskan hipotesis yang jelas berdasarkan tujuan Anda dan variabel yang dipilih. Misalnya, "Mengubah tombol ajakan bertindak dari biru menjadi merah akan meningkatkan rasio klik-tayang."
4. **Pilih alat Anda.** Tergantung pada platform dan skalanya, Anda mungkin menggunakan alat seperti Google Optimize, Optimizely, VWO, atau lainnya untuk tes berbasis web. Untuk kampanye email, platform seperti Mailchimp atau HubSpot mungkin cocok.
5. **Segmentasikan audiens Anda.** Bagi audiens Anda menjadi dua kelompok. Satu kelompok (kelompok kontrol) akan melihat versi saat ini, dan kelompok lainnya (kelompok variasi) akan melihat versi baru.
6. **Penugasan acak.** Pastikan pengguna ditugaskan secara acak ke setiap kelompok untuk menghindari bias seleksi.

7. **Tetapkan durasi pengujian dan ukuran sampel.** Sebelum memulai pengujian, hitung ukuran sampel yang dibutuhkan untuk mencapai signifikansi statistik. Selain itu, tentukan durasi pengujian, pastikan Anda menangkap siklus bisnis lengkap.
8. **Luncurkan pengujian.** Setelah semuanya siap, luncurkan pengujian Anda. Pastikan pemantauan waktu nyata untuk memeriksa anomali atau masalah apa pun.
9. **Pembahasan hasil.** Di akhir periode pengujian, kumpulkan dan pembahasan data. Hitung metrik seperti rasio konversi untuk kedua kelompok, perbedaan rasio tersebut, dan signifikansi statistik dari perbedaan tersebut.
10. **Tarik kesimpulan.** Berdasarkan hasil pembahasan, putuskan apakah hipotesis tersebut didukung atau disangkal.
11. **Terapkan pembelajaran.** Jika versi baru berkinerja lebih baik daripada versi lama dan Anda mencapai signifikansi statistik, pertimbangkan untuk menerapkan perubahan tersebut. Jika pengujian tidak meyakinkan atau versi baru tidak berkinerja baik, gunakan wawasan yang diperoleh untuk menginformasikan pengujian di masa mendatang.
12. **Dokumentasikan semuanya.** Untuk referensi di masa mendatang dan untuk membangun kurva pembelajaran organisasi, dokumentasikan pengaturan pengujian, hipotesis, hasil, dan poin-poin penting.
13. **Ulangi prosesnya.** Keindahan pengujian A/B terletak pada sifat iteratifnya. Gunakan wawasan dari satu pengujian untuk menginformasikan pengujian di masa mendatang, terus meningkatkan dan mengoptimalkan upaya pemasaran Anda.

Pengaturan yang tepat sangat penting untuk keberhasilan pengujian A/B. Ini memastikan bahwa hasil yang diperoleh valid, dapat ditindaklanjuti, dan selaras dengan tujuan bisnis. Dengan mengikuti langkah-langkah yang diuraikan secara sistematis, pemasar dapat memanfaatkan kekuatan pengujian A/B untuk membuat keputusan yang berdasarkan data dan informasi.

Signifikansi Statistik dalam Uji A/B

Signifikansi statistik menunjukkan seberapa yakin kita bahwa hasil yang diamati dalam uji A/B bukan disebabkan oleh kebetulan semata.

- Nilai P. Ini adalah metrik yang umum digunakan dalam pengujian A/B. Nilai p yang rendah (biasanya $< 0,05$) menunjukkan bahwa hasilnya signifikan secara statistik.
- Kesalahan Tipe I dan Tipe II. Penting untuk menyadari potensi positif palsu (percaya ada efek padahal tidak ada) dan negatif palsu (percaya tidak ada efek padahal ada).
- Kekuatan uji. Ini mengacu pada probabilitas mendeteksi perbedaan jika ada. Kekuatan standar yang diinginkan adalah 0,8 atau 80%.

Teknik Pengujian A/B Tingkat Lanjut

Seiring bisnis menjadi lebih berbasis data dan lanskap digital berkembang, ranah pengujian A/B telah menyaksikan pengenalan teknik yang lebih canggih. Mari kita selami lebih dalam metode-metode canggih ini dan pahami perbedaannya dengan pendekatan pengujian A/B dasar.

Pengujian Multivariat

Apa itu? Tidak seperti pengujian A/B, di mana hanya satu variabel yang diubah pada satu waktu, pengujian multivariat (MVT) melibatkan pengujian beberapa perubahan/variasi secara bersamaan untuk melihat kombinasi mana yang menghasilkan hasil terbaik.

Bagaimana perbedaannya dengan pengujian A/B? Meskipun pengujian A/B membandingkan versi A dengan versi B, MVT mungkin membandingkan kombinasi versi A1, B1, C1 dengan A2, B2, C2, dan seterusnya, mengeksplorasi interaksi antar variabel.

Aplikasi. Misalnya, jika situs e-commerce ingin menguji warna tombol ajakan bertindak dan teks di dalamnya secara bersamaan, MVT akan menilai berbagai kombinasi warna dan teks untuk mengidentifikasi campuran yang paling efektif.

Pengujian Sekuensial

Apa itu? Pengujian sekuensial adalah metode dinamis di mana ukuran sampel tidak ditetapkan sebelumnya. Pengujian terus dipantau, dan berdasarkan hasil yang masuk, keputusan untuk berhenti (karena temuan signifikan) atau melanjutkan dibuat.

Bagaimana perbedaannya dengan pengujian A/B? Pengujian A/B tradisional memiliki ukuran sampel dan durasi yang telah ditentukan sebelumnya. Dalam pengujian sekuensial, pengujian mungkin berakhir lebih awal jika pemenang yang jelas diidentifikasi atau mungkin diperpanjang jika hasilnya tetap tidak meyakinkan.

Aplikasi. Jika sebuah perusahaan sedang menguji fitur situs web yang penting, dan hasil awal sudah menunjukkan peningkatan yang signifikan dengan varian baru, mereka mungkin memutuskan untuk menghentikan pengujian dan menerapkan perubahan lebih cepat, menghemat sumber daya dan waktu.

Pengujian A/B Bayesian

Apa itu? Pengujian A/B Bayesian adalah pendekatan yang memperbarui probabilitas hipotesis menjadi benar seiring tersedianya lebih banyak data, memberikan pembahasan yang lebih intuitif dan fleksibel.

Bagaimana perbedaannya dengan pengujian A/B? Pengujian A/B tradisional, berdasarkan statistik frekuensi, memberikan nilai p yang menunjukkan apakah ada perbedaan yang signifikan secara statistik. Namun, pengujian Bayesian memberikan distribusi probabilitas, yang menunjukkan seberapa besar kemungkinan hasil atau ukuran efek tertentu.

Aplikasi. Jika tim pemasaran ingin memahami potensi dampak dari dua desain iklan, pendekatan Bayesian akan memberi tahu mereka tidak hanya apakah satu iklan lebih baik, tetapi seberapa jauh lebih baiknya dan tingkat kepastian dari perkiraan tersebut.

Personalisasi dan Pengujian Berbasis Segmen

Apa itu? Pendekatan ini berfokus pada pembuatan pengujian yang disesuaikan untuk segmen audiens tertentu alih-alih memperlakukan seluruh audiens sebagai satu entitas tunggal.

Bagaimana perbedaannya dengan pengujian A/B? Meskipun pengujian A/B dasar mungkin memberikan hasil untuk pengguna rata-rata, pengujian berbasis segmen menggali lebih dalam, mengungkap wawasan untuk kelompok tertentu seperti pengunjung baru, pelanggan tetap, atau pengguna dari lokasi tertentu.

Aplikasi. Layanan streaming dapat melakukan pengujian berbasis segmen untuk memahami preferensi konten. Alih-alih menggeneralisasi bahwa UI baru lebih baik untuk semua pengguna, mereka mungkin menemukan bahwa pengguna yang lebih muda lebih menyukai satu gaya, sedangkan pengguna yang lebih tua memiliki kecenderungan yang berbeda.

Kesimpulan. Meskipun pengujian A/B tradisional menawarkan wawasan yang berharga, teknik-teknik canggih ini memungkinkan pemahaman yang lebih bernuansa, mengoptimalkan berbagai aspek kampanye dan memenuhi kebutuhan audiens yang beragam. Seiring pertumbuhan bisnis dan perluasan kumpulan data, mengintegrasikan metodologi canggih ini dapat menghasilkan strategi pemasaran yang lebih canggih, efektif, dan personal.

Potensi Jebakan dalam Pengujian A/B

Meskipun pengujian A/B adalah metodologi yang ampuh, ada jebakan umum yang perlu diwaspadai pemasar untuk memastikan hasil yang valid dan dapat ditindaklanjuti:

- **Efek carryover.** Terkadang, pengguna yang terpapar satu versi (misalnya, Versi A) mungkin kemudian terpapar versi lain (Versi B), yang menyebabkan potensi bias dalam perilaku mereka.
- **Efek kebaruan.** Desain atau fitur baru mungkin awalnya berkinerja lebih baik hanya karena baru dan menarik perhatian, bukan karena secara inheren lebih baik.
- **Faktor eksternal.** Peristiwa di luar pengujian, seperti hari libur, peristiwa berita, atau masalah teknis, dapat memengaruhi hasil.
- **Mengintip terlalu dini.** Sangat menggoda untuk menghentikan pengujian lebih awal ketika hasilnya tampak menjanjikan, tetapi ini dapat menyebabkan kesimpulan yang salah. Pengujian harus dijalankan sepenuhnya untuk memastikan validitas statistik.
- **Menguji terlalu banyak elemen sekaligus.** Meskipun pengujian multivariat dapat bermanfaat, menguji terlalu banyak perubahan secara bersamaan dapat menyulitkan untuk menentukan perubahan mana yang menyebabkan hasil yang diamati.
- **Mengabaikan siklus bisnis.** Tidak mempertimbangkan siklus bisnis mingguan atau bulanan dapat menyebabkan data yang bias. Misalnya, pengecer online mungkin melihat perilaku yang berbeda pada hari kerja dibandingkan dengan akhir pekan.

Menginterpretasikan Hasil Uji A/B

Setelah pengujian selesai, interpretasi dan penerapan hasil sangat penting:

- **Ukuran efek.** Bahkan jika hasilnya signifikan secara statistik, sangat penting untuk menentukan apakah perbedaannya signifikan secara praktis. Misalnya, peningkatan 0,01% dalam rasio klik-tayang mungkin tidak sepadan dengan investasi, meskipun signifikan secara statistik.
- **Interval kepercayaan.** Alih-alih hanya melihat estimasi titik, periksa rentang di mana metrik sebenarnya berada dengan tingkat kepercayaan tertentu. Ini memberikan pandangan yang lebih holistik tentang potensi hasil.
- **Kontekstualisasikan hasil.** Selalu interpretasikan hasil dalam konteks bisnis dan audiens. Perubahan yang berhasil untuk satu demografi atau produk mungkin tidak berhasil untuk yang lain.

- **Tes lanjutan.** Jika hasil tes tidak terduga atau bertentangan dengan intuisi, pertimbangkan untuk menjalankan tes lanjutan untuk memvalidasi temuan atau mengeksplorasi fenomena tersebut lebih lanjut.
- **Dokumentasikan dan bagikan.** Memelihara repositori pengujian masa lalu, desainnya, hasilnya, dan pelajaran yang dipetik dapat sangat berharga untuk kampanye di masa mendatang dan untuk mendidik tim yang lebih luas.

Pengujian A/B, jika dieksekusi dengan benar, dapat mengungkap wawasan mendalam tentang perilaku, preferensi, dan motivasi pelanggan. Meskipun ini adalah proses teknis, ini juga merupakan seni yang mengharuskan pemasar untuk menyeimbangkan ketelitian statistik dengan pemahaman intuitif tentang audiens dan konteks pasar mereka. Dengan menyadari potensi jebakan, memperbarui metodologi pengujian secara teratur, dan menempatkan hasil dalam konteks bisnis yang tepat, pemasar dapat secara signifikan meningkatkan strategi dan eksekusi mereka.

Singkatnya, pengujian A/B adalah alat ampuh dalam persenjataan pemasar. Dengan menerapkan teknik ilmiah yang ketat, pemasar dapat memastikan bahwa keputusan mereka didasarkan pada data, mengurangi risiko, dan meningkatkan pengembalian investasi pemasaran mereka.

4.6 PENGUJIAN HIPOTESIS DALAM PEMASARAN

Pengantar Pengujian Hipotesis

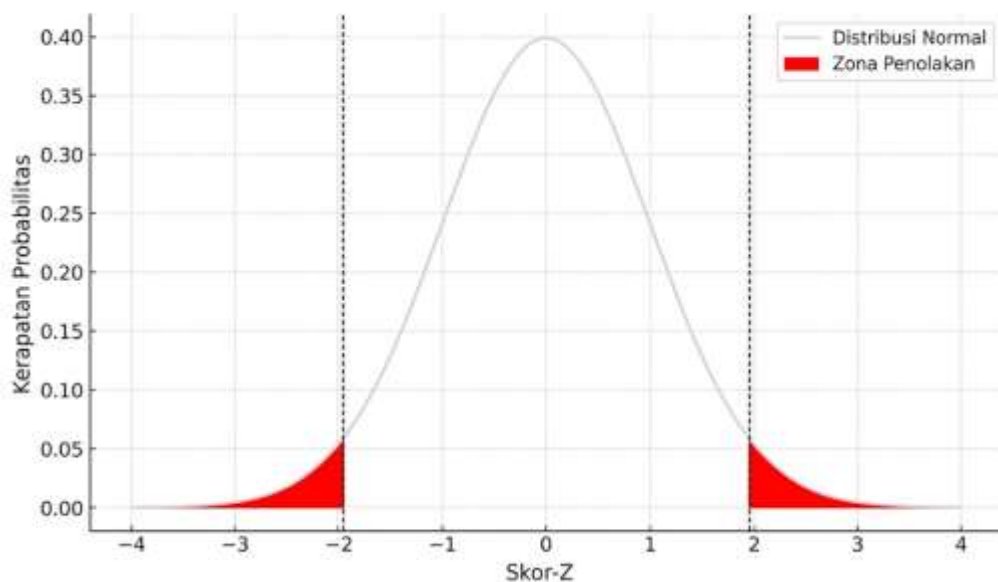
Pengujian hipotesis adalah teknik fundamental dalam statistik inferensial yang memungkinkan peneliti untuk mengevaluasi validitas asumsi mereka dan membuat keputusan berdasarkan data. Dalam konteks pemasaran, pengujian hipotesis dapat digunakan untuk menilai efektivitas kampanye pemasaran, membandingkan segmen pelanggan, atau mengevaluasi dampak perubahan harga, di antara aplikasi lainnya.

Proses pengujian hipotesis melibatkan perumusan hipotesis nol (H_0) dan hipotesis alternatif (H_1). Hipotesis nol mewakili asumsi tidak adanya efek atau hubungan antara variabel, dan hipotesis alternatif menyatakan bahwa ada efek atau hubungan. Misalnya, seorang pemasar mungkin ingin menilai apakah kampanye promosi baru memiliki dampak positif pada penjualan. Dalam hal ini, hipotesis nol akan menyatakan bahwa tidak ada perbedaan penjualan antara kampanye baru dan kampanye lama, dan hipotesis alternatif akan menyatakan bahwa ada perbedaan.

Untuk menilai hipotesis, peneliti menggunakan data sampel dan menghitung statistik uji, seperti statistik t atau statistik chi-kuadrat, yang mengukur perbedaan antara data yang diamati dan hipotesis nol. Statistik uji kemudian dibandingkan dengan nilai kritis, yang ditentukan berdasarkan tingkat signifikansi yang dipilih (α) dan distribusi probabilitas yang terkait dengan statistik uji. Jika statistik uji melebihi nilai kritis, hipotesis nol ditolak dan hipotesis alternatif diterima, yang menunjukkan bahwa ada bukti yang mendukung klaim bahwa kampanye pemasaran baru berdampak pada penjualan.

Pengujian hipotesis telah banyak digunakan dalam riset dan praktik pemasaran untuk mengevaluasi strategi pemasaran dan membuat keputusan berdasarkan data. Dengan

menerapkan teknik pengujian hipotesis, pemasar dapat memperoleh wawasan berharga dan mengoptimalkan upaya pemasaran mereka, yang pada akhirnya mengarah pada peningkatan hasil dan peningkatan ROI. Gambar 4.5 memberikan representasi grafis dari pengujian hipotesis dua arah, dengan wilayah yang disorot menunjukkan zona penolakan dan non-penolakan. Alat bantu visual ini sangat berguna untuk menggambarkan bagaimana statistik uji dibandingkan dengan nilai kritis untuk menentukan apakah akan menolak atau mempertahankan hipotesis nol.



Gambar 4.5 Uji Hipotesis Dua Arah dengan Wilayah yang Disorot Menunjukkan Zona Penolakan dan Non-Penolakan.

Uji Hipotesis Umum dalam Pemasaran

Dalam pemasaran, berbagai uji hipotesis digunakan untuk membahas data dan memperoleh wawasan. Beberapa uji hipotesis yang paling umum meliputi uji t, uji chi-kuadrat, ANOVA, dan uji korelasi dan regresi. Pada bagian ini, kita akan membahas secara singkat masing-masing uji ini dan aplikasinya dalam pemasaran, dengan referensi yang relevan (lihat Tabel 4.3).

Tabel 4.3 Uji Hipotesis Umum dalam Pemasaran dengan Aplikasi dan Asumsinya.

Uji Hipotesis	Aplikasi	Asumsi
Uji t	Membandingkan rata-rata dua kelompok (misalnya, kelompok kontrol vs. kelompok perlakuan)	Data terdistribusi normal, varians sama
Uji chi-kuadrat	Menguji hubungan antar variabel kategorikal (misalnya, jenis iklan vs. rasio klik-tayang)	Frekuensi yang diamati cukup besar
Pembahasan varians	Membandingkan rata-rata tiga kelompok atau lebih	Terdistribusi normal, varians sama, pengamatan independen

Pembahasan regresi	Memprediksi nilai suatu variabel berdasarkan nilai satu atau lebih variabel lainnya	Hubungan linier, normalitas multivariat, tidak ada multikolinearitas

Uji T

Uji T adalah keluarga uji statistik yang digunakan untuk membandingkan rata-rata dua kelompok. Dalam pemasaran, uji t dapat diterapkan untuk membandingkan rata-rata penjualan atau skor kepuasan pelanggan antara dua kampanye pemasaran atau segmen pelanggan yang berbeda. Terdapat beberapa jenis uji t, termasuk uji t sampel independen, uji t sampel berpasangan, dan uji t satu sampel, masing-masing dirancang untuk skenario pembelajaran tertentu.

Uji Chi-kuadrat

Uji chi-kuadrat adalah uji non-parametrik yang digunakan untuk memeriksa hubungan antara dua variabel kategorikal. Dalam pemasaran, uji chi-kuadrat dapat digunakan untuk membahas hubungan antara demografi pelanggan (misalnya, usia, jenis kelamin, pendapatan) dan preferensi mereka terhadap produk atau layanan tertentu. Statistik uji, chi-kuadrat (χ^2), dihitung berdasarkan frekuensi yang diamati dan diharapkan dalam tabel kontingensi dan dibandingkan dengan nilai kritis untuk menentukan signifikansi hubungan tersebut.

Pembahasan Varians

ANOVA adalah teknik statistik yang digunakan untuk membandingkan rata-rata tiga kelompok atau lebih. Dalam pemasaran, ANOVA dapat digunakan untuk membahas efektivitas beberapa kampanye pemasaran, strategi penetapan harga, atau penawaran promosi. ANOVA menguraikan total variasi dalam data menjadi variasi antar kelompok dan dalam kelompok dan menghitung rasio F untuk menguji hipotesis nol bahwa semua rata-rata kelompok adalah sama.

Uji Korelasi dan Regresi

Uji korelasi dan regresi adalah alat pembahasan yang ampuh yang mengungkap hubungan antara variabel kontinu. Dalam pembelajaran pemasaran, uji ini tidak hanya memastikan hubungan tetapi juga memprediksi hasil di masa depan berdasarkan berbagai faktor yang memengaruhi.

Pembahasan korelasi. Koefisien korelasi Pearson (r) mengukur kekuatan dan arah hubungan linier antara dua variabel. Ini dapat sangat berguna dalam pemasaran untuk melakukan hal-hal berikut:

- Mengidentifikasi apakah ada hubungan antara pengeluaran iklan dan penjualan.
- Memahami korelasi antara skor kepuasan pelanggan dan pembelian berulang.

Namun, meskipun korelasi dapat mengungkapkan bahwa dua variabel bergerak bersama, hal itu tidak selalu menyiratkan sebab akibat. Korelasi yang tinggi antara pengeluaran iklan dan penjualan tidak membuktikan bahwa iklan tersebut menyebabkan peningkatan penjualan.

Pembahasan regresi. Regresi membawa ini selangkah lebih maju dengan membangun model prediktif. Dalam pemasaran, pembahasan regresi merupakan alat yang sangat

diperlukan karena kemampuannya untuk memprediksi hasil berdasarkan perubahan variabel prediktor.

- **Regresi linier sederhana.** Ini melibatkan satu variabel independen yang memprediksi variabel dependen, misalnya, memprediksi penjualan (variabel dependen) berdasarkan anggaran pemasaran (variabel independen).
- **Regresi berganda.** Di sini kita memodelkan hubungan antara variabel dependen dan beberapa variabel independen. Pertimbangkan sebuah perusahaan e-commerce. Mereka mungkin ingin memprediksi penjualan produk (variabel dependen) menggunakan beberapa prediktor seperti pengeluaran iklan online, jumlah ulasan, harga produk, dan lalu lintas situs web.

Penerapan Regresi dalam Pemasaran

- **Alokasi anggaran.** Dengan memahami aktivitas pemasaran mana (iklan online, kampanye influencer, pemasaran email) yang mendorong penjualan, bisnis dapat mengoptimalkan alokasi anggaran mereka untuk ROI maksimum.
- **Strategi penetapan harga.** Jika harga produk adalah salah satu variabel independen dalam model regresi, bisnis dapat membahas dampaknya terhadap penjualan dan dengan demikian menyempurnakan strategi penetapan harga.
- **Prediksi nilai seumur hidup pelanggan.** Dengan menggunakan regresi, perusahaan dapat memprediksi total laba bersih dari pelanggan sepanjang hubungan mereka. Prediktor dapat mencakup nilai pembelian rata-rata, frekuensi pembelian, dan umur pelanggan.
- **Rekomendasi produk.** Dengan membahas perilaku pembelian dan data pelanggan lainnya, model regresi dapat membantu memprediksi produk mana yang mungkin diminati pelanggan selanjutnya, mendorong strategi upsell dan cross-sell.

Kesimpulannya, meskipun korelasi memberikan wawasan tentang hubungan, regresi adalah alat yang ampuh dalam persenjataan pemasar untuk memprediksi hasil dan mengoptimalkan strategi. Ini bukan hanya tentang memahami apa yang terjadi tetapi juga memanfaatkan data untuk membentuk strategi dan keputusan di masa depan.

Tingkat Signifikansi dan Nilai P

Dalam pengujian hipotesis, tingkat signifikansi (α) dan nilai p memainkan peran penting dalam menentukan apakah akan menolak atau mempertahankan hipotesis nol. Konsep-konsep ini membantu peneliti mengukur kemungkinan memperoleh hasil yang diamati jika hipotesis nol benar. Pada bagian ini, kita akan membahas tingkat signifikansi dan nilai p secara detail, dengan referensi yang relevan.

Tingkat Signifikansi (α)

Tingkat signifikansi, yang dilambangkan dengan α , adalah probabilitas menolak hipotesis nol ketika hipotesis tersebut sebenarnya benar (kesalahan Tipe I). Tingkat signifikansi yang umum digunakan dalam pembelajaran adalah 0,05, 0,01, dan 0,001, yang mewakili probabilitas maksimum yang dapat diterima untuk melakukan kesalahan Tipe I. Tingkat signifikansi yang dipilih menentukan nilai kritis, yang digunakan untuk membandingkan

statistik uji. Jika statistik uji melebihi nilai kritis, hipotesis nol ditolak dan hipotesis alternatif diterima.

Nilai P

Nilai P , yang merupakan singkatan dari nilai probabilitas, mewakili probabilitas memperoleh statistik uji yang ekstrem atau lebih ekstrem daripada nilai yang diamati, dengan asumsi bahwa hipotesis nol benar. Nilai p yang lebih kecil menunjukkan bukti yang lebih kuat terhadap hipotesis nol, sedangkan nilai p yang lebih besar menunjukkan bukti yang lebih lemah. Untuk menentukan hasil uji hipotesis, nilai p dibandingkan dengan tingkat signifikansi yang dipilih (α). Jika nilai p kurang dari atau sama dengan α , hipotesis nol ditolak, yang menunjukkan bahwa hasil yang diamati signifikan secara statistik dan memberikan bukti yang mendukung hipotesis alternatif.

Konsep tingkat signifikansi dan nilai p sangat penting untuk pengambilan keputusan yang tepat dalam pengujian hipotesis. Dengan menetapkan tingkat signifikansi yang sesuai dan menafsirkan nilai p dengan benar, peneliti dapat mengendalikan risiko membuat kesimpulan yang salah dan meningkatkan keandalan temuan mereka. Dalam pemasaran, pemahaman konsep-konsep ini sangat penting untuk mengevaluasi efektivitas strategi pemasaran dan membuat keputusan berbasis data.

4.7 SEGMENTASI DAN PENGOLAHAN PELANGGAN

Saat kita beralih dari pengujian hipotesis ke segmentasi pelanggan, mungkin tampak seolah-olah kita melompat dari satu ranah ke ranah lain. Namun, dalam jalinan rumit pemasaran berbasis data, kedua area ini saling terkait dan saling melengkapi.

Pembahasan inferensial sebagai fondasinya. Pada intinya, pembahasan inferensial, terutama melalui alat-alat seperti pengujian A/B, memungkinkan pemasar untuk menarik kesimpulan tentang strategi atau aset tertentu. Kita membuat hipotesis, menguji, dan kemudian menyimpulkan berdasarkan data sampel apa yang mungkin benar untuk seluruh populasi target kita. Ini membekali pemasar dengan pengetahuan tentang apa yang berhasil dan apa yang tidak dalam melibatkan audiens tertentu.

Masuklah segmentasi pelanggan. Meskipun pembelajaran inferensial membantu pemasar memahami efektivitas strategi tertentu, segmentasi pelanggan berfokus pada pemahaman audiens itu sendiri. Ini mengkategorikan beragam konsumen ke dalam kelompok yang lebih mudah dikelola dan koheren berdasarkan berbagai atribut seperti perilaku, kebutuhan, atau demografi. Tetapi mengapa ini relevan dalam konteks pembelajaran inferensial?

- **Hipotesis yang disesuaikan.** Dengan pemahaman yang lebih baik tentang segmen pelanggan yang berbeda, pemasar dapat merumuskan hipotesis yang lebih tepat. Misalnya, alih-alih membuat hipotesis tentang efektivitas strategi untuk seluruh basis pengguna, seseorang dapat fokus pada dampaknya pada segmen tertentu, seperti milenial atau pelanggan tetap.

- **Peningkatan relevansi pengujian.** Pembahasan inferensial, ketika diterapkan pada segmen tertentu, memastikan bahwa pengujian lebih relevan. Menguji fitur baru untuk pengguna yang melek teknologi mungkin menghasilkan hasil yang berbeda dibandingkan dengan mengujinya untuk pengguna yang tidak begitu mahir teknologi.
- **Mendorong personalisasi.** Karena pengujian memvalidasi strategi mana yang berhasil untuk segmen mana, hal itu membuka jalan bagi kampanye pemasaran yang dipersonalisasi. Alih-alih pendekatan satu ukuran untuk semua, bisnis dapat memenuhi kebutuhan dan preferensi unik dari setiap segmen.
- **Alokasi sumber daya yang dioptimalkan.** Dengan memahami segmen mana yang paling responsif terhadap strategi mana, bisnis dapat mengalokasikan sumber daya mereka secara lebih efektif. Ini memastikan ROI yang lebih tinggi dan meminimalkan pemborosan upaya pada segmen yang kurang responsif.

Pada intinya, meskipun pembelajaran inferensial memberi pemasar alat untuk menguji dan memvalidasi strategi, segmentasi pelanggan menawarkan lensa untuk melihat audiens mereka dengan cara yang lebih bernuansa. Bersama-sama, keduanya memberdayakan bisnis untuk tidak hanya melibatkan audiens mereka tetapi juga melibatkan mereka dengan cara yang paling efektif. Saat kita mempelajari lebih dalam tentang segmentasi pelanggan di bagian ini, ingatlah fondasi inferensial yang telah kita bangun, dan pertimbangkan bagaimana setiap segmen mungkin merespons secara berbeda terhadap berbagai strategi pemasaran.

Di era digital, memahami pelanggan pada tingkat granular sangat penting untuk keberhasilan pemasaran. Namun, dengan banyaknya data pelanggan yang tersedia, memproses dan mensegmentasikan informasi ini dengan cara yang bermakna menjadi upaya yang menantang. Segmentasi pelanggan adalah proses membagi basis pelanggan yang luas menjadi kelompok yang lebih mudah dikelola dan homogen, berdasarkan kriteria atau perilaku tertentu, untuk menyesuaikan upaya pemasaran secara efektif. Bab ini akan membahas berbagai metodologi yang ampuh untuk segmentasi pelanggan dan pemrosesan data. Kita akan mulai dengan memahami seluk-beluk pengelompokan k-means, metode partisi yang membagi data menjadi kelompok-kelompok yang berbeda.

Selanjutnya, kita akan mengeksplorasi struktur hierarki kelompok pelanggan melalui pengelompokan hierarki, yang memberikan pandangan bertingkat tentang segmen pelanggan. Terakhir, bab ini akan membedah pembahasan RFM (*recency, frequency, monetary*), metode segmentasi perilaku yang menawarkan lensa komprehensif ke dalam nilai dan keterlibatan pelanggan. Melalui teknik-teknik ini, pemasar dapat mencapai pemahaman yang lebih mendalam tentang audiens mereka, memastikan strategi pemasaran tepat dan berdampak.

Pengelompokan K-Means

Pengelompokan K-means adalah salah satu metode partisi paling populer dalam pembelajaran tanpa pengawasan, terutama dalam konteks pemasaran untuk segmentasi pelanggan. Ide inti di balik pengelompokan k-means adalah membagi dataset menjadi k jumlah cluster yang saling eksklusif di mana setiap titik data termasuk ke dalam cluster dengan mean terdekat.

Proses Pengelompokan K-Means

1. **Inisialisasi.** Pilih k pusat cluster awal (biasanya k titik data dari dataset).
2. **Penugasan.** Tetapkan setiap titik data ke pusat klaster terdekat.
3. **Perhitungan Ulang.** Hitung rata-rata baru untuk setiap klaster.
4. **Ulangi.** Ulangi langkah penugasan dan perhitungan ulang hingga penugasan klaster tidak lagi berubah.

Manfaat bagi Pemasar

Efisiensi segmentasi. Dengan menggunakan k-means, kumpulan data pelanggan yang besar dapat dengan cepat disegmentasikan berdasarkan karakteristik atau perilaku yang dipilih, membantu dalam pemasaran target.

Fleksibilitas. Pemasar dapat menentukan jumlah segmen pelanggan yang diinginkan (k) berdasarkan kebutuhan bisnis, meskipun penting untuk memilih k yang optimal menggunakan teknik seperti metode siku.

Pembuatan profil. Setelah klaster didefinisikan, pemasar dapat membuat profil setiap segmen untuk memahami karakteristik yang mendefinisikannya, mendorong upaya pemasaran yang dipersonalisasi.

Pertimbangan

Penskalasian fitur. Penskalasian fitur, yang mencakup teknik standardisasi dan normalisasi, merupakan langkah pra-pemrosesan yang penting ketika menggunakan algoritma seperti pengelompokan k-means. Alasan perlunya hal ini sangat terkait dengan cara kerja algoritma:

- **Algoritma berbasis jarak.** Pengelompokan k-means pada dasarnya adalah algoritma berbasis jarak, di mana titik data dikelompokkan berdasarkan jaraknya dari centroid suatu klaster. Jika satu fitur memiliki rentang nilai yang luas (misalnya, pendapatan tahunan berkisar dari ribuan hingga jutaan) dan fitur lain memiliki rentang yang sempit (misalnya, usia berkisar dari 20 hingga 70), algoritma mungkin akan memberikan bobot yang lebih besar pada fitur dengan rentang yang lebih luas. Pembobotan yang tidak proporsional ini dapat mendistorsi pembentukan klaster.
- **Keseragaman dalam pengukuran.** Fitur dalam suatu dataset dapat diukur dalam berbagai satuan (misalnya, kilogram, mil, dolar). Membandingkan secara langsung satuan-satuan ini tanpa penskalasian dapat menghasilkan klaster yang tidak bermakna. Dengan menskalakan fitur, kita memastikan fitur tersebut berada pada skala pengukuran yang seragam, sehingga memungkinkan perbandingan yang bermakna.
- **Meningkatkan konvergensi.** Banyak algoritma optimasi (seperti penurunan gradien) yang digunakan dalam pembelajaran mesin berkonvergensi lebih cepat dengan data yang distandarisasi atau dinormalisasi. Dalam konteks k-means, penskalaan fitur dapat membantu mencapai konvergensi yang lebih cepat ke centroid optimal.
- **Peningkatan interpretasi.** Setelah pengelompokan, jika Anda bertujuan untuk membuat profil atau memahami karakteristik setiap klaster, memiliki fitur yang diskalakan dapat memudahkan interpretasi. Pentingnya relatif setiap fitur dalam menentukan keanggotaan klaster menjadi lebih jelas.

Pada intinya, agar pengelompokan k-means menghasilkan hasil yang bermakna dan akurat, penting untuk memastikan bahwa setiap fitur berkontribusi secara proporsional terhadap perhitungan jarak. Penskalaan fitur memastikan keseimbangan ini, menjadikannya langkah yang diperlukan dalam proses pengelompokan.

- **Jebakan inisialisasi acak.** K-means dapat menghasilkan hasil yang berbeda tergantung pada pusat kluster awal. Hal ini dapat diatasi dengan menggunakan teknik seperti k-means++ untuk inisialisasi yang lebih cerdas.
- **Memilih k.** Menemukan jumlah kluster yang tepat sangat penting. Teknik seperti metode siku atau pembahasan siluet dapat membantu dalam penentuan ini.
- **Aplikasi dalam pemasaran.** Sebuah perusahaan ritel dapat menggunakan pengelompokan k-means untuk mensegmentasikan pelanggannya berdasarkan perilaku pembelian, frekuensi, dan preferensi produk. Dengan demikian, mereka dapat menyesuaikan penawaran promosi atau iklan khusus untuk karakteristik setiap segmen, memaksimalkan peluang respons pelanggan yang positif.

Pengelompokan Hierarki dalam Segmentasi Pelanggan

Pengelompokan hierarki adalah metode yang digunakan untuk membangun hierarki atau pohon kluster. Tidak seperti pengelompokan k-means, yang membagi data menjadi kluster yang berbeda tanpa urutan atau hubungan intrinsik, pengelompokan hierarki menetapkan hierarki kluster bertingkat. Ini dapat memberikan pandangan yang lebih rinci tentang segmentasi pelanggan.

Dasar-dasar Pengelompokan Hierarki

Pengelompokan hierarki beroperasi berdasarkan prinsip pengelompokan titik data yang serupa ke dalam kluster, memastikan titik data dalam satu kluster lebih mirip dibandingkan dengan yang ada di kluster lain. Ada dua pendekatan utama untuk ini:

- **Pengelompokan aglomeratif.** Ini adalah pendekatan bottom-up di mana setiap titik data dimulai sebagai kluster individual. Kluster-kluster ini kemudian digabungkan secara iteratif berdasarkan kesamaannya hingga hanya ada satu kluster besar yang berisi semua titik data.
- **Pengelompokan divisif.** Kebalikan dari pendekatan aglomeratif, pengelompokan divisif dimulai dengan satu kluster besar yang berisi semua titik data. Kemudian, kluster tersebut dibagi secara iteratif menjadi kluster yang lebih kecil berdasarkan ketidakmiripan, berlanjut hingga setiap titik data berdiri sendiri sebagai satu kluster tunggal.

Cara Kerjanya

- **Pembuatan matriks kesamaan.** Langkah pertama adalah menghitung matriks yang berisi jarak (atau ketidakmiripan) antara setiap pasangan titik data.
- **Pembentukan kluster.** Dengan menggunakan matriks jarak, kluster dibentuk menggunakan kriteria keterkaitan. Kriteria keterkaitan umum meliputi:
 - **Keterkaitan tunggal.** Jarak antara dua kluster adalah jarak terpendek antara dua titik data di setiap kluster.

- **Keterkaitan lengkap.** Jarak antara dua klaster adalah jarak terpanjang antara dua titik data di setiap klaster.
- **Keterkaitan rata-rata.** Jarak antara dua klaster adalah jarak rata-rata antara titik data di kedua klaster.
- **Pembuatan pohon (dendrogram).** Diagram seperti pohon yang disebut dendrogram dibuat, yang mewakili bagaimana klaster digabungkan (aglomeratif) atau dipisahkan (divisif).
- **Memotong dendrogram.** Dengan memotong dendrogram pada tingkat tertentu, pemasar dapat menentukan jumlah klaster yang mereka inginkan untuk segmentasi.

Penerapan dalam Pemasaran

Pengelompokan hierarkis menawarkan perspektif multi-resolusi segmentasi pelanggan. Misalnya, bisnis e-commerce dapat mengelompokkan pelanggan berdasarkan perilaku pembelian mereka:

- Pada tingkat yang lebih tinggi, mereka dapat mengidentifikasi kelompok seperti 'Pembeli Sering', 'Pembeli Musiman', dan 'Pembeli Jarang'.
- Lebih detail lagi, 'Pembeli Sering' dapat dibagi lagi menjadi 'Pengeluaran Tinggi' dan 'Pemburu Diskon'.

Dengan memvisualisasikan ini dalam dendrogram, bisnis mendapatkan pandangan hierarkis yang jelas tentang segmen pelanggan, yang mengarah pada strategi pemasaran yang lebih bernuansa.

Keuntungan dan Keterbatasan

Keuntungan

- **Struktur hierarkis.** Salah satu manfaat utama pengelompokan hierarkis adalah kemampuan untuk memvisualisasikan dan memahami pengelompokan bersarang. Ini seringkali berharga dalam skenario dunia nyata di mana hubungan hierarkis penting, seperti mengkategorikan produk atau memahami struktur organisasi.
- **Tidak perlu klaster yang telah ditentukan sebelumnya.** Tidak seperti beberapa metode pengelompokan lainnya yang memerlukan jumlah klaster yang telah ditentukan sebelumnya, pengelompokan hierarkis tidak memerlukan masukan ini, sehingga lebih mudah untuk memulai tanpa asumsi sebelumnya.

Keterbatasan

- **Intensif secara komputasi.** Pengelompokan hierarkis lebih menuntut secara komputasi daripada beberapa algoritma pengelompokan lainnya, terutama untuk dataset yang lebih besar. Algoritma harus mengevaluasi dan menggabungkan titik data atau klaster secara bertahap, yang menyebabkan biaya komputasi yang lebih tinggi.
- **Kurangnya reproduksibilitas dan determinisme.** Pengelompokan hierarkis tidak memasukkan keacakan dalam prosesnya. Oleh karena itu, orang mungkin berasumsi bahwa itu akan selalu menghasilkan hasil yang sama untuk dataset yang sama. Namun, masalahnya terletak pada nuansanya:

- Perangkat lunak atau alat yang berbeda dapat mengimplementasikan pengelompokan hierarkis dengan sedikit variasi, yang menyebabkan hasil yang berbeda.
- Urutan di mana titik data disajikan ke algoritma, atau urutan penggabungan, dapat memengaruhi hierarki yang dihasilkan. Ini berarti bahwa kecuali prosesnya dikontrol dengan cermat agar deterministik (yaitu, tindakan yang sama dilakukan dalam urutan yang sama setiap kali), eksekusi atau aplikasi yang berbeda mungkin menghasilkan hierarki kluster yang berbeda.
- **Kompleksitas dendrogram.** Dendrogram adalah alat utama untuk memvisualisasikan hasil pengelompokan hierarki. Meskipun dendrogram dapat memberikan banyak informasi, ia memiliki tantangan:
 - **Interpretasi.** Seiring bertambahnya jumlah titik data atau kluster, dendrogram dapat menjadi kompleks dan sulit untuk diinterpretasikan. Menentukan di mana harus memotong dendrogram untuk mendefinisikan kluster tidak selalu mudah.
 - **Sensitivitas skala.** Dendrogram dapat sensitif terhadap skala data. Jika data tidak distandarisasi atau dinormalisasi dengan benar, dendrogram mungkin salah menggambarkan hubungan hierarki yang sebenarnya.
 - **Detail yang berlebihan.** Terutama untuk kumpulan data besar, dendrogram dapat menampilkan detail yang berlebihan, sehingga sulit untuk segera membedakan pola atau hubungan yang bermakna.

Kesimpulannya, meskipun pengelompokan hierarkis menawarkan keunggulan unik, penting untuk menyadari keterbatasannya, terutama ketika berurusan dengan kumpulan data yang lebih besar atau ketika membutuhkan hasil yang konsisten dan dapat direproduksi. Menginterpretasikan dendrogram dengan benar juga membutuhkan tingkat keahlian dan pengalaman.

Contoh Dunia Nyata: Kebiasaan Konsumsi Media

Sebuah layanan streaming ingin memahami kebiasaan menonton audiensnya. Dengan menggunakan pengelompokan hierarkis, mereka memsegmentasikan pengguna berdasarkan genre yang ditonton. Pada tingkat yang lebih luas, kelompok seperti 'Pecinta Aksi', 'Penggemar Drama', dan 'Penonton Dokumenter' diidentifikasi. Lebih jauh lagi, 'Pecinta Aksi' terbagi menjadi 'Penggemar Film Superhero' dan 'Penggemar Film Aksi Klasik'.

Berdasarkan wawasan ini, layanan tersebut dapat merekomendasikan konten yang lebih terkurasi kepada pengguna, sehingga meningkatkan keterlibatan pengguna. Pengelompokan hierarkis menawarkan pendekatan unik untuk memahami perilaku dan preferensi pelanggan. Kemampuannya untuk memberikan perspektif segmentasi bertingkat menjadikannya sangat berharga bagi bisnis yang mencari wawasan mendalam. Meskipun memiliki tantangan, jika diterapkan dengan bijaksana, hal ini dapat secara signifikan meningkatkan strategi pemasaran, memastikan strategi tersebut disesuaikan dan ditargetkan.

Pembahasan Kekinian, Frekuensi, dan Nilai Moneter dalam Pemasaran

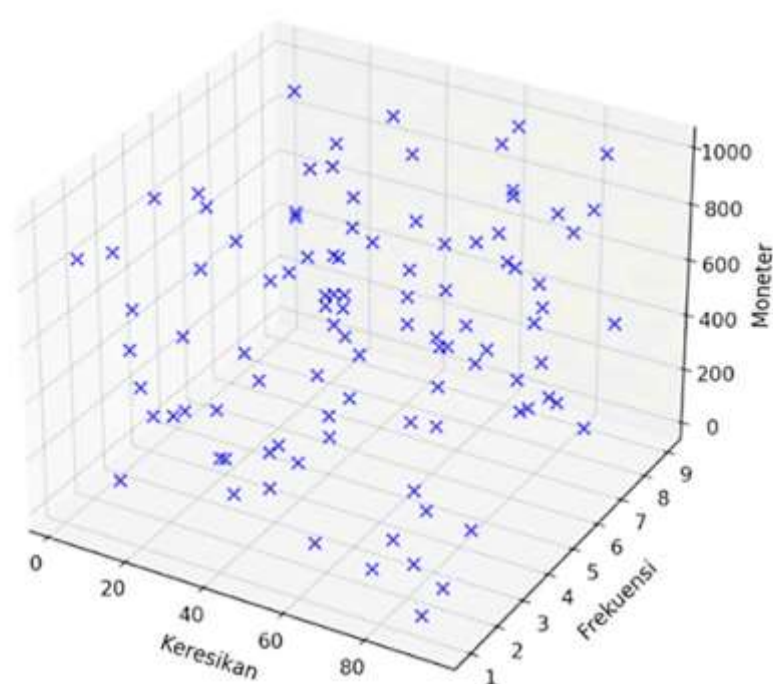
Pembahasan RFM adalah metode dasar dalam dunia segmentasi pelanggan berdasarkan atribut perilaku. Kekinian, frekuensi, dan nilai moneter mewakili tiga dimensi

penting untuk memahami dan memprediksi perilaku pelanggan. Dengan membahas dan mensegmentasi pelanggan berdasarkan parameter ini, bisnis dapat menyesuaikan strategi pemasaran mereka secara lebih efektif dan meningkatkan keterlibatan pelanggan. Gambar 4.6 menyajikan scatterplot 3D yang menunjukkan hasil penerapan segmen RFM pada kumpulan data pelanggan sampel. Representasi visual ini secara efektif menggambarkan distribusi pelanggan di ketiga dimensi RFM, memberikan pandangan yang jelas dan komprehensif tentang bagaimana parameter ini berpotongan untuk membentuk segmen pelanggan yang berbeda.

Dengan mengamati scatterplot 3D ini, pemasar dapat memperoleh pemahaman yang lebih dalam tentang basis pelanggan mereka, mengamati pola dan tren yang menginformasikan strategi pemasaran yang ditargetkan dan berkontribusi pada peningkatan keterlibatan pelanggan. Alat visual seperti itu sangat penting dalam menerjemahkan aspek teoritis pembahasan RFM menjadi wawasan yang dapat ditindaklanjuti.

Memahami Kekinian, Frekuensi, dan Nilai Moneter

- **Kekinian (R).** Mengacu pada waktu sejak transaksi atau interaksi terakhir pelanggan. Pelanggan yang telah berinteraksi atau membeli baru-baru ini lebih cenderung merespons secara positif terhadap penawaran baru dan umumnya dianggap lebih loyal.
- **Frekuensi (F).** Menunjukkan seberapa sering pelanggan bertransaksi atau berinteraksi dengan merek dalam jangka waktu tertentu. Pelanggan dengan frekuensi tinggi adalah pembeli yang konsisten dan sangat penting untuk keberlangsungan bisnis.



Gambar 4.6 Hasil Segmentasi RFM pada Kumpulan Data Pelanggan Sampel.

- **Moneter (M).** Mewakili jumlah total uang yang telah dihabiskan pelanggan dengan merek selama periode tertentu. Pelanggan dengan nilai moneter tinggi adalah

pembelanja tinggi, seringkali membentuk segmen yang memberikan kontribusi besar terhadap pendapatan bisnis.

Manfaat Pembahasan Recency, Frequency, Monetary

- **Pemasaran yang ditargetkan.** Dengan melakukan segmentasi basis pelanggan berdasarkan kriteria RFM, bisnis dapat membuat kampanye yang sangat tertarget. Misalnya, pelanggan dengan recency tinggi tetapi frekuensi rendah dapat ditargetkan dengan kampanye khusus untuk meningkatkan frekuensi pembelian mereka.
- **Peningkatan retensi pelanggan.** Memahami pelanggan mana yang berisiko berhenti berlangganan memungkinkan bisnis untuk mengambil tindakan pencegahan, menawarkan promosi khusus atau manfaat loyalitas untuk mempertahankan mereka.
- **Anggaran pemasaran yang dioptimalkan.** Alih-alih menghabiskan uang secara sembarangan, bisnis dapat menyalurkan anggaran pemasaran mereka secara lebih efektif ke segmen yang akan menghasilkan ROI tertinggi.
- **Personalisasi.** Pembahasan RFM dapat mendorong strategi personalisasi, memastikan bahwa pelanggan menerima konten dan penawaran yang paling relevan dengan perilaku dan pola pengeluaran mereka.

Menentukan Ambang Batas Penilaian Kekinian, Frekuensi, dan Moneter

Menentukan ambang batas untuk penilaian RFM sangat penting karena memengaruhi bagaimana pelanggan disegmentasikan dan ditargetkan. Berikut cara bisnis biasanya menetapkan ambang batas ini:

- **Penilaian berbasis persentil.** Salah satu metode umum adalah membagi dataset menjadi persentil. Misalnya, 20% pelanggan teratas dalam hal kekinian (yaitu, yang paling baru) mungkin diberi skor 5, 20% berikutnya diberi skor 4, dan seterusnya. Metode ini memastikan bahwa penilaian relatif terhadap dataset saat ini dan dapat disesuaikan seiring pertumbuhan bisnis atau perubahan perilaku pelanggan.
- **Aturan bisnis.** Beberapa bisnis mungkin memiliki tolok ukur atau target khusus. Misalnya, jika suatu bisnis percaya bahwa setiap pelanggan yang belum membeli dalam 30 hari terakhir berisiko berhenti berlangganan, mereka mungkin memberi pelanggan tersebut skor kekinian 1, terlepas dari bagaimana hal itu selaras dengan persentil.
- **Deviasi standar.** Metode lain melibatkan penghitungan rata-rata dan deviasi standar untuk kekinian, frekuensi, dan nilai moneter. Pelanggan dapat diberi skor berdasarkan seberapa banyak deviasi standar dari rata-rata perilaku mereka.
- **Pembahasan historis.** Dengan melihat data masa lalu, bisnis dapat mengidentifikasi pola atau tolok ukur. Jika secara historis, pelanggan yang membeli lebih dari lima kali dalam sebulan cenderung loyal, frekuensi di atas angka ini dapat diberi skor 5.
- **Pendekatan iteratif.** Beberapa bisnis mungkin memulai dengan sistem penilaian awal, kemudian menyesuaikannya dari waktu ke waktu berdasarkan hasilnya. Jika ambang batas penilaian tidak efektif dalam mengelompokkan pelanggan yang berharga, ambang batas tersebut dapat dikalibrasi ulang.

- **Melibatkan pemangku kepentingan.** Khusus untuk menetapkan ambang batas moneter, bisnis dapat melibatkan tim keuangan atau penjualan untuk memahami apa yang dianggap sebagai pembelanja "tinggi".
- **Pentingnya kalibrasi.** Sangat penting untuk meninjau kembali dan mengkalibrasi ulang ambang batas ini secara teratur. Perilaku pelanggan, model bisnis, dan faktor eksternal dapat berkembang, sehingga memerlukan penyesuaian pada penilaian.

Dengan memahami dan menetapkan ambang batas ini dengan cermat, bisnis dapat memastikan bahwa pembahasan RFM mereka disesuaikan dengan kebutuhan unik mereka dan dapat menghasilkan wawasan yang dapat ditindaklanjuti. Ini bukan hanya tentang angka; ini tentang memahami apa arti angka-angka tersebut dalam konteks tujuan dan tantangan bisnis.

Menerapkan Pembahasan Kekinian, Frekuensi, dan Moneter

1. **Pengumpulan data.** Kumpulkan data transaksi pelanggan. Pastikan data tersebut bersih, terbaru, dan akurat.
2. **Pemberian skor.** Berikan skor biasanya dari 1 hingga 5 (dengan 5 sebagai skor tertinggi) berdasarkan nilai RFM. Pelanggan dengan skor 555 adalah pelanggan bernilai tinggi, yang telah berinteraksi baru-baru ini, sering, dan telah menghabiskan sejumlah besar uang.
3. **Segmentasi.** Berdasarkan skor RFM, segmentasikan pelanggan ke dalam kategori seperti 'Pelanggan Unggulan', 'Berisiko', 'Tidak Boleh Kehilangan Mereka', 'Pelanggan Baru', 'Hilang', dan sebagainya. Setiap segmen akan sesuai dengan kombinasi skor RFM yang unik.
4. **Strategi pemasaran yang disesuaikan.** Kembangkan dan terapkan strategi pemasaran yang spesifik untuk setiap segmen. Misalnya, pelanggan 'Berisiko' mungkin mendapat manfaat dari penawaran eksklusif atau poin loyalitas untuk menarik kembali mereka.
5. **Pemantauan berkelanjutan.** RFM bukanlah pembahasan sekali saja. Pantau dan sesuaikan skor secara berkelanjutan seiring dengan perkembangan perilaku pelanggan.

Studi Kasus: Implementasi E-Commerce

Sebuah bisnis e-commerce memperhatikan penurunan penjualan dan peningkatan tingkat churn pelanggan. Setelah menerapkan pembahasan RFM, mereka mengidentifikasi segmen pelanggan (dengan skor 515) yang telah melakukan pembelian signifikan di masa lalu, tetapi sudah cukup lama sejak transaksi terakhir mereka, meskipun mereka cukup sering berbelanja. Menyadari bahwa pelanggan ini 'Berisiko', situs e-commerce meluncurkan kampanye email "Kami Merindukan Anda" dengan diskon eksklusif. Hal ini menghasilkan tingkat keterlibatan kembali sebesar 25% dari segmen tersebut, yang meningkatkan penjualan.

Pertimbangan dan Keterbatasan

RFM adalah pembahasan kuantitatif dan mungkin tidak memperhitungkan aspek kualitatif perilaku pelanggan. Sangat penting untuk memilih kerangka waktu yang tepat untuk pembahasan, yang mungkin bervariasi berdasarkan model bisnis dan industri. Tidak semua pelanggan dengan pengeluaran tinggi menguntungkan. RFM harus dikombinasikan dengan pembahasan profitabilitas untuk pandangan holistik. Pembahasan RFM, dengan

kesederhanaan dan efektivitasnya, tetap menjadi alat penting bagi bisnis modern. Dengan berfokus pada tiga metrik kunci, metode ini membantu mengungkap seluk-beluk perilaku pelanggan, memungkinkan bisnis untuk menjalin hubungan yang lebih kuat dan lebih personal dengan klien mereka. Dalam dunia pembelajaran data dan strategi pemasaran yang luas, RFM menonjol sebagai mercusuar, membimbing bisnis menuju interaksi pelanggan yang lebih bermakna dan peningkatan profitabilitas.

Meskipun pembahasan RFM menawarkan pendekatan yang lugas dan efektif untuk segmentasi pelanggan, penting juga untuk mempertimbangkan bagaimana perbandingannya dengan metode segmentasi lain seperti k-means dan pengelompokan hierarkis. Tabel 4.4 menyajikan perbandingan komprehensif dari ketiga teknik ini, mencakup kriteria seperti jenis data yang paling cocok untuk setiap metode, ukuran dataset yang dapat mereka tangani, skalabilitas, interpretasi, dan kasus penggunaan spesifik.

Pandangan komparatif ini memungkinkan kita untuk menghargai kekuatan dan keterbatasan unik dari setiap pendekatan dan memahami di mana pembahasan RFM berada dalam lanskap strategi segmentasi berbasis data yang lebih luas. Dengan mempelajari tabel ini, bisnis dapat membuat keputusan yang lebih tepat mengenai metode segmentasi mana yang paling sesuai dengan kebutuhan spesifik mereka dan sifat data mereka, sehingga meningkatkan efektivitas strategi pemasaran mereka.

Tabel 4.4 Perbandingan K-Means, Pengelompokan Hierarki, dan Pembahasan RFM.

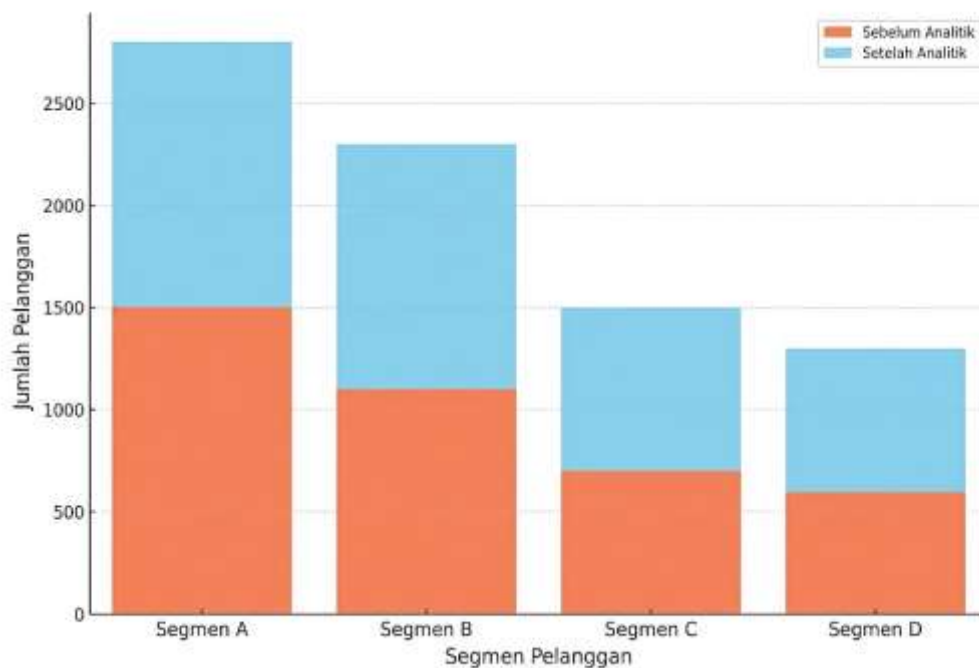
Kriteria	Pengelompokan K-Means	Pengelompokan Hierarkis	Pembahasan RFM
Jenis data	Numerik, seringkali distandarisasi atau dinormalisasi	Numerik, dapat menangani tipe data campuran dalam beberapa variasi	Data transaksional (tanggal, frekuensi, jumlah)
Ukuran dataset	Bekerja dengan baik pada kumpulan data besar	Lebih cocok untuk dataset yang lebih kecil karena kompleksitas komputasi	Relatif fleksibel; dapat menangani kumpulan data kecil dan besar
Masalah bisnis	Segmentasi umum berdasarkan atribut atau perilaku pelanggan	Memahami pengelompokan bersarang atau hierarki dalam data; menghasilkan dendrogram untuk representasi visual	Segmentasi nilai pelanggan berdasarkan perilaku pembelian
Jumlah kluster	Perlu ditentukan terlebih dahulu (meskipun metode seperti metode siku dapat membantu menentukan angka optimal)	Tidak perlu ditentukan di awal; membuat pohon kluster	Segmen biasanya berdasarkan kombinasi skor (misalnya, 111 hingga 555)

Skalabilitas	Sangat mudah diskalakan	Kurang mudah diskalakan karena intensitas komputasi	Sangat mudah diskalakan, terutama dengan alat otomatis
Interpretabilitas	Bergantung pada fitur yang digunakan; mungkin memerlukan pembuatan profil setelah pengelompokan	Dendrogram bisa sulit diinterpretasikan, terutama dengan dataset besar	Sangat mudah diinterpretasikan karena penilaiannya yang lugas dan fokus pada tiga parameter penting
Contoh kasus penggunaan	Segmentasi pengunjung situs web berdasarkan perilaku; mengidentifikasi segmen pelanggan berdasarkan preferensi produk	Memahami hubungan hierarkis dalam kategori produk, mengelompokkan item serupa dalam sistem rekomendasi konten	Menargetkan pelanggan bernilai tinggi dengan penawaran premium, melibatkan kembali pelanggan yang belum melakukan pembelian baru-baru ini
Hasil deterministik	Mungkin tidak deterministik karena inisialisasi centroid secara acak	Deterministik karena tidak memiliki unsur keacakan	Deterministik, berdasarkan penilaian tetap dari parameter RFM
Fleksibilitas	Pengelompokan tujuan umum; dapat disesuaikan dengan berbagai dataset dan masalah	Lebih spesifik, terutama ketika ada struktur hierarkis dalam data	Dirancang khusus untuk data transaksional; kurang mudah beradaptasi dengan skenario non-transaksional

4.8 MENENTUKAN ATRIBUT KUNCI PELANGGAN

Pembelajaran Inferensial untuk Segmentasi Pelanggan

Segmentasi pelanggan adalah strategi pemasaran mendasar yang melibatkan pembagian basis pelanggan menjadi kelompok-kelompok yang lebih kecil dengan karakteristik, preferensi, atau perilaku yang serupa. Pembelajaran inferensial memainkan peran penting dalam segmentasi pelanggan karena memungkinkan pemasar untuk mengidentifikasi pola dan hubungan yang mendasari dalam data pelanggan, menguji hipotesis tentang segmen yang berbeda, dan membuat keputusan yang tepat tentang penargetan dan personalisasi.



Gambar 4.7 Bagaimana Pembahasan Inferensial Meningkatkan Segmentasi Pelanggan untuk Merek Tertentu.

Pada bagian ini, kita akan membahas penerapan pembelajaran inferensial dalam segmentasi pelanggan, dengan referensi yang relevan. Dalam mengilustrasikan peran penting pembelajaran inferensial dalam meningkatkan segmentasi pelanggan, Gambar 4.7 menyajikan studi kasus yang menarik. Grafik ini menunjukkan bagaimana penerapan metode pembelajaran inferensial telah secara signifikan meningkatkan segmentasi pelanggan untuk merek tertentu.

Mengidentifikasi Atribut Pelanggan Utama

Langkah pertama dalam segmentasi pelanggan adalah mengidentifikasi atribut utama yang membedakan pelanggan, seperti demografi, psikografi, perilaku pembelian, dan preferensi. Teknik pembelajaran inferensial, seperti pembahasan korelasi dan regresi, dapat membantu pemasar menentukan atribut mana yang secara signifikan terkait dengan nilai pelanggan, loyalitas, atau kepuasan. Dengan memahami hubungan antara atribut pelanggan dan hasil pemasaran, pemasar dapat memilih variabel yang paling relevan untuk segmentasi dan menyesuaikan strategi mereka sesuai dengan itu.

Pembahasan Kluster untuk Segmentasi

Pembahasan kluster adalah teknik pembelajaran inferensial yang banyak digunakan untuk segmentasi pelanggan, yang bertujuan untuk mengelompokkan pelanggan berdasarkan kesamaan mereka di berbagai atribut yang dipilih. Berbagai algoritma pengelompokan, seperti pengelompokan hierarkis, k-means, dan pengelompokan berbasis model, dapat diterapkan untuk membagi data pelanggan ke dalam segmen homogen. Dengan membandingkan sifat statistik (misalnya, rata-rata, varians) dari kluster yang dihasilkan, pemasar dapat memperoleh

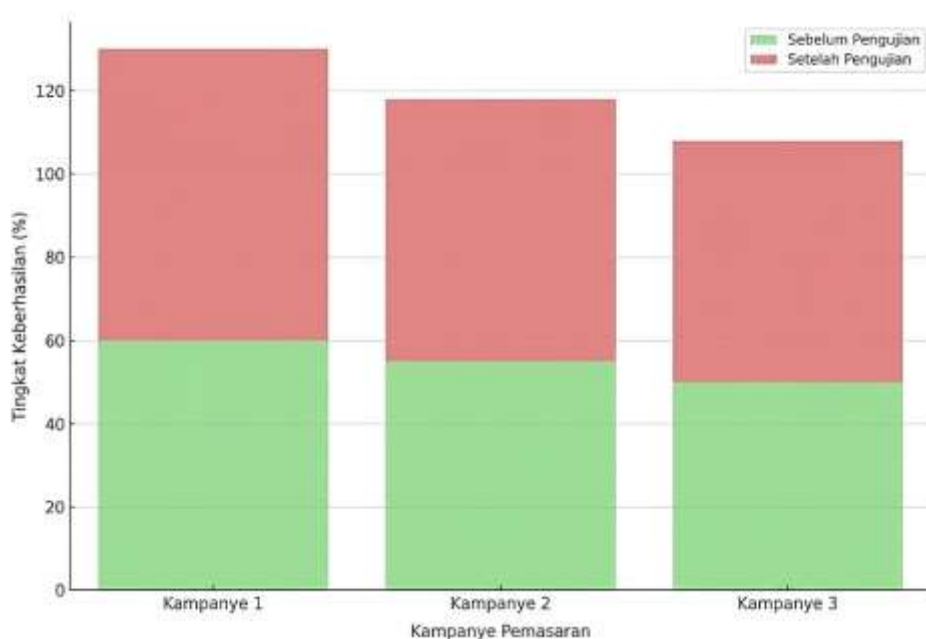
wawasan tentang segmen pelanggan yang berbeda dan mengembangkan strategi pemasaran yang ditargetkan untuk setiap kelompok.

Pengujian Hipotesis untuk Validasi Segmentasi

Setelah segmen pelanggan diidentifikasi, pengujian hipotesis dapat digunakan untuk memvalidasi hasil segmentasi dan memastikan bahwa perbedaan antar segmen signifikan secara statistik. Uji T, ANOVA, atau uji chi-square dapat digunakan untuk membandingkan rata-rata atau proporsi hasil pemasaran utama (misalnya, penjualan, kepuasan pelanggan, tingkat konversi) di berbagai segmen. Jika hipotesis nol tentang kesamaan rata-rata atau proporsi ditolak, pemasar dapat memiliki kepercayaan yang lebih besar pada hasil segmentasi dan menerapkan strategi yang ditargetkan untuk mengatasi kebutuhan dan preferensi unik dari setiap segmen pelanggan.

Pengujian Hipotesis untuk Kinerja Kampanye Pemasaran

Mengevaluasi kinerja kampanye pemasaran sangat penting untuk mengoptimalkan strategi pemasaran, alokasi anggaran, dan ROI. Pengujian hipotesis memainkan peran kunci dalam proses evaluasi ini, memungkinkan pemasar untuk menilai signifikansi statistik dari perbedaan yang diamati dalam hasil pemasaran antara kampanye, saluran, atau segmen pelanggan. Pada bagian ini, kita akan membahas penerapan pengujian hipotesis untuk kinerja kampanye pemasaran, dengan referensi yang relevan. Gambar 4.8 memberikan representasi grafis yang membandingkan kinerja kampanye pemasaran sebelum dan sesudah penerapan teknik pengujian hipotesis.



Gambar 4.8 Membandingkan Kinerja Kampanye Pemasaran Sebelum dan Setelah Menerapkan Pengujian Hipotesis.

Pengujian A/B dan Desain Eksperimental

Pengujian A/B, juga dikenal sebagai uji coba terkontrol acak atau pengujian terpisah, adalah pendekatan populer untuk mengevaluasi efektivitas berbagai taktik pemasaran, seperti

materi iklan, halaman arahan, atau baris subjek email. Dalam pengujian A/B, sampel pelanggan dibagi secara acak menjadi dua atau lebih kelompok, masing-masing terpapar versi stimulus pemasaran yang berbeda (misalnya, perlakuan vs. kontrol). Uji hipotesis, seperti uji t atau uji chi-kuadrat, digunakan untuk membandingkan rata-rata atau proporsi indikator kinerja utama (KPI), seperti tingkat konversi, tingkat klik-tayang, atau pendapatan, antar kelompok. Jika hipotesis nol tentang kesamaan rata-rata atau proporsi ditolak, pemasar dapat menyimpulkan bahwa perbedaan yang diamati signifikan secara statistik dan menerapkan taktik pemasaran yang lebih efektif.

Pemodelan Bauran Pemasaran dan Pembahasan Regresi

Pemodelan bauran pemasaran adalah teknik pembelajaran inferensial yang menggunakan pembahasan regresi berganda untuk mengukur dampak berbagai aktivitas pemasaran terhadap penjualan atau hasil pemasaran lainnya. Dengan memperkirakan koefisien variabel pemasaran (misalnya, periklanan, promosi, penetapan harga) dalam model regresi, pemasar dapat menentukan kontribusi relatif setiap aktivitas terhadap kinerja keseluruhan dan menguji hipotesis tentang signifikansinya. Berdasarkan wawasan ini, pemasar dapat mengoptimalkan bauran pemasaran mereka dan mengalokasikan sumber daya secara lebih efisien untuk memaksimalkan ROI.

Pembahasan Deret Waktu untuk Kinerja Pemasaran

Pembahasan deret waktu, seperti model autoregresif terintegrasi rata-rata bergerak atau model ruang keadaan penghalusan eksponensial, dapat digunakan untuk memprediksi kinerja pemasaran dan menilai signifikansi intervensi pemasaran dari waktu ke waktu. Dengan menggabungkan uji hipotesis, seperti pembahasan intervensi atau uji kausalitas Granger, pemasar dapat mengevaluasi apakah peristiwa pemasaran tertentu (misalnya, peluncuran produk, promosi, atau kampanye iklan) memiliki dampak signifikan pada penjualan atau KPI lainnya. Temuan ini dapat membantu pemasar memahami efektivitas upaya pemasaran mereka dan membuat keputusan berbasis data untuk meningkatkan kinerja di masa mendatang.

Pemodelan Atribusi dan Pengujian Hipotesis

Pemodelan atribusi adalah metode yang digunakan untuk membahas perjalanan pelanggan dan memberikan kredit kepada berbagai titik kontak pemasaran yang berkontribusi pada konversi atau penjualan. Berbagai model atribusi, seperti model sentuhan terakhir, sentuhan pertama, atau multi-sentuhan, dapat digunakan untuk mengalokasikan kredit ke saluran atau kampanye pemasaran. Pengujian hipotesis dapat membantu pemasar menentukan apakah perbedaan yang diamati dalam kinerja saluran atau kampanye pemasaran, sebagaimana diukur oleh model atribusi, signifikan secara statistik. Dengan memvalidasi hasil atribusi, pemasar dapat mengoptimalkan bauran pemasaran mereka dan membuat keputusan berbasis data untuk meningkatkan efisiensi investasi pemasaran mereka.

4.9 INTEGRASI DATA UNTUK STRATEGI MASA DEPAN

Dalam lanskap pemasaran yang berkembang pesat, memanfaatkan data untuk mendapatkan wawasan yang dapat ditindaklanjuti tidak pernah sepenting ini. Bab ini

membahas secara mendalam ranah pembelajaran inferensial dan pengujian hipotesis, mengungkap peran pentingnya dalam disiplin pemasaran. Pembelajaran inferensial, dengan kemampuannya untuk mengekstrapolasi wawasan dari sampel ke populasi yang lebih besar, memungkinkan bisnis untuk membuat keputusan yang kuat tanpa kesulitan mensurvei seluruh basis pelanggan. Ini memberi bisnis kekuatan statistik untuk memahami perilaku, preferensi, dan pola pelanggan pada tingkat makro, membentuk arah strategi pemasaran holistik. Namun, pengujian hipotesis adalah landasan untuk memvalidasi asumsi pemasaran. Seiring dunia pemasaran semakin didorong oleh data, metode ini menawarkan cara terstruktur untuk menguji dan mengadopsi atau menolak strategi. Ini bukan hanya tentang membuat pilihan, tetapi tentang membuat pilihan yang berakar pada bukti, memberikan dasar empiris untuk keputusan pemasaran.

Implementasi praktis yang dibahas, mulai dari segmentasi pelanggan dengan pembahasan kluster hingga evaluasi efektivitas kampanye pemasaran, semakin memperkuat relevansi teknik-teknik ini. Ini bukan sekadar konstruksi teoretis, tetapi alat yang dapat ditindaklanjuti yang dapat digunakan pemasar, baik untuk membedah basis pelanggan menjadi segmen yang dapat ditindaklanjuti atau untuk mengukur dampak kampanye iklan baru. Penting juga untuk mencatat tantangan yang melekat pada teknik-teknik ini. Mulai dari menentukan jumlah kluster optimal dalam pengelompokan k-means hingga menghindari jebakan inisialisasi acak, praktisi harus mendekati alat-alat ini dengan perpaduan ketelitian analitis dan pertimbangan yang cermat.

Singkatnya, seiring semakin terjalannya persimpangan antara pemasaran dan ilmu data, para profesional yang dilengkapi dengan pengetahuan tentang pembelajaran inferensial dan pengujian hipotesis akan lebih siap untuk menavigasi perpaduan ini. Mereka akan menjadi pembawa obor, memimpin bisnis menuju strategi pemasaran yang lebih informatif, berdampak, dan inovatif. Seiring kita melangkah lebih dalam ke era data, biarlah bab ini berfungsi sebagai kompas dan katalis, membimbing dan memotivasi para pemasar untuk memanfaatkan potensi sebenarnya dari ilmu data dalam pekerjaan mereka.

LATIHAN 4.1

INFERENSI BAYESIAN UNTUK PEMASARAN YANG DIPERSONALISASI

Tujuan:

Menggunakan inferensi Bayesian untuk memperkirakan kemungkinan pelanggan tertarik pada elektronik berdasarkan perilaku dan demografi mereka di masa lalu.

Tugas:

1. **Eksplorasi Data:** Membahas dataset untuk memahami demografi pelanggan dan perilaku mereka di masa lalu.
2. **Pembahasan Bayesian:**
 - a. Menghitung probabilitas prior (minat umum pada elektronik).
 - b. Menghitung kemungkinan (probabilitas mengklik konten terkait elektronik).
 - c. Menghitung probabilitas posterior menggunakan teorema Bayes.

3. **Interpretasi:** Menginterpretasikan hasil untuk memahami segmen pelanggan mana yang lebih mungkin tertarik pada elektronik.
4. **Aplikasi:** Menyarankan strategi kampanye email yang dipersonalisasi berdasarkan hasil inferensi Bayesian.

Langkah-langkah:

1. Mengimpor Pustaka yang Diperlukan:

```
1. import pandas as pd
```

- **Pandas** digunakan untuk manipulasi dan pembahasan data.

2. Memuat dan Menampilkan Data — Bayesian_Inference:Customer_Data.csv:

```
2. # Loading the data for Exercise 1: Bayesian Inference for Personalized Marketing
3. df_bayesian = pd.read_csv('path_to_csv_file')
4. # Displaying the first few rows of the DataFrame
5. df_bayesian.head()
```

- **pd.read_csv():** Membaca file CSV ke dalam DataFrame pandas.
- **df_bayesian.head():** Menampilkan lima baris pertama DataFrame untuk gambaran singkat struktur data.

Setelah data dimuat, langkah selanjutnya akan melibatkan penghitungan probabilitas prior, kemungkinan, dan probabilitas posterior menggunakan teorema Bayes. Mari kita lanjutkan untuk melakukan perhitungan ini.

3. Hitung Probabilitas Prior:

- Langkah selanjutnya adalah menghitung probabilitas prior. Ini adalah kemungkinan umum seorang pelanggan tertarik pada elektronik berdasarkan data historis yang kita miliki. Berikut cara kita menghitungnya:

```
6. # Total number of customers who clicked on electronics email
7. total_clicked_electronics=
   df_bayesian['ClickedOnElectronicsEmail'].sum()
8. # Total number of customers
9. total_customers = len(df_bayesian)
10. # Prior probability: P(E) - Probability of being interested in
    electronics
11. prior_probability = total_clicked_electronics/ total_customers
```

- **df_bayesian['ClickedOnElectronicsEmail'].sum ():** Menghitung jumlah pelanggan yang mengklik email terkait elektronik.
- **len(df_bayesian):** Menentukan jumlah total pelanggan dalam dataset.
- **prior_probability:** Rasio pelanggan yang menunjukkan minat pada elektronik terhadap jumlah total pelanggan, yang memberi kita probabilitas prior P(E).

Dalam dataset kita, probabilitas prior P(E) seorang pelanggan yang tertarik pada elektronik adalah sekitar 0,312, atau 31,2%.

Selanjutnya, kita akan menghitung kemungkinan dan probabilitas total untuk mengklik konten terkait elektronik, yang diperlukan untuk menemukan probabilitas posterior menggunakan teorema Bayes.

4. Perhitungan Kemungkinan:

Kemungkinan $P(F|E)$ adalah probabilitas seorang pelanggan mengklik email terkait elektronik dengan asumsi mereka tertarik pada elektronik. Untuk contoh ini, mari kita asumsikan bahwa pelanggan yang tertarik pada elektronik dua kali lebih mungkin untuk mengklik email elektronik dibandingkan dengan pelanggan rata-rata. Begini cara kami menghitungnya:

```
12. # Likelihood: P(F|E) - Probability of clicking on electronics email
    given they are interested in electronics
13. # Assuming customers interested in electronics are twice as likely to
    click on electronics emails
14. likelihood = 2 * prior_probability
```

5. Perhitungan Probabilitas Total:

Probabilitas total $P(F)$ adalah probabilitas keseluruhan dari setiap pelanggan yang mengklik konten terkait elektronik. Ini pada dasarnya adalah tingkat klik rata-rata pada email elektronik dalam dataset kami:

```
15. # Total Probability: P(F) - Overall probability of clicking on
    electronics email
16. # This is the average click rate on electronics emails
17. total_probability= df_bayesian['ClickedOnElectronicsEmail'].mean()
```

Berdasarkan data:

- Kemungkinan $P(F | E)$ sekitar 0,624, atau 62,4%.
- Probabilitas total $P(F)$ sekitar 0,312, atau 31,2%.

Selanjutnya, kita akan menggunakan nilai-nilai ini untuk menghitung probabilitas posterior, yang akan memberi tahu kita probabilitas seorang pelanggan tertarik pada elektronik mengingat mereka telah mengklik konten yang berkaitan dengan elektronik. Langkah terakhir adalah menghitung probabilitas posterior menggunakan teorema Bayes. Probabilitas posterior $P(E | F)$ mewakili probabilitas seorang pelanggan tertarik pada elektronik, mengingat mereka telah mengklik konten yang berkaitan dengan elektronik. Berikut rumus dan perhitungannya:

6. Teorema Bayes

$$P(E|F) = \frac{P(F|E) \times P(E)}{P(F)}$$

- $P(E | F)$: Probabilitas posterior (apa yang ingin kita cari)
- $P(F | E)$: Kemungkinan mengklik email elektronik jika mereka tertarik pada elektronik (kemungkinan)
- $P(E)$: Probabilitas prior tertarik pada elektronik
- $P(F)$: Total probabilitas mengklik email elektronik

```
18. # Calculating the Posterior Probability using Bayes's Theorem
19. posterior_probability = (likelihood * prior_probability) /
    total_probability
```

Dalam dataset kami, probabilitas posterior $P(E|F)$ adalah sekitar 0,624, atau 62,4%. Ini berarti bahwa jika pelanggan mengklik konten terkait elektronik, ada kemungkinan

62,4% bahwa mereka tertarik pada elektronik. Ini merupakan peningkatan signifikan dari probabilitas awal sebesar 31,2%, menunjukkan bahwa mengklik konten terkait elektronik merupakan indikator kuat minat pada elektronik.

Ini melengkapi latihan inferensi Bayesian, yang menunjukkan cara menggunakan Python untuk menerapkan teorema Bayes untuk pembelajaran pemasaran. Pendekatan ini memungkinkan pemasar untuk menyempurnakan strategi mereka berdasarkan keyakinan yang diperbarui tentang preferensi pelanggan, yang mengarah pada kampanye pemasaran yang lebih efektif dan tepat sasaran.

LATIHAN 4.2

PENGUJIAN A/B UNTUK EVALUASI KAMPANYE PEMASARAN

Tujuan

Mengevaluasi efektivitas dua kampanye pemasaran menggunakan pengujian A/B.

Tugas:

1. **Desain Eksperimental:** Memahami desain pengujian A/B (penugasan acak, durasi, ukuran sampel).
2. **Pembahasan Statistik:**
 - Hitung metrik kinerja utama untuk kedua kampanye.
 - Lakukan pengujian hipotesis (misalnya, uji-t) untuk menentukan apakah ada perbedaan yang signifikan secara statistik dalam efektivitas kedua kampanye.
3. **Interpretasi Hasil:** Pembahasan dan interpretasikan hasil uji A/B.
4. **Pengambilan Keputusan:** Berikan rekomendasi tentang kampanye mana yang harus diadopsi berdasarkan hasil uji.

Langkah-langkah:

1. Mengimpor Pustaka yang Diperlukan:

```
1. import pandas as pd
```

- **Pandas** digunakan untuk manipulasi dan pembahasan data.

2. Memuat dan Menampilkan Data—B_Testing_Campaign_Data.csv:

```
2. # Loading the data for Exercise 2: A/B Testing for Marketing Campaign  
Evaluation
```

```
3. df_ab_testing = pd.read_csv('path_to_csv_file')
```

```
4. # Displaying the first few rows of the DataFrame
```

```
5. df_ab_testing.head()
```

- **pd.read_csv():** Membaca file CSV ke dalam DataFrame pandas.
- **df_ab_testing.head():** Menampilkan lima baris pertama DataFrame untuk gambaran umum struktur data.

Data terdiri dari grup kampanye 'A' dan 'B', dengan metrik yang akan membantu kita membandingkan efektivitas kampanye ini. Selanjutnya, kita akan memisahkan data untuk setiap grup kampanye dan menghitung metrik kinerja utama. Mari kita lanjutkan dengan perhitungan ini.

Langkah selanjutnya dalam pembahasan pengujian A/B melibatkan pemisahan data untuk setiap kampanye dan perhitungan metrik kinerja utama. Berikut cara kita melakukannya di Python:

3. Memisahkan Data Kampanye:

Pertama-tama kita membagi data menjadi dua subset, satu untuk setiap grup kampanye (A dan B):

```
6. # Separating the data for Campaign A and Campaign B
7. df_campaign_a = df_ab_testing[df_ab_testing['CampaignGroup'] == 'A']
8. df_campaign_b = df_ab_testing[df_ab_testing['CampaignGroup'] == 'B']
```

4. Menghitung Metrik Utama:

Kemudian kami menghitung rata-rata 'Rasio Klik-Tayang (CTR)' dan 'Rasio Konversi' untuk setiap kampanye:

```
9. # Mean Click-Through Rate (CTR) and Conversion Rate for each campaign
10. mean_ctr_a = df_campaign_a['ClickThroughRate'].mean()
11. mean_ctr_b = df_campaign_b['ClickThroughRate'].mean()
12. mean_conversion_rate_a = df_campaign_a['ConversionRate'].mean()
13. mean_conversion_rate_b = df_campaign_b['ConversionRate'].mean()
```

Berdasarkan data:

- **Kampanye A:**

- Rata-rata 'Rasio Klik-Tayang (CTR)': Sekitar 0,1108 (atau 11,08%)
- Rata-rata 'Rasio Konversi': Sekitar 0,0264 (atau 2,64%)

- **Kampanye B:**

- Rata-rata 'Rasio Klik-Tayang (CTR)': Sekitar 0,1048 (atau 10,48%)
- Rata-rata 'Rasio Konversi': Sekitar 0,0243 (atau 2,43%)

Metrik ini memberi kita indikasi awal kinerja setiap kampanye. Selanjutnya, kita akan melakukan uji statistik (seperti uji-t) untuk menentukan apakah perbedaan yang diamati dalam metrik ini antara kedua kampanye signifikan secara statistik. Mari kita lanjutkan pembahasan ini.

Untuk menentukan signifikansi statistik dari perbedaan yang diamati antara kedua kampanye, kita melakukan uji-t pada 'Rasio Klik-Tayang (CTR)' dan 'Rasio Konversi'. Berikut rincian bagian pembahasan ini:

5. Melakukan Uji-T:

Kita menggunakan fungsi **ttest_ind** dari modul **scipy.stats**, yang melakukan uji-t dua sampel independen. Tes ini membandingkan rata-rata dari dua kelompok independen (dalam hal ini, Kampanye A dan Kampanye B) untuk menentukan apakah ada perbedaan yang signifikan secara statistik di antara keduanya.

```
14. from scipy import stats
15. # T-test for Click-Through Rates
16. t_stat_ctr, p_value_ctr = stats.ttest_ind(
    df_campaign_a['ClickThroughRate'], df_campaign_b['ClickThroughRate'])
17. # T-test for Conversion Rates
```

```
18. t_stat_conversion, p_value_conversion =  
stats.ttest_ind(df_campaign_a['ConversionRate'], df_campaign_b['ConversionRate'])
```

- **stats.ttest_ind()**: Melakukan uji t untuk rata-rata dua sampel independen.
- **t_stat_ctr, p_value_ctr**: Statistik t dan nilai p untuk perbandingan 'Rasio Klik-Tayang'.
- **t_stat_conversion, p_value_conversion**: Statistik t dan nilai p untuk perbandingan 'Rasio Konversi'.

Berdasarkan data:

- Statistik t untuk perbandingan 'Rasio Klik-Tayang' adalah sekitar 1,231 dengan nilai p sekitar 0,219.
- Statistik t untuk perbandingan 'Rasio Konversi' adalah sekitar 1,654 dengan nilai p sekitar 0,099.

6. Interpretasi:

Nilai p menunjukkan probabilitas mengamati data jika hipotesis nol (tidak ada perbedaan antara kampanye) benar.

- Untuk 'Click-Through Rate' dan 'Conversion Rate', nilai p lebih besar dari tingkat alfa tipikal 0,05, menunjukkan bahwa kita tidak memiliki cukup bukti untuk menolak hipotesis nol pada tingkat signifikansi 5%.
- Oleh karena itu, berdasarkan data ini, kita tidak dapat menyimpulkan bahwa ada perbedaan yang signifikan secara statistik antara Kampanye A dan Kampanye B dalam hal 'Click-Through Rate' dan 'Conversion Rate'.

Ini menyelesaikan bagian pembahasan statistik dari latihan pengujian A/B, menunjukkan cara menggunakan Python untuk membandingkan efektivitas dua kampanye pemasaran. Hasilnya menunjukkan bahwa, dalam skenario ini, mungkin tidak ada perbedaan kinerja yang signifikan antara kedua kampanye tersebut.

BAB 5

PEMBAHASAN PREDIKTIF DAN PEMBELAJARAN MESIN

5.1 GAMBARAN UMUM PEMBAHASAN PREDIKTIF

Pembahasan prediktif adalah cabang dari pembahasan tingkat lanjut yang menggunakan data, algoritma statistik, dan teknik pembelajaran mesin untuk memprediksi hasil di masa depan. Tujuannya adalah untuk melampaui sekadar mengetahui apa yang telah terjadi, yaitu memberikan perkiraan terbaik tentang apa yang akan terjadi di masa depan. Hal ini dicapai dengan memanfaatkan berbagai teknik, termasuk penambangan data, statistik, pemodelan, pembelajaran mesin, dan kecerdasan buatan. Pembahasan prediktif didasarkan pada penangkapan hubungan antara variabel penjelas dan variabel yang diprediksi dari kejadian masa lalu dan menggunakannya untuk memprediksi hasil di masa depan. Pembahasan ini memanfaatkan pola yang ditemukan dalam data historis dan transaksional untuk mengidentifikasi risiko dan peluang, sehingga memberikan wawasan yang memandu pengambilan keputusan di berbagai sektor.

Dalam pemasaran, pembahasan prediktif digunakan untuk membahas data terkini dan fakta historis untuk lebih memahami pelanggan, produk, dan mitra, serta untuk mengidentifikasi potensi risiko dan peluang. Pembelajaran prediktif dapat digunakan untuk memprediksi perilaku pelanggan, mendeteksi dan mencegah penipuan, mengoptimalkan kampanye pemasaran, meningkatkan operasional, mengurangi risiko, memenuhi kebutuhan pelanggan secara lebih efektif, dan meningkatkan profitabilitas. Kemampuan pembelajaran prediktif untuk memberikan wawasan yang dapat ditindaklanjuti memberikan keunggulan kompetitif dan membantu organisasi membuat keputusan yang terinformasi dan berwawasan ke depan. Hal ini memberi bisnis kekuatan untuk memprediksi apa yang akan dilakukan atau diinginkan pelanggan mereka di masa depan, meningkatkan kinerja bisnis dan mendorong keputusan manajemen strategis.

Pembelajaran Mesin dalam Pemasaran

Pembelajaran mesin, bagian dari kecerdasan buatan, memberikan sistem kemampuan untuk secara otomatis belajar dan meningkatkan diri dari pengalaman tanpa diprogram secara eksplisit. Ini berfokus pada pengembangan program komputer yang dapat mengakses data dan menggunakannya untuk belajar sendiri. Pembelajaran mesin adalah penggerak utama pembelajaran prediktif, menyediakan algoritma yang memungkinkan prediksi.

Dalam konteks pemasaran, pembelajaran mesin dapat dimanfaatkan dalam berbagai cara untuk mendorong pengambilan keputusan yang lebih efektif, meningkatkan pengalaman pelanggan, dan memberikan nilai tambah yang lebih besar. Algoritma pembelajaran mesin dapat digunakan untuk memprediksi perilaku pelanggan, seperti pola pembelian atau kemungkinan pelanggan berhenti berlangganan. Algoritma ini dapat membantu mengelompokkan pelanggan ke dalam kelompok yang bermakna, sehingga memungkinkan strategi pemasaran yang lebih terarah dan personal. Pembelajaran mesin juga dapat diterapkan untuk mengoptimalkan harga, memperkirakan permintaan, dan meningkatkan

rekomendasi, di antara aplikasi lainnya. Misalnya, algoritma pembelajaran terawasi seperti regresi linier atau mesin vektor pendukung (SVM) dapat digunakan untuk memprediksi hasil tertentu, seperti nilai seumur hidup pelanggan (CLV) atau respons terhadap kampanye pemasaran. Algoritma pembelajaran tak terawasi seperti pengelompokan dapat digunakan untuk mengidentifikasi segmen atau kelompok dalam basis pelanggan Anda. Pembelajaran penguatan (RL), cabang lain dari pembelajaran mesin, dapat diterapkan untuk mengoptimalkan strategi penetapan harga dinamis atau rekomendasi yang dipersonalisasi.

Munculnya pembelajaran mesin dan bidang AI yang lebih luas telah membuka peluang baru bagi pemasar untuk lebih memahami pelanggan mereka, mengantisipasi kebutuhan mereka, dan memberikan pengalaman yang lebih menarik dan personal. Namun, penerapan pembelajaran mesin yang sukses dalam pemasaran membutuhkan pemahaman yang baik tentang konteks bisnis dan aspek teknis pembelajaran mesin.

Tantangan Umum dalam Pembahasan Prediktif dan Pembelajaran Mesin dalam Pemasaran

Dalam lanskap pembelajaran pemasaran, pembelajaran prediktif dan pembelajaran mesin telah menjadi alat yang sangat diperlukan. Namun, efektivitasnya dapat terhambat secara signifikan oleh berbagai tantangan. Berikut ini menguraikan hambatan umum yang dihadapi saat memanfaatkan teknik canggih ini dalam pemasaran:

- **Kualitas dan pra-pemrosesan data.** Salah satu tantangan yang paling umum adalah mendapatkan data berkualitas tinggi. Data yang tidak akurat, hilang, atau usang dapat berdampak buruk pada kinerja model prediktif. Langkah-langkah pra-pemrosesan seperti pembersihan data, normalisasi, dan transformasi sangat penting untuk memastikan data yang dimasukkan ke dalam model memiliki kualitas optimal.
- **Interpretasi model.** Terutama dengan model yang kompleks seperti jaringan saraf dalam atau metode ensemble, memahami bagaimana model membuat keputusan bisa menjadi tantangan. Ini bisa menjadi masalah dalam pemasaran ketika pemangku kepentingan ingin mengetahui alasan di balik sebuah prediksi.
- **Overfitting.** Dalam keinginan untuk menangkap setiap pola dari data, ada risiko membuat model yang terlalu kompleks dan overfitting data pelatihan. Model seperti itu berkinerja buruk dalam skenario dunia nyata karena pada dasarnya mereka telah menghafal data pelatihan daripada melakukan generalisasi darinya.
- **Menangani data yang tidak seimbang.** Dalam banyak aplikasi pemasaran, seperti prediksi churn, kelas mungkin tidak seimbang dengan jumlah pelanggan yang berhenti berlangganan jauh lebih sedikit daripada pelanggan yang tidak berhenti berlangganan. Hal ini dapat menyebabkan model bias terhadap kelas mayoritas, sehingga kehilangan wawasan berharga dari kelas minoritas.
- **Masalah skalabilitas.** Seiring pertumbuhan bisnis, volume data juga meningkat. Model perlu tangguh dan skalabel untuk menangani peningkatan beban data tanpa mengorbankan kecepatan atau akurasi.
- **Kekhawatiran privasi dan etika.** Menggunakan data konsumen untuk melatih model menimbulkan kekhawatiran etika dan privasi. Memastikan data dianonimkan dan

digunakan secara etis adalah hal yang terpenting. Tantangan ini dibahas secara komprehensif di Bab 13.

- **Perilaku konsumen yang terus berkembang.** Perilaku konsumen tidak statis. Terutama di era digital saat ini, perilaku konsumen dapat berubah dengan cepat. Model mungkin menjadi usang jika tidak beradaptasi dengan perubahan perilaku ini.

Pengelolaan tantangan ini secara efektif sangat penting bagi bisnis yang bertujuan untuk menggunakan pembelajaran prediktif dan pembelajaran mesin dalam merancang strategi pemasaran yang sukses. Bab-bab selanjutnya akan membahas strategi dan praktik terbaik untuk mengurangi tantangan ini dan memanfaatkan pembelajaran prediktif secara maksimal dalam pemasaran.

Kesalahpahaman dalam Pembahasan Prediktif dan Pembelajaran Mesin dalam Pemasaran

Mengungkap kesalahpahaman umum tentang pembahasan prediktif dan pembelajaran mesin dapat membantu pemasar menetapkan ekspektasi yang realistis dan menggunakan alat-alat ini secara lebih efektif. Berikut adalah beberapa mitos yang umum di bidang ini:

- **Model yang kompleks selalu lebih baik.** Meskipun model kompleks seperti pembelajaran mendalam dapat menangkap pola yang rumit, model tersebut tidak selalu diperlukan. Terkadang, model yang lebih sederhana seperti regresi linier sudah cukup dan lebih mudah diinterpretasikan.
- **Lebih banyak data selalu lebih baik.** Meskipun memiliki kumpulan data yang lebih besar biasanya membantu, kualitas data adalah yang terpenting. Selain itu, menambahkan data yang tidak relevan dapat mengurangi kinerja model.
- **Pembelajaran mesin dapat menyelesaikan segalanya.** Ekspektasi terhadap pembelajaran mesin dan pembahasan prediktif terkadang tidak realistis. Keduanya adalah alat, dan seperti alat lainnya, efisiensinya bergantung pada cara penggunaannya.
- **Akurasi tinggi berarti model yang baik.** Meskipun akurasi adalah metrik yang penting, itu bukan satu-satunya. Tergantung pada aplikasinya, metrik seperti presisi, recall, atau skor F1 mungkin lebih relevan.
- **Model berjalan secara otomatis.** Setelah diterapkan, model perlu dipantau dan diperbarui secara berkala. Model dapat bergeser seiring waktu karena perubahan pola data dan mungkin perlu dilatih ulang.
- **Setiap masalah membutuhkan pembelajaran prediktif.** Meskipun pembelajaran prediktif dapat menawarkan wawasan yang berharga, tidak setiap masalah pemasaran membutuhkannya. Sangat penting untuk menentukan apakah biaya dan upaya penerapan pembelajaran prediktif melebihi manfaatnya.

Dengan mengatasi kesalahpahaman ini, pemasar dapat menerapkan pembelajaran prediktif dan teknik pembelajaran mesin dengan lebih bijaksana, memastikan bahwa alat-alat canggih ini memberikan hasil terbaik untuk tujuan pemasaran mereka.

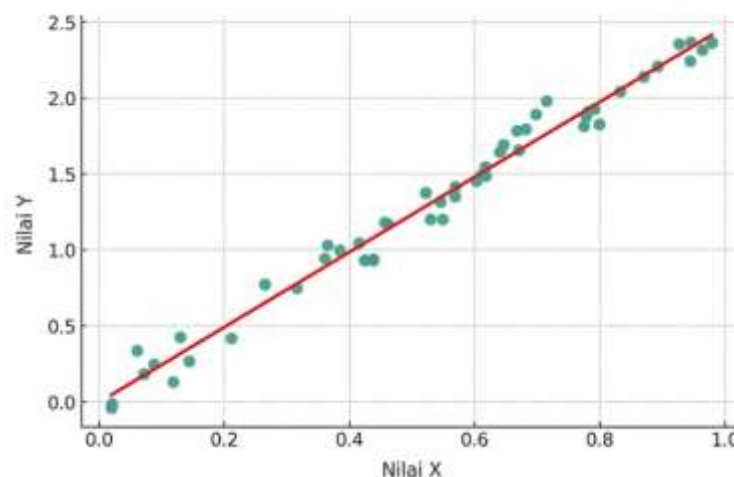
5.2 TANTANGAN PERAMALAN DAN PERTIMBANGAN BISNIS

Regresi Linier dan Logistik

Regresi linier dan logistik adalah teknik dasar dalam pembelajaran prediktif dan pembelajaran mesin. Keduanya merupakan bentuk pemodelan prediktif yang membahas hubungan antara variabel hasil (juga dikenal sebagai variabel dependen) dan satu atau lebih prediktor (juga dikenal sebagai variabel independen).

Regresi linier digunakan ketika variabel hasil bersifat kontinu, artinya dapat mengambil nilai apa pun dalam rentang tertentu. Regresi linier memodelkan hubungan antara variabel hasil dan prediktor sebagai garis lurus, oleh karena itu disebut regresi linier. Misalnya, dalam pemasaran, sebuah bisnis dapat menggunakan regresi linier untuk memprediksi pendapatan penjualan berdasarkan pengeluaran iklan. Gambar 5.1 memberikan representasi visual dari konsep ini, khususnya menggambarkan scatterplot dengan garis regresi linier. Gambar ini menunjukkan bagaimana regresi linier digunakan untuk memodelkan hubungan antara variabel hasil dan prediktor. Diagram sebaran menunjukkan titik data individual, dengan garis regresi linier mewakili kecocokan terbaik melalui titik-titik ini, menggambarkan tren dan arah hubungan. Namun, regresi logistik digunakan ketika variabel hasil bersifat biner, artinya hanya dapat mengambil dua nilai yang mungkin, seperti 0 atau 1, ya atau tidak, benar atau salah. Regresi logistik memodelkan log odds dari probabilitas hasil sebagai kombinasi linier dari prediktor. Misalnya, perusahaan telekomunikasi mungkin menggunakan regresi logistik untuk memprediksi apakah pelanggan akan berhenti berlangganan (1) atau tidak (0) berdasarkan pola penggunaan dan demografi mereka.

Meskipun regresi linier dan logistik adalah alat yang ampuh, keduanya juga memiliki keterbatasan. Misalnya, keduanya mengasumsikan hubungan linier antara hasil dan prediktor, yang mungkin tidak selalu berlaku dalam skenario dunia nyata. Selain itu, keduanya mungkin tidak berkinerja baik ketika berurusan dengan data kompleks berdimensi tinggi atau ketika hubungan yang mendasarinya bersifat nonlinier atau melibatkan interaksi antar prediktor. Dalam kasus-kasus ini, teknik pembelajaran mesin yang lebih canggih, seperti pohon keputusan atau jaringan saraf, mungkin lebih tepat.



Gambar 5.1 Scatterplot dengan Garis Regresi Linier.

Peramalan Deret Waktu

Peramalan deret waktu merupakan komponen penting dari pembelajaran prediktif dan pembelajaran mesin, khususnya relevan di bidang pemasaran. Deret waktu adalah urutan titik data yang dikumpulkan atau dicatat dalam urutan waktu pada interval waktu yang sama. Peramalan melibatkan prediksi nilai masa depan dari deret berdasarkan data historis.

Peramalan deret waktu digunakan secara luas dalam pemasaran untuk tugas-tugas seperti memprediksi penjualan, lalu lintas web, dan permintaan pelanggan. Ramalan ini dapat membantu bisnis merencanakan masa depan, mengelola sumber daya dengan lebih efisien, dan membuat keputusan strategis yang lebih baik. Misalnya, bisnis ritel dapat menggunakan peramalan deret waktu untuk memprediksi penjualan di masa mendatang dan menyesuaikan tingkat persediaan sesuai dengan itu.

Terdapat banyak metode untuk peramalan deret waktu, masing-masing dengan kekuatan dan kelemahannya. Beberapa metode yang paling umum meliputi model *autoregressive integrated moving average* (ARIMA), model exponential smoothing, dan model state space. Baru-baru ini, metode pembelajaran mesin seperti jaringan saraf berulang juga telah diterapkan pada peramalan deret waktu.

Terlepas dari kegunaannya, peramalan deret waktu bukannya tanpa tantangan. Hal ini memerlukan penanganan komponen seperti tren, musiman, dan noise, yang dapat mempersulit proses pemodelan. Hal ini juga mengasumsikan bahwa masa depan akan mengikuti pola yang sama seperti masa lalu, yang mungkin tidak selalu berlaku, terutama jika terjadi peristiwa yang mengganggu atau perubahan kondisi pasar. Oleh karena itu, peramalan deret waktu yang sukses tidak hanya membutuhkan keahlian teknis tetapi juga pemahaman yang baik tentang konteks bisnis dan dinamika pasar. Karena kita mempertimbangkan pentingnya peramalan deret waktu dalam domain pembelajaran pemasaran, memvisualisasikan penerapan model-model ini pada data historis sangatlah berharga.



Gambar 5.2 Data Deret Waktu dengan Overlay Model Peramalan.

Gambar 5.2 menawarkan penggambaran yang jelas tentang hal ini, dengan plot data deret waktu dan overlay model peramalan. Gambar tersebut mengilustrasikan bagaimana titik data historis digunakan untuk memproyeksikan tren masa depan, memberikan representasi nyata dari proses peramalan. Lapisan ini memungkinkan pemasar untuk membandingkan prediksi model dengan nilai historis aktual, menyoroti kemampuan model untuk menangkap pola dan tren yang mendasarinya. Alat visual semacam ini tidak hanya penting untuk memahami dinamika data deret waktu, tetapi juga untuk mengkomunikasikan hasil perkiraan kepada pemangku kepentingan, membantu dalam perencanaan strategis dan pengambilan keputusan.

5.3 TEKNIK PEMBELAJARAN MESIN

Pembelajaran Terawasi untuk Pemasaran

Pembelajaran terawasi adalah salah satu kategori utama pembelajaran mesin. Ini melibatkan penggunaan serangkaian contoh berlabel (pasangan input-output) untuk melatih model yang dapat membuat prediksi untuk contoh baru yang belum pernah dilihat sebelumnya. "Pengawasan" datang dalam bentuk label output yang diberikan untuk data pelatihan. Dalam konteks pemasaran, pembelajaran terawasi dapat diterapkan pada berbagai macam tugas. Misalnya, dapat digunakan untuk memprediksi CLV, meramalkan penjualan, melakukan segmentasi pelanggan, atau memperkirakan probabilitas pelanggan berhenti berlangganan. Masing-masing tugas ini dapat dibingkai sebagai masalah pembelajaran terawasi di mana tujuannya adalah untuk memprediksi output (misalnya, CLV, penjualan, probabilitas berhenti berlangganan) berdasarkan serangkaian input (misalnya, demografi pelanggan, riwayat transaksi, metrik keterlibatan).

Terdapat banyak algoritma pembelajaran terawasi yang tersedia, masing-masing dengan kekuatan dan kelemahannya. Beberapa yang paling umum meliputi regresi linier, regresi logistik, pohon keputusan, hutan acak, penguatan gradien, SVM, dan jaringan saraf. Pilihan algoritma bergantung pada berbagai faktor, seperti sifat tugas, jenis data, dan kendala bisnis. Misalnya, regresi linier dapat digunakan untuk peramalan penjualan, di mana tujuannya adalah untuk memprediksi hasil kontinu (penjualan) berdasarkan serangkaian prediktor (misalnya, pengeluaran iklan, musiman, indikator ekonomi). Namun, regresi logistik atau pohon keputusan dapat digunakan untuk prediksi churn, di mana tujuannya adalah untuk memprediksi hasil biner (churn atau tidak churn) berdasarkan serangkaian prediktor (misalnya, pola penggunaan, kepuasan pelanggan, riwayat penagihan). Meskipun pembelajaran terawasi dapat memberikan wawasan berharga dan mendorong pengambilan keputusan yang efektif dalam pemasaran, hal ini juga memerlukan pertimbangan cermat terhadap isu-isu seperti overfitting, underfitting, interpretasi model, dan privasi data.

Pohon Keputusan dalam Pembelajaran Terawasi

Pohon keputusan adalah algoritma pembelajaran mesin populer yang digunakan terutama untuk tugas klasifikasi dan regresi. Pada intinya, pohon keputusan membagi data menjadi subset berdasarkan nilai fitur input. Proses ini menghasilkan model keputusan seperti

pohon, di mana setiap node mewakili fitur, setiap cabang mewakili aturan keputusan, dan setiap daun mewakili hasil atau kelas.

Pembentukan pohon keputusan melibatkan langkah-langkah berikut:

1. **Pemilihan atribut.** Pilih atribut terbaik untuk membagi data. Keputusan ini sering melibatkan metrik seperti 'Information Gain', 'Gini Impurity', atau 'Variance Reduction'.
2. **Pemisahan.** Bagi dataset menjadi subset berdasarkan nilai atribut yang dipilih. Ini menghasilkan sebuah node dalam pohon.
3. **Pemisahan rekursif.** Untuk setiap subset, ulangi langkah 1 dan 2 hingga salah satu kondisi penghentian terpenuhi, seperti mencapai kedalaman maksimum pohon atau simpul memiliki jumlah sampel kurang dari jumlah minimum.
4. **Penugasan simpul daun.** Setelah pohon dibangun, tetapkan kelas keluaran ke setiap simpul daun, yang dapat digunakan untuk membuat prediksi untuk data baru.

Bayangkan sebuah perusahaan telekomunikasi ingin memprediksi pelanggan yang akan berhenti berlangganan berdasarkan berbagai atribut pelanggan. Pohon keputusan dapat disusun sebagai berikut:

- Node akar dapat membagi data berdasarkan fitur 'Durasi Kontrak' (misalnya, bulanan vs. satu tahun).
 - Untuk pelanggan dengan kontrak bulanan:
 - Node berikutnya dapat dibagi lebih lanjut berdasarkan 'Biaya Bulanan', dengan ambang batas, misalnya, Rp500.000.
 - Pelanggan yang membayar lebih dari Rp500.000 mungkin memiliki tingkat berhenti berlangganan yang lebih tinggi, yang mengarah ke node daun berlabel 'Berhenti Berlangganan.'
 - Pelanggan yang membayar kurang dari atau sama dengan Rp500.000 dapat dibagi lebih lanjut berdasarkan fitur lain, seperti 'Panggilan Dukungan Pelanggan'.
 - Untuk pelanggan dengan kontrak satu tahun:
 - Pohon tersebut dapat dibagi berdasarkan atribut yang berbeda, seperti 'Jenis Layanan Internet'.

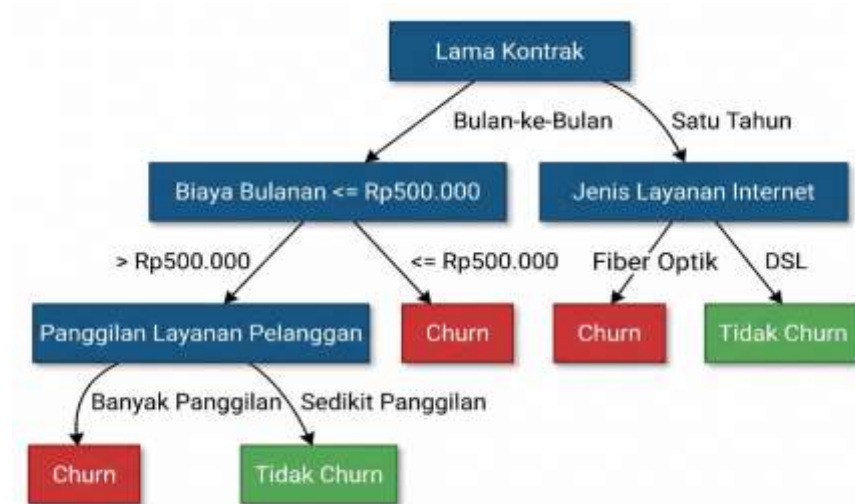
Mengikuti cabang-cabang pohon dari akar ke daun memberikan aturan keputusan yang mengarah pada hasil prediksi (dalam hal ini, apakah pelanggan cenderung berhenti berlangganan atau tidak). Lihat Gambar 5.3 untuk contohnya. Pohon keputusan disukai dalam banyak aplikasi bisnis karena interpretasinya.

Setiap jalur dalam pohon mewakili aturan keputusan, dan dengan demikian, mereka memberikan alasan yang jelas untuk setiap prediksi. Namun, mereka rentan terhadap overfitting, terutama ketika sangat dalam. Untuk mengurangi hal ini, teknik seperti pemangkasan atau metode ensemble seperti random forest dapat digunakan.

Random Forest dalam Pembelajaran Terawasi

Random forest adalah metode pembelajaran ensemble yang terutama digunakan untuk tugas klasifikasi dan regresi. Mereka beroperasi dengan membangun beberapa pohon keputusan selama pelatihan dan menghasilkan modus kelas (untuk klasifikasi) atau prediksi rata-rata (untuk regresi) dari masing-masing pohon untuk data yang belum pernah dilihat.

Pada dasarnya, ini adalah "hutan" pohon, di mana setiap pohon memberikan "suara" untuk suatu kelas, dan kelas mayoritas dipilih sebagai prediksi akhir.



Gambar 5.3 Struktur Pohon Keputusan.

Pembentukan hutan acak melibatkan langkah-langkah berikut:

1. **Pengambilan sampel bootstrap.** Pilih sampel secara acak dari dataset dengan penggantian, membuat beberapa subset.
2. **Pembuatan pohon.** Untuk setiap subset, kembangkan pohon keputusan. Namun, alih-alih menggunakan semua fitur untuk pemisahan di setiap node, subset fitur acak dipilih. Keacakan ini memastikan pohon-pohon tersebut beragam.
3. **Agregasi.** Untuk tugas klasifikasi, setiap pohon dalam hutan memprediksi kelas (memberikan suara untuk suatu kelas). Kelas yang mendapatkan suara terbanyak adalah prediksi hutan. Untuk tugas regresi, prediksi hutan adalah rata-rata dari prediksi semua pohon.
4. **Output.** Hasilkan prediksi berdasarkan hasil mayoritas (klasifikasi) atau rata-rata (regresi) dari semua pohon dalam hutan.

Pertimbangkan platform e-commerce yang bertujuan untuk memprediksi apakah pengguna akan membeli suatu produk berdasarkan fitur-fitur seperti riwayat penelusuran, waktu yang dihabiskan di situs, riwayat pembelian sebelumnya, dan sebagainya.

Dengan menggunakan random forest, platform akan melakukan hal berikut:

- Membuat beberapa subset data menggunakan bootstrap sampling.
- Untuk setiap subset, pohon keputusan dibangun, tetapi setiap node dalam pohon hanya menggunakan subset fitur acak untuk membuat keputusan (misalnya, hanya 'Riwayat Penjelajahan' dan 'Waktu yang Dihabiskan di Situs' untuk satu pohon, dan 'Riwayat Pembelian Sebelumnya' dan 'Waktu dalam Sehari' untuk pohon lainnya).
- Setelah hutan pohon dibangun, prediksi untuk pengguna baru dibuat dengan meminta setiap pohon dalam hutan memprediksi 'beli' atau 'tidak beli' berdasarkan fitur pengguna. Prediksi akhir adalah prediksi yang dipilih oleh mayoritas pohon.

Random forest sangat ampuh karena dapat menangkap pola nonlinier yang kompleks dalam data, kurang rentan terhadap overfitting dibandingkan dengan pohon keputusan individual, dan dapat menangani campuran fitur numerik dan kategorikal. Namun, mereka mungkin kehilangan sebagian interpretasi yang ditawarkan oleh pohon keputusan tunggal. Meskipun demikian, akurasi tinggi mereka dalam banyak tugas seringkali lebih penting daripada kekurangan ini dalam praktiknya.

Gradient Boosting dalam Pembelajaran Terawasi

Gradient boosting adalah teknik pembelajaran mesin ensemble yang digunakan untuk masalah regresi dan klasifikasi. Teknik ini membangun model secara bertahap, dan menggeneralisasinya dengan memungkinkan optimasi fungsi kerugian yang dapat didiferensiasi secara sembarang. Pada intinya, gradient boosting melibatkan pembangunan dan penggabungan serangkaian model lemah (biasanya pohon keputusan) untuk menciptakan model prediktif yang kuat.

Proses gradient boosting melibatkan tahapan-tahapan berikut:

1. **Inisialisasi.** Mulailah dengan model awal, yang membuat prediksi, seringkali hanya rata-rata sederhana atau model dasar lainnya.
2. **Hitung residual.** Hitung perbedaan (atau residual) antara nilai yang diamati dan prediksi model saat ini.
3. **Sesuaikan dengan residual.** Bangun model baru yang memprediksi residual dari langkah sebelumnya.
4. **Perbarui prediksi.** Sesuaikan prediksi model utama menggunakan prediksi dari model residual. Ini biasanya dilakukan dengan menambahkan sebagian kecil prediksi model baru ke prediksi model utama.
5. **Iterasi.** Ulangi langkah 2–4 untuk jumlah iterasi yang telah ditentukan atau hingga residual diminimalkan di bawah ambang batas.
6. **Keluaran.** Prediksi akhir adalah jumlah prediksi dari semua model.

Misalkan sebuah perusahaan ingin memprediksi kemungkinan pelanggan membeli produk baru berdasarkan fitur-fitur seperti usia, pendapatan, perilaku pembelian sebelumnya, dan keterlibatan dengan kampanye pemasaran sebelumnya.

Dengan menggunakan gradient boosting, Anda dapat mencapai hal berikut:

- Perusahaan memulai dengan model dasar, mungkin memprediksi bahwa setiap pelanggan memiliki kemungkinan rata-rata untuk membeli.
- Perbedaan antara perilaku pembelian aktual dan prediksi model awal ini dihitung (ini adalah residual).
- Model baru dilatih untuk memprediksi residual ini berdasarkan fitur pelanggan.
- Prediksi model baru ini digabungkan dengan prediksi model awal untuk menghasilkan serangkaian prediksi yang direvisi.
- Proses ini diulang, setiap kali menyempurnakan prediksi dengan melatih model baru pada residual dari model gabungan sebelumnya.
- Setelah beberapa iterasi, perusahaan memiliki model prediktif yang ampuh yang secara efektif menggabungkan wawasan dari semua model individual.

Gradient boosting dikenal karena akurasinya yang tinggi dan kemampuannya untuk menangani berbagai jenis dan struktur data. Salah satu kekuatan utamanya adalah kemampuannya untuk fokus pada kasus-kasus yang sulit diprediksi (kasus dengan residual besar) dalam iterasi berturut-turut, yang mengarah pada peningkatan kinerja keseluruhan. Namun, penyetelan parameter yang cermat diperlukan untuk mencegah overfitting dan memastikan kinerja optimal. Lihat Gambar 5.4 untuk representasi visual dari struktur gradient boosting, yang menggambarkan bagaimana model secara berurutan membangun pembelajar yang lebih lemah untuk membentuk model prediktif yang kuat.

Support Vector Machines dalam Pembelajaran Terawasi

SVM adalah algoritma pembelajaran mesin terawasi yang digunakan terutama untuk tugas klasifikasi, tetapi juga dapat digunakan untuk regresi. Pada intinya, SVM mencoba menemukan hyperplane (atau batas keputusan) terbaik yang memisahkan data ke dalam kelas yang berbeda. Hyperplane terbaik adalah hyperplane yang memaksimalkan margin antara dua kelas. Dalam ruang berdimensi lebih tinggi, SVM mencari hyperplane yang paling baik memisahkan data ke dalam kelas.

Mekanisme SVM dapat dipecah menjadi beberapa tahap:

1. **Memaksimalkan margin.** Temukan hyperplane yang memiliki jarak maksimum ke titik data pelatihan terdekat dari kelas mana pun. Titik data yang terletak paling dekat dengan permukaan keputusan disebut vektor pendukung.
2. **Tangani nonlinieritas dengan trik kernel.** Dalam kasus di mana data tidak dapat dipisahkan secara linier, SVM menggunakan fungsi yang disebut trik kernel untuk mengubah data input ke ruang berdimensi lebih tinggi di mana hyperplane dapat digunakan untuk memisahkan data. Kernel umum meliputi polinomial, fungsi basis radial (RBF), dan sigmoid.
3. **Margin lunak dan regularisasi.** Untuk menangani data yang bising di mana pemisahan lengkap mungkin tidak optimal, SVM memperkenalkan konsep yang dikenal sebagai margin lunak, yang memungkinkan beberapa kesalahan klasifikasi sebagai imbalan untuk margin yang lebih luas dan lebih dapat digeneralisasi. Parameter regularisasi, yang sering disebut sebagai 'C', menentukan pertukaran antara memaksimalkan margin dan meminimalkan kesalahan klasifikasi.
4. **Prediksi.** Untuk input baru, tentukan sisi hyperplane mana yang akan digunakan untuk mengklasifikasikannya ke dalam suatu kategori.

Bayangkan sebuah perusahaan elektronik yang memiliki data tentang interaksi pelanggan dengan kampanye emailnya. Fitur-fitur tersebut mungkin termasuk waktu yang dibutuhkan untuk membuka email, apakah tautan di dalamnya diklik, perilaku pembelian sebelumnya, dan sebagainya. Perusahaan ingin mengklasifikasikan pelanggan ke dalam dua kelompok: mereka yang kemungkinan besar akan membeli produk baru dan mereka yang tidak.

Dengan menggunakan SVM, hal berikut dapat terjadi:

- Algoritma akan mencoba menemukan hyperplane terbaik yang memisahkan pelanggan yang telah melakukan pembelian dari mereka yang belum, berdasarkan fitur interaksi email mereka.

- Jika batas keputusan antara kedua kelompok ini tidak linier, kernel RBF dapat mengubah fitur-fitur tersebut menjadi ruang di mana kelompok-kelompok tersebut dapat dipisahkan oleh hyperplane.
- Setelah model SVM dilatih, ketika kampanye email baru dikirim, perusahaan dapat menggunakan model tersebut untuk memprediksi, berdasarkan metrik interaksi awal, pelanggan mana yang kemungkinan besar akan melakukan pembelian.

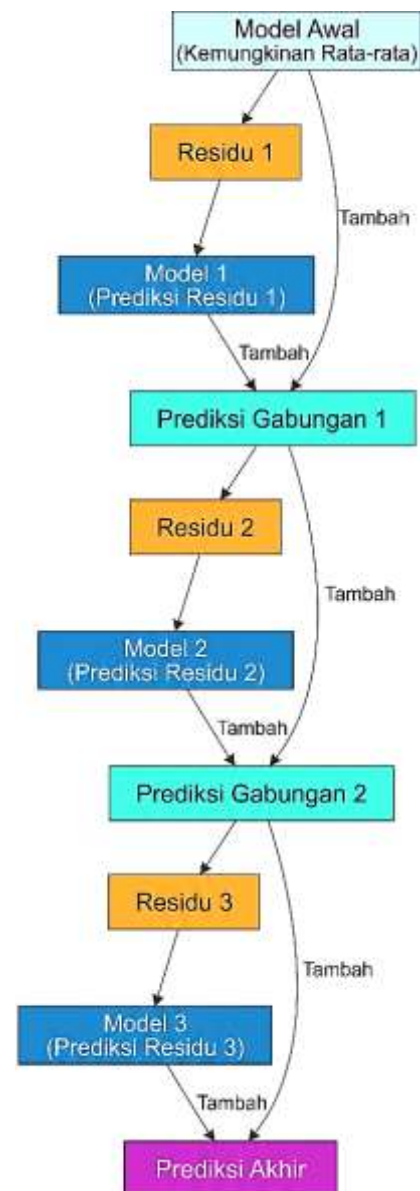
SVM dikenal karena akurasi yang tinggi, kemampuannya untuk menangani data berdimensi tinggi, dan fleksibilitas dalam memodelkan berbagai sumber data. Namun, SVM dapat membutuhkan komputasi yang intensif, terutama dengan sejumlah besar fitur, dan mungkin memerlukan penyetelan dan validasi yang signifikan untuk dioptimalkan untuk aplikasi tertentu.

Jaringan Saraf dalam Pembelajaran Terawasi

Jaringan saraf adalah model komputasi yang terinspirasi oleh struktur dan fungsi sistem saraf biologis. Jaringan saraf terdiri dari node atau "neuron" yang saling terhubung yang memproses dan mengirimkan informasi. Deep learning, subbidang dari machine learning, sering menggunakan jaringan saraf dengan banyak lapisan (dikenal sebagai jaringan saraf dalam) untuk melakukan tugas-tugas seperti pengenalan gambar, pemrosesan bahasa alami, dan banyak lagi. Jaringan ini dapat mempelajari pola-pola rumit dari sejumlah besar data.

Pengoperasian jaringan saraf dapat dipecah menjadi beberapa tahap berikut:

1. **Arsitektur.** Jaringan saraf terdiri dari lapisan input, satu atau lebih lapisan tersembunyi, dan lapisan output. Setiap neuron terhubung ke neuron lain melalui bobot, yang disesuaikan selama pelatihan.
2. **Fungsi aktivasi.** Ketika sebuah neuron menerima input, ia memprosesnya menggunakan fungsi aktivasi. Fungsi aktivasi umum meliputi sigmoid, ReLU (*rectified linear unit*), dan fungsi tanh.
3. **Propagasi maju.** Data input bergerak melalui jaringan, mengalami penjumlahan berbobot dan aktivasi. Proses ini menghasilkan prediksi pada lapisan output.
4. **Perhitungan biaya.** Setelah mendapatkan prediksi, jaringan menghitung kesalahan atau biaya dengan membandingkan prediksi dengan label sebenarnya. Tujuannya adalah untuk meminimalkan kesalahan ini.



Gambar 5.4 Struktur Gradient Boosting.

5. **Propagasi balik.** Kesalahan disebarkan mundur melalui jaringan, menyesuaikan bobot menggunakan algoritma optimasi seperti penurunan gradien.
6. **Iterasi pelatihan.** Langkah 3 hingga 5 diulang secara iteratif hingga model mencapai kinerja yang memuaskan atau sejumlah iterasi (*epoch*) tercapai.

Pertimbangkan platform e-commerce fesyen yang bertujuan untuk memberikan rekomendasi pakaian yang dipersonalisasi kepada penggunanya. Platform tersebut mengumpulkan berbagai data pengguna, seperti perilaku penelusuran, pembelian sebelumnya, kueri pencarian, dan peringkat umpan balik.

Dengan menggunakan jaringan saraf, hal-hal berikut dapat terjadi:

- Fitur input mungkin termasuk pola aktivitas pengguna, seperti seberapa sering mereka mengunjungi situs, kategori mana yang paling sering mereka telusuri, waktu rata-rata yang mereka habiskan di halaman produk, dan sebagainya.
- Jaringan saraf, setelah dilatih dengan data historis, dapat memprediksi produk atau gaya yang kemungkinan besar akan dibeli oleh pengguna tertentu.
- Seiring pengguna terus berinteraksi dengan platform, umpan balik dapat lebih menyempurnakan rekomendasi, memastikan rekomendasi tersebut semakin akurat dan disesuaikan dari waktu ke waktu.

Jaringan saraf menawarkan kemampuan untuk menangkap pola dan hubungan yang rumit dalam data, menjadikannya sangat ampuh untuk tugas-tugas dengan kompleksitas tinggi. Namun, jaringan saraf seringkali membutuhkan data dalam jumlah besar, daya komputasi yang cukup besar, dan penyetelan yang cermat untuk menghindari masalah seperti overfitting. Sifat "kotak hitam" mereka juga dapat menimbulkan tantangan bagi interpretasi dan transparansi dalam pengambilan keputusan.

Pembelajaran Tanpa Pengawasan untuk Pemasaran

Pembelajaran tanpa pengawasan adalah cabang utama lain dari pembelajaran mesin. Tidak seperti pembelajaran terawasi, yang belajar dari contoh berlabel, pembelajaran tanpa pengawasan menemukan pola dan struktur dalam data yang tidak berlabel. Pembelajaran ini berupaya memodelkan distribusi data yang mendasari atau menemukan kelompok contoh serupa dalam data. Di bidang pemasaran, pembelajaran tanpa pengawasan memiliki berbagai macam aplikasi. Metode ini digunakan dalam segmentasi pelanggan, deteksi anomali, pembahasan keranjang belanja, dan banyak lagi. Masing-masing tugas ini dapat dibingkai sebagai masalah pembelajaran tanpa pengawasan di mana tujuannya adalah untuk menemukan pola atau struktur dalam data tanpa target prediksi tertentu. Salah satu penggunaan pembelajaran tanpa pengawasan yang paling umum dalam pemasaran adalah segmentasi pelanggan, di mana pelanggan dikelompokkan berdasarkan kesamaan mereka di berbagai dimensi seperti demografi, perilaku pembelian, atau pola keterlibatan. Pengelompokan K-means dan pengelompokan hierarkis adalah metode populer untuk tugas ini.

Pembahasan keranjang belanja (*Market Basket Analysis*/MBA), aplikasi kunci lain dari pembelajaran tanpa pengawasan dalam pemasaran, bertujuan untuk menemukan asosiasi antar produk berdasarkan data transaksi. Hal ini sering dicapai menggunakan metode

pembelajaran aturan asosiasi seperti algoritma a priori atau algoritma FP-growth. Meskipun pembelajaran tanpa pengawasan dapat mengungkap wawasan berharga yang mungkin terlewatkan dengan metode pengawasan, ia memiliki serangkaian tantangan tersendiri. Ini termasuk menentukan jumlah kluster optimal dalam pembahasan pengelompokan, menangani data yang bising atau berdimensi tinggi, dan menafsirkan hasilnya dengan cara yang bermakna.

Pembelajaran Penguatan untuk Pemasaran

RL adalah jenis pembelajaran mesin di mana agen belajar untuk membuat keputusan dengan berinteraksi dengan lingkungan. Agen mengambil tindakan, menerima umpan balik dalam bentuk hadiah atau hukuman, dan menyesuaikan perilakunya untuk memaksimalkan hadiah kumulatif dari waktu ke waktu. Tujuan RL adalah untuk menemukan kebijakan optimal, yaitu strategi yang menentukan tindakan terbaik yang harus diambil di setiap keadaan lingkungan.

Dalam konteks pemasaran, RL dapat digunakan untuk mempersonalisasi interaksi pelanggan, mengoptimalkan kampanye pemasaran, dan membuat keputusan penawaran waktu nyata dalam periklanan online, di antara aplikasi lainnya. Masing-masing tugas ini dapat dibingkai sebagai masalah RL di mana tujuannya adalah untuk membuat serangkaian keputusan yang memaksimalkan tujuan jangka panjang, seperti CLV, ROI kampanye, atau pendapatan iklan. Misalnya, dalam pemasaran yang dipersonalisasi, lingkungannya bisa berupa pelanggan, keadaannya bisa berupa perilaku dan atribut pelanggan, tindakannya bisa berupa intervensi pemasaran (misalnya, mengirim email, menawarkan diskon), dan imbalannya bisa berupa respons pelanggan terhadap intervensi (misalnya, melakukan pembelian, mengklik tautan). Agen RL kemudian akan belajar mempersonalisasi intervensi pemasaran berdasarkan perilaku dan atribut pelanggan untuk memaksimalkan respons pelanggan dari waktu ke waktu.

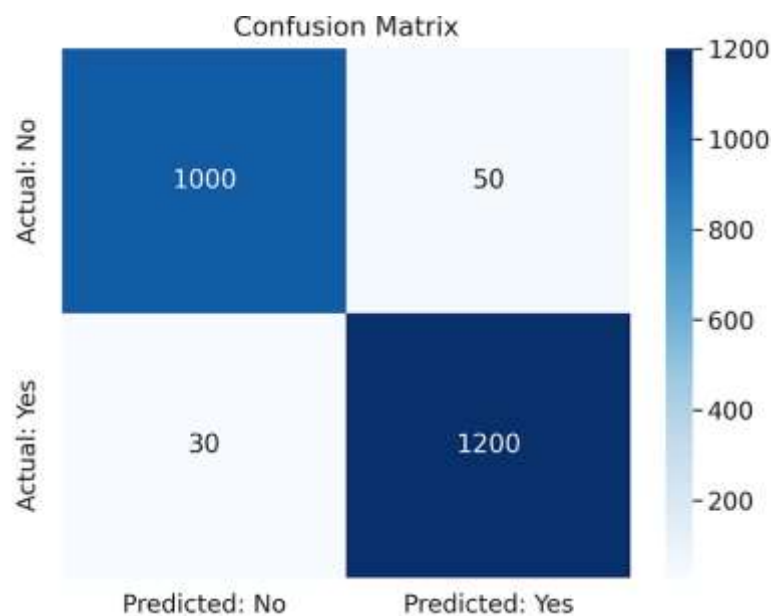
RL memiliki beberapa keunggulan dibandingkan metode pemasaran tradisional. Ia dapat menangani lingkungan yang dinamis, mempertimbangkan efek jangka panjang, dan terus belajar serta beradaptasi dengan informasi baru. Namun, ia juga menghadirkan tantangan, seperti trade-off eksplorasi-eksploitasi, masalah penugasan kredit, dan kebutuhan akan sejumlah besar data interaksi. Tabel 5.1 menyajikan perbedaan yang ditabulasi antara pembelajaran terawasi, tidak terawasi, dan penguatan, membandingkannya di berbagai dimensi seperti persyaratan data, aplikasi, dan hasil. Kerangka kerja komparatif ini membantu memperjelas peran unik RL dalam pemasaran, terutama dalam hal proses pembelajaran berbasis data dan potensinya untuk mengembangkan strategi yang memaksimalkan imbalan jangka panjang. Tabel ini juga membandingkan prasyarat data dan aplikasi tipikal dari setiap paradigma pembelajaran, memberikan perbedaan yang jelas tentang di mana dan bagaimana setiap metode dapat diterapkan dalam bidang pemasaran.

Tabel 5.1 Perbedaan Antara Pembelajaran Terawasi, Tak Terawasi, dan Penguatan dalam Hal Persyaratan Data, Aplikasi, dan Hasil.

Aspek	Pembelajaran Terawasi	Pembelajaran Tanpa Pengawasan	Pembelajaran Penguatan
Persyaratan data	Data berlabel (pasangan input-output)	Data tanpa label	Lingkungan, agen, imbalan
Aplikasi	Klasifikasi Regresi	Pengelompokan Pengurangan Dimensi	Robotika Bermain game Keputusan waktu nyata
Hasil yang diharapkan	Model prediktif berdasarkan data masa lalu	Pola, struktur dalam data	Strategi untuk memaksimalkan imbalan

5.4 EVALUASI DAN SELEKSI MODEL

Setelah model prediktif dilatih, sangat penting untuk mengevaluasi kinerjanya untuk memastikan keandalan dan efektivitasnya dalam membuat prediksi. Memilih metrik dan teknik evaluasi yang tepat tidak hanya membantu dalam menilai kinerja model tetapi juga membantu dalam pemilihan model yang paling sesuai untuk masalah tertentu. Pada bagian ini, kita akan mengeksplorasi berbagai metrik dan teknik evaluasi model.



Gambar 5.5 Matriks Konfusi dan Komponennya.

Akurasi, Presisi, dan Recall Model

Matriks Konfusi

Sebelum membahas lebih lanjut tentang akurasi, presisi, dan recall, ada baiknya untuk memahami konsep matriks konfusi. Matriks konfusi adalah tabel yang digunakan untuk menggambarkan kinerja model klasifikasi pada sekumpulan data yang nilai sebenarnya diketahui (lihat Gambar 5.5). Matriks ini terdiri dari empat nilai:

- **Positif Sejati (TP).** Jumlah positif yang diklasifikasikan dengan benar.
- **Negatif Sejati (TN).** Jumlah negatif yang diklasifikasikan dengan benar.
- **Positif Palsu (FP).** Jumlah negatif yang salah diklasifikasikan sebagai positif.
- **Negatif palsu (FN).** Jumlah kasus positif yang salah diklasifikasikan sebagai negatif.

Contoh Praktis: Kampanye Pemasaran Email

Bayangkan sebuah perusahaan baru saja menjalankan kampanye pemasaran email yang menargetkan 1.000 pelanggannya, mempromosikan produk baru. Tujuan utama kampanye ini adalah untuk membuat pelanggan membeli produk tersebut, sehingga perusahaan telah memberi label pada penerima kampanye berdasarkan tindakan mereka: 'Membeli' (jika mereka membeli produk) atau 'Tidak Membeli' (jika mereka tidak membeli produk). Sekarang, tim pemasaran menggunakan model pembelajaran mesin untuk memprediksi sebelumnya pelanggan mana yang akan membeli produk berdasarkan perilaku pembelian sebelumnya, interaksi dengan email sebelumnya, dan sebagainya. Jadi, untuk setiap pelanggan, model memprediksi 'Akan Membeli' atau 'Tidak Akan Membeli'.

Setelah kampanye berakhir, kita dapat membuat matriks kebingungan untuk memahami seberapa baik prediksi model sesuai dengan hasil aktual (lihat Tabel 5.2).

Tabel 5.2 Matriks Kebingungan Kampanye Pemasaran Email.

	Aktual: Dibeli	Aktual: Tidak Dibeli
Prediksi: Akan Membeli	300 (TP)	50 (FP)
Prediksi: Tidak Akan Membeli	20 (FN)	630 (TN)

Dari matriks ini kita dapat membedakan hal-hal berikut:

- **Positif Sejati (TP).** Tiga ratus pelanggan diprediksi dengan benar akan membeli produk tersebut, dan memang mereka membelinya.
- **Negatif sejati (TN).** Enam ratus tiga puluh pelanggan diprediksi dengan benar tidak akan membeli produk tersebut, dan memang mereka tidak membelinya.
- **Positif palsu (FP).** Lima puluh pelanggan diprediksi akan membeli produk tersebut, tetapi mereka tidak membelinya. Ini berarti model tersebut terlalu optimis untuk pelanggan-pelanggan ini.
- **Negatif palsu (FN).** Dua puluh pelanggan diprediksi tidak akan membeli produk tersebut, tetapi mereka akhirnya membelinya. Ini adalah peluang yang terlewatkan karena model tidak mengharapkan mereka untuk melakukan konversi, tetapi mereka melakukannya.

Dengan memahami angka-angka ini, tim pemasaran dapat menyempurnakan strategi mereka. Misalnya, mereka mungkin ingin menyelidiki lebih lanjut profil dari 50 pelanggan FP: Mengapa model tersebut mengira mereka akan membeli? Apakah ada karakteristik atau perilaku umum di antara mereka? Wawasan semacam ini dapat membantu mengoptimalkan kampanye di masa mendatang dan meningkatkan model prediksi. Sekarang, mari kita lanjutkan dengan metriknya.

Akurasi

Akurasi adalah salah satu metrik yang paling mudah dipahami dalam bidang pemodelan prediktif. Ini mengukur proporsi prediksi yang benar dari total prediksi yang dibuat:

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi yang Benar}}{\text{Jumlah Total Prediksi}}$$

Meskipun akurasi adalah metrik yang umum digunakan, mungkin tidak selalu menjadi pilihan terbaik, terutama ketika berurusan dengan dataset yang tidak seimbang di mana satu kelas secara signifikan lebih banyak daripada kelas lainnya.

Presisi

Presisi adalah rasio pengamatan positif yang diprediksi dengan benar terhadap total prediksi positif. Ini menjawab pertanyaan: Dari semua label positif yang kita prediksi, berapa banyak yang benar?

$$\text{Presisi} = \frac{\text{Positif Sejati}}{\text{Positif Sejati} + \text{Positif Palsu}}$$

Recall (Sensitivitas)

Recall menghitung rasio pengamatan positif yang diprediksi dengan benar terhadap semua pengamatan positif yang sebenarnya. Pertanyaannya adalah: Dari semua label positif yang sebenarnya, berapa banyak yang berhasil kita prediksi dengan benar?

$$\text{Ingat} = \frac{\text{Positif Sejati}}{\text{Positif Sejati} + \text{Negatif Palsu}}$$

Dalam banyak skenario, terdapat pertukaran antara presisi dan recall. Presisi tinggi menunjukkan tingkat false positive yang rendah, sedangkan recall tinggi menunjukkan bahwa pengklasifikasi menangkap sebagian besar instance positif.

Tantangan dengan Akurasi dalam Dataset yang Tidak Seimbang

Bayangkan sebuah dataset di mana 95% sampel termasuk dalam Kelas A, dan hanya 5% termasuk dalam Kelas B. Bahkan model naif yang selalu memprediksi Kelas A akan mencapai akurasi 95%. Namun, model ini akan sepenuhnya tidak efektif dalam memprediksi Kelas B, yang mungkin sangat menarik (misalnya, memprediksi timbulnya penyakit di mana sebagian besar sampel adalah "sehat").

Memperkenalkan F1-Score. Untuk mengatasi kekurangan akurasi dalam skenario seperti itu, kami memperkenalkan F1-score. F1-score adalah rata-rata harmonik dari presisi dan recall, dan menawarkan ukuran yang lebih seimbang ketika kelas tidak seimbang.

$$\text{Skor FI} = 2 \times \frac{\text{Penarikan Presisi}}{\text{Presisi} + \text{Ingat}}$$

di mana

- Presisi adalah jumlah hasil positif yang benar dibagi dengan jumlah semua hasil positif (termasuk yang salah diklasifikasikan).
- Recall (atau sensitivitas) adalah jumlah hasil positif yang benar dibagi dengan jumlah hasil positif yang seharusnya dikembalikan.

Nilai F1-score berkisar dari 0 hingga 1, di mana 1 menunjukkan presisi dan recall yang sempurna, dan 0 menunjukkan tidak ada presisi maupun recall. F1-score memberikan pandangan yang lebih holistik tentang kinerja model, terutama ketika data tidak seimbang.

Teknik over-sampling minoritas sintetis (SMOTE). Pendekatan lain untuk mengatasi dataset yang tidak seimbang adalah dengan menyeimbangkannya dengan membuat sampel sintetis. SMOTE adalah salah satu teknik tersebut.

Berikut adalah gambaran singkat tentang cara kerja SMOTE:

1. Untuk setiap instance dalam kelas minoritas, dipilih sekumpulan tetangga terdekatnya.
2. Berdasarkan tetangga-tetangga ini, sampel sintetis dibuat dengan memilih perbedaan antara fitur-fitur dari instance yang sedang dipertimbangkan dan tetangganya, mengalikan perbedaan ini dengan angka acak antara 0 dan 1, lalu menambahkannya ke instance tersebut.
3. Hal ini secara efektif menciptakan instance sintetis yang sedikit berbeda dari aslinya.

Dengan mengulangi metode ini, SMOTE menciptakan dataset yang seimbang di mana kelas minoritas telah di-oversampling. Setelah itu, model apa pun dapat dilatih pada dataset baru ini.

Jaringan adversarial generatif (GAN) untuk dataset yang tidak seimbang. Dalam mengatasi tantangan yang terkait dengan dataset yang tidak seimbang, GAN telah muncul sebagai alat yang ampuh. GAN terdiri dari dua jaringan saraf generator dan diskriminator yang dilatih secara bersamaan dalam pengaturan kompetitif di mana generator bertujuan untuk menciptakan instance data yang tidak dapat dibedakan dari yang sebenarnya, dan diskriminator berupaya untuk membedakan antara keduanya.

Berikut adalah garis besar bagaimana GAN dapat digunakan untuk menangani dataset yang tidak seimbang:

1. Jaringan generator menerima noise acak dan menghasilkan titik data sintetis, bertujuan untuk mereplikasi distribusi kelas minoritas.
2. Jaringan diskriminator dilatih untuk membedakan antara instance nyata dari kelas minoritas dan instance sintetis yang dibuat oleh generator.
3. Melalui pelatihan iteratif, generator belajar untuk menghasilkan data yang semakin realistis, sementara diskriminator menjadi lebih baik dalam membedakan data sintetis dari data nyata.
4. Pada akhirnya, generator menghasilkan data sintetis berkualitas tinggi yang dapat ditambahkan ke kelas minoritas, sehingga memperkaya dataset dan membantu menyeimbangkan kelas.
5. Dataset yang diperkaya kemudian dapat digunakan untuk melatih model prediktif yang kurang bias terhadap kelas mayoritas dan memiliki kinerja generalisasi yang lebih baik pada data yang belum pernah dilihat sebelumnya.

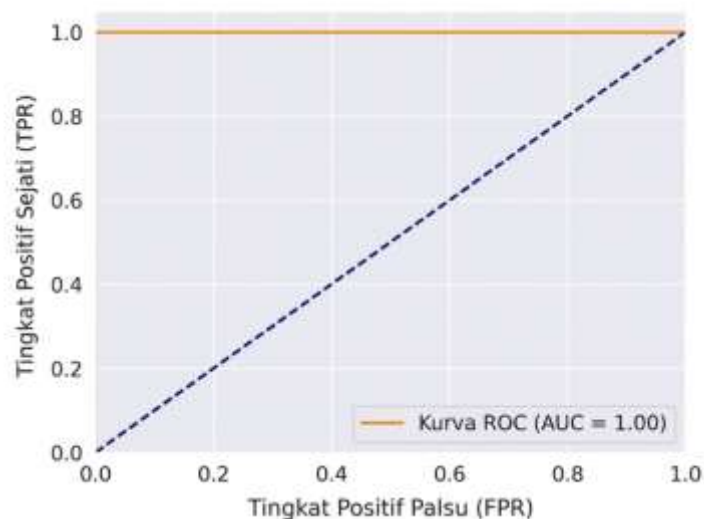
Penggunaan GAN untuk menghasilkan instance sintetis dari kelas minoritas dapat meningkatkan sensitivitas model terhadap kelas minoritas tanpa kehilangan spesifisitas. Teknik ini telah menunjukkan potensi dalam berbagai aplikasi di mana ketidakseimbangan kelas merupakan masalah yang signifikan.

Meskipun akurasi adalah metrik yang berharga, penting untuk mewaspadai keterbatasannya, terutama dalam konteks dataset yang tidak seimbang. Memanfaatkan metrik lain seperti skor F1 dan teknik seperti SMOTE dapat memberikan pemahaman yang lebih komprehensif tentang kinerja model dan meningkatkan kemampuan prediksi pada dataset tersebut. Selain itu, GAN menghadirkan pendekatan inovatif untuk menghasilkan data sintetis untuk kelas minoritas, meningkatkan keseimbangan dalam dataset. Aplikasi dan metodologi GAN dalam konteks pemasaran dibahas lebih detail di Bab 12.

Kurva ROC dan AUC

Kurva karakteristik operasi penerima (ROC). Kurva ROC adalah representasi grafis dari kinerja pengklasifikasi biner. Kurva ini memplot tingkat positif sejati (*recall*) terhadap tingkat positif palsu di berbagai nilai ambang batas.

Area di bawah kurva (AUC). AUC mengukur seluruh area dua dimensi di bawah seluruh kurva ROC. AUC memberikan ukuran agregat kinerja di semua ambang klasifikasi yang mungkin. Model dengan kapasitas prediksi sempurna akan memiliki AUC 1, dan model yang memprediksi murni secara kebetulan akan memiliki AUC 0,5. Lihat Gambar 5.6 untuk ilustrasi kurva ROC dengan nilai AUC yang sesuai, yang menunjukkan bagaimana pengaturan ambang batas yang berbeda memengaruhi pertukaran antara tingkat positif sejati dan positif palsu.



Gambar 5.6 Kurva ROC dan AUC.

Teknik Validasi Silang

Metode Holdout

Metode holdout melibatkan pemisahan dataset menjadi dua set yang berbeda: set pelatihan dan set pengujian. Model dilatih pada set pelatihan dan dievaluasi pada set pengujian. Metode ini sederhana dan cepat, tetapi evaluasinya dapat memiliki varians yang

tinggi karena penilaian sangat bergantung pada titik data mana yang berakhir di set pengujian. Untuk mengurangi hal ini, seringkali pemisahan dilakukan beberapa kali dan hasilnya dirata-ratakan, atau alternatifnya, metode yang lebih canggih seperti validasi silang digunakan.

Validasi Silang K-Fold

Dalam validasi silang K-fold, dataset dipartisi secara acak menjadi k subset berukuran sama. Dari k subset, satu dipertahankan untuk menguji model, dan k-1 subset yang tersisa digunakan untuk pelatihan. Proses ini diulang k kali, dengan setiap subset berfungsi sebagai set pengujian sekali. Hasil k tersebut kemudian dirata-ratakan untuk menghasilkan satu estimasi tunggal.

Validasi Silang Leave-One-Out

Dalam validasi silang *leave-one-out* (LOOCV), satu titik data digunakan sebagai set pengujian sementara titik data yang tersisa merupakan set pelatihan. Proses ini diulang secara iteratif sehingga setiap titik data berfungsi sebagai titik pengujian sekali. Meskipun LOOCV dapat memakan biaya komputasi yang tinggi, metode ini memanfaatkan data secara efisien karena setiap titik data digunakan untuk pelatihan dan pengujian.

Validasi Silang Bertingkat

Untuk dataset di mana variabel hasil tidak seimbang, validasi silang bertingkat memastikan bahwa setiap lipatan merupakan representasi yang baik dari keseluruhan dengan memiliki proporsi sampel yang sama dari setiap kelas target seperti set lengkap.

Pemisahan Deret Waktu

Dalam pemisahan deret waktu, data diurutkan secara kronologis dan dibagi menjadi set pelatihan dan pengujian beberapa kali. Pembagian awal mungkin menggunakan 70% data pertama sebagai data pelatihan dan 30% berikutnya sebagai data pengujian. Untuk iterasi berikutnya, set pelatihan mungkin bergeser maju dalam waktu, termasuk bagian data berikutnya, dan seterusnya. Metode ini sangat penting untuk data deret waktu di mana asumsinya adalah bahwa informasi masa lalu digunakan untuk memprediksi peristiwa di masa mendatang. Tidak seperti metode validasi silang standar, metode ini memastikan bahwa set pelatihan selalu mendahului set pengujian dalam waktu, dengan menghormati urutan temporal data. Lihat Tabel 5.3 untuk perbandingan komprehensif tentang manfaat dan kekurangan berbagai teknik validasi silang, termasuk metode pembagian deret waktu dan bagaimana metode ini secara unik diadaptasi untuk menangani data dengan ketergantungan temporal.

Tabel 5.3 Manfaat dan Kekurangan Berbagai Teknik Validasi Silang.

Teknik	Keunggulan	Kekurangan
Holdout	Sederhana dan cepat	Varians tinggi jika dataset tidak besar
K-fold	Mengurangi overfitting Estimasi lebih andal daripada holdout	Mahal secara komputasi untuk k besar
Leave-one-out	Memanfaatkan semua titik data Baik untuk dataset kecil	Sangat mahal secara komputasi untuk dataset besar

Stratified K-fold	Mempertahankan distribusi kelas, mengurangi bias	Mahal secara komputasi
Pemisahan deret waktu	Mempertimbangkan urutan waktu Ideal untuk data deret waktu	Tidak mengacak data Mungkin tidak menangkap semua pola

Kompleksitas Model dan Overfitting

Kompleksitas model mengacu pada jumlah parameter dalam suatu model dan kerumitan strukturnya. Model yang lebih kompleks akan memiliki lebih banyak parameter dan dapat menyesuaikan berbagai fungsi yang lebih luas. Namun, kompleksitas yang berlebihan mungkin tidak selalu bermanfaat. Overfitting terjadi ketika suatu model terlalu kompleks dan mulai menangkap noise dalam data daripada pola yang mendasarinya. Model yang mengalami overfitting akan memiliki kesalahan pelatihan yang sangat rendah tetapi akan berkinerja buruk pada data yang belum pernah dilihat sebelumnya karena telah menyesuaikan diri terlalu dekat dengan dataset pelatihan. Untuk mengurangi overfitting, teknik seperti regularisasi, pemangkasan, dan penggunaan model yang lebih sederhana dapat diterapkan. Singkatnya, evaluasi dan pemilihan model merupakan langkah penting dalam alur kerja pembelajaran mesin. Menggunakan metrik dan teknik yang tepat memastikan bahwa model yang dikembangkan kuat, andal, dan sesuai untuk tugas yang dimaksudkan, membuka jalan bagi prediksi yang efektif dan hasil yang berwawasan.

5.5 DAMPAK CHURN PADA BISNIS

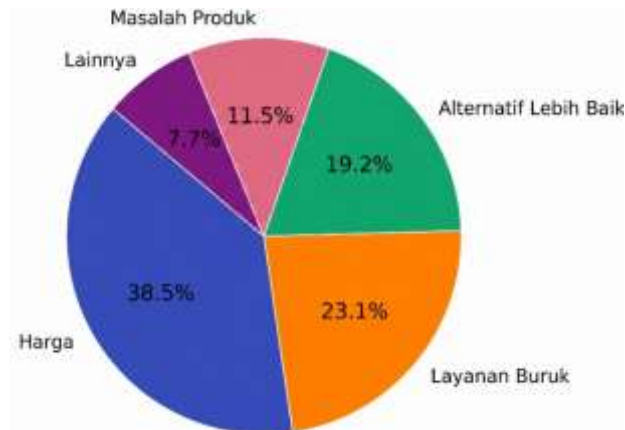
Peran ilmu data dalam pemasaran telah melengkapi teknik tradisional, memungkinkan wawasan yang lebih tepat dan dapat ditindaklanjuti tentang perilaku pelanggan. Di antara ini, prediksi churn, penilaian CLV, dan pemodelan kecenderungan menonjol karena nilainya dalam memahami, mempertahankan, dan memaksimalkan potensi basis pelanggan. Mari kita telusuri konsep-konsep ini dan signifikansinya dalam lanskap pemasaran berbasis data saat ini.

Memahami Churn dan Pentingnya

Churn, dalam konteks perilaku pelanggan, mengacu pada saat pelanggan berhenti berbisnis atau mengakhiri hubungan dengan perusahaan. Tingkat churn, metrik penting, menunjukkan persentase pelanggan yang berhenti berlangganan selama periode waktu tertentu. Memahami dan mengurangi pelanggan berhenti berlangganan sangat penting karena beberapa alasan:

- **Efisiensi biaya.** Memperoleh pelanggan baru seringkali lebih mahal daripada mempertahankan pelanggan yang sudah ada. Oleh karena itu, mengurangi pelanggan berhenti berlangganan dapat menghasilkan penghematan biaya yang signifikan.
- **Dampak pendapatan.** Pelanggan tetap seringkali memberikan kontribusi lebih besar terhadap pendapatan perusahaan. Mereka mungkin membeli lebih banyak dan bahkan dapat bertindak sebagai pendukung merek, mendatangkan pelanggan baru.
- **Umpan balik.** Membahas alasan di balik pelanggan berhenti berlangganan dapat memberikan wawasan berharga tentang area peningkatan, baik dalam penawaran produk, layanan pelanggan, atau aspek operasional lainnya.

Dengan memprediksi pelanggan mana yang paling mungkin berhenti berlangganan, bisnis dapat secara proaktif mengatasi masalah dan menerapkan strategi retensi yang disesuaikan dengan kebutuhan pelanggan individu (lihat Gambar 5.7).



Gambar 5.7 Alasan Umum Pelanggan Berhenti Berlangganan.

Perhitungan dan Aplikasi CLV

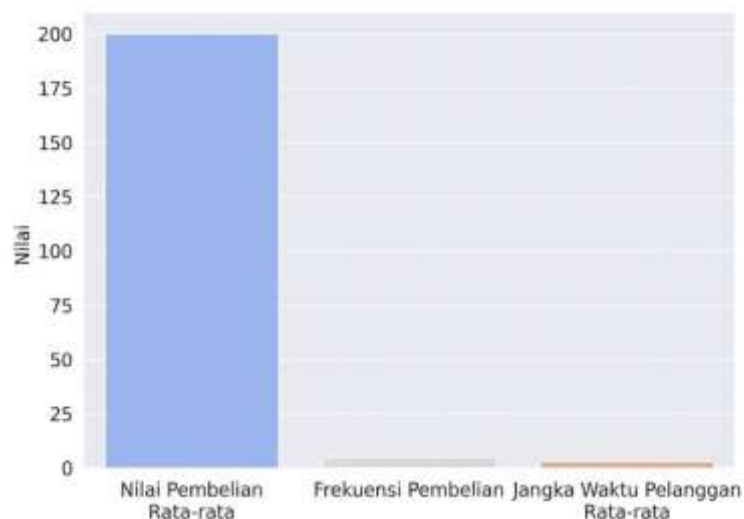
CLV menunjukkan total laba bersih yang diantisipasi perusahaan dari pelanggan tertentu selama masa hubungan mereka. Ini memberikan estimasi moneter atas nilai yang dibawa pelanggan sepanjang masa hidup mereka sebagai pelanggan suatu bisnis tertentu.

Perhitungan CLV Sederhana

Cara dasar untuk menghitung CLV adalah dengan menggunakan rumus:

$$CLV = (\text{Nilai Pesanan Rata-rata}) \times (\text{Jumlah Penjualan Berulang}) \times (\text{Rata-rata Waktu Retensi dalam Bulan atau Tahun})$$

Meskipun rumus ini memberikan cara yang mudah untuk memperkirakan CLV, perlu dicatat bahwa banyak bisnis sering menggunakan model yang lebih rumit untuk mendapatkan pemahaman yang lebih mendalam (lihat Gambar 5.8).



Gambar 5.8 Komponen Nilai Seumur Hidup Pelanggan (*Customer Lifetime Value/CLV*).

Model CLV Tingkat Lanjut

- **CLV Historis.** Menggunakan laba kotor dari semua pembelian pelanggan di masa lalu untuk memprediksi perilaku di masa mendatang.
- **CLV Prediktif.** Menggunakan teknik statistik tingkat lanjut dan pembelajaran mesin untuk memperkirakan potensi nilai pelanggan di masa mendatang berdasarkan perilaku mereka di masa lalu dan saat ini.
- **Model RFM (*recency, frequency, monetary*) tradisional.** Berfokus pada tiga faktor: seberapa baru pelanggan melakukan pembelian, seberapa sering mereka melakukan pembelian, dan berapa banyak yang mereka belanjakan rata-rata.
- **Model gamma-gamma nilai moneter.** Diterapkan ketika ada transaksi berulang dengan nilai moneter yang bervariasi.

Memahami CLV dapat mengubah bisnis secara signifikan:

- **Segmentasi dan personalisasi.** Pelanggan dengan CLV yang lebih tinggi dapat diidentifikasi dan diperlakukan dengan strategi pemasaran khusus, memastikan keterlibatan dan loyalitas mereka yang berkelanjutan.
- **Alokasi sumber daya.** Perusahaan dapat memutuskan di mana mengalokasikan sumber daya mereka secara lebih efisien, lebih fokus pada pelanggan dengan CLV tinggi.
- **Pengembangan produk.** Wawasan yang diperoleh dari CLV dapat memandu pengembangan produk, menyesuaikan penawaran dengan segmen pelanggan yang paling berharga.
- **Manajemen risiko.** Perusahaan dapat mengantisipasi fluktuasi pendapatan berdasarkan perubahan yang diproyeksikan dalam CLV.

Meskipun rumus sederhana memberikan pemahaman dasar tentang CLV, menyelami lebih dalam model-model canggih dapat menawarkan bisnis perspektif yang lebih halus dan komprehensif, memungkinkan mereka untuk melayani pelanggan mereka dengan lebih baik dan mengoptimalkan operasi mereka.

Penilaian Kecenderungan dan Aplikasi Pemasarannya

Skor kecenderungan mengukur probabilitas bahwa pelanggan akan melakukan tindakan tertentu. Bagi pemasar, ini biasanya diterjemahkan ke kemungkinan pelanggan melakukan pembelian, menanggapi iklan atau kampanye, atau memutuskan untuk pergi (*churn*).

Menghitung skor kecenderungan. Proses menghitung skor kecenderungan sering dimulai dengan mengumpulkan data pelanggan yang relevan, seperti riwayat pembelian mereka, demografi, interaksi dengan merek, dan sinyal perilaku lainnya.

Salah satu metodologi yang umum digunakan untuk menentukan skor kecenderungan adalah regresi logistik. Berikut adalah gambaran singkatnya:

- Regresi logistik dalam penilaian kecenderungan. Regresi logistik adalah metode statistik yang memprediksi probabilitas hasil biner. Ketika diterapkan dalam pemasaran, hasilnya bisa berupa 'Akan Membeli' versus 'Tidak Akan Membeli', 'Akan

Merespons' versus 'Tidak Akan Merespons', atau 'Akan Berhenti Berlangganan' versus 'Akan Tetap Berlangganan', dan lain sebagainya.

Hasilnya (atau variabel dependen) bersifat biner, dan prediktornya (atau variabel independen) dapat berupa kontinu, kategorikal, atau campuran keduanya. Model ini memberikan koefisien untuk setiap prediktor, yang menunjukkan kekuatan dan arah hubungan dengan hasil. Probabilitas yang dihitung dari model regresi logistik membentuk skor kecenderungan (lihat Tabel 5.4).

Tabel 5.4 Kasus Penggunaan dan Manfaat Penilaian Kecenderungan dalam Pemasaran.

Aplikasi	Keuntungan
Kampanye pemasaran yang ditargetkan	Membantu menargetkan pelanggan yang lebih mungkin merespons, meningkatkan ROI
Alokasi sumber daya	Memastikan sumber daya diarahkan kepada pelanggan bernilai tinggi
Strategi retensi pelanggan	Mengidentifikasi pelanggan yang berisiko dan memberikan informasi untuk strategi retensi
Rekomendasi produk	Mempersonalisasi pengalaman belanja berdasarkan kemungkinan pembelian
Strategi penetapan harga	Mengoptimalkan harga berdasarkan kemungkinan pembelian pelanggan

Aplikasi Skor Kecenderungan

- **Kampanye pemasaran yang ditargetkan.** Dengan mengidentifikasi pelanggan dengan kecenderungan tinggi untuk merespons secara positif terhadap suatu kampanye, bisnis dapat mengoptimalkan pengeluaran dan upaya pemasaran mereka.
- **Penjualan silang dan penjualan tambahan.** Memahami kecenderungan pelanggan untuk membeli produk terkait dapat memberikan informasi tentang strategi penjualan silang dan penjualan tambahan, memaksimalkan pendapatan.
- **Strategi retensi.** Pelanggan yang diidentifikasi memiliki kecenderungan tinggi untuk berhenti berlangganan dapat ditargetkan dengan kampanye atau penawaran retensi khusus.
- **Rekomendasi produk.** Platform e-commerce sering menggunakan skor kecenderungan untuk merekomendasikan produk yang kemungkinan besar akan dibeli pelanggan berdasarkan perilaku penelusuran dan pembelian mereka.

Singkatnya, alat-alat seperti prediksi churn, evaluasi CLV, dan pemodelan kecenderungan merupakan komponen penting dari perangkat kerja pemasar. Terjalin dengan metodologi berbasis data lainnya, alat-alat ini memberikan pemahaman holistik tentang dinamika konsumen, memberdayakan bisnis untuk memprediksi kebutuhan, menyempurnakan pendekatan, dan meningkatkan kepuasan pelanggan serta hasil keuangan.

5.6 LANSKAP KOMERSIAL

Lanskap komersial telah menyaksikan lonjakan variasi dan pilihan produk yang belum pernah terjadi sebelumnya selama beberapa dekade terakhir. Secara historis, konsep pembahasan pasar berakar pada toko-toko fisik tradisional. Peritel, yang ingin memaksimalkan penjualan, sering mengamati kebiasaan pembelian konsumen untuk mengetahui pola-pola tertentu. Misalnya, mereka mungkin memperhatikan bahwa orang yang membeli roti juga cenderung membeli mentega. Pengamatan tersebut memengaruhi tata letak toko, dengan produk pelengkap ditempatkan bersama untuk mendorong pembelian impulsif. Memasuki era digital, pengamatan tradisional ini berubah menjadi pembahasan berbasis data yang canggih. Dengan platform belanja online yang merekam setiap klik, tampilan, dan pembelian, terdapat harta karun data yang menunggu untuk digali guna menemukan pola.

Mengapa aturan asosiasi sangat penting. Aturan asosiasi, yang membentuk inti dari pembahasan keranjang belanja, berfungsi sebagai jembatan antar produk, mengungkapkan hubungan dan pola tersembunyi dalam kumpulan data besar. Pentingnya aturan ini dapat diringkas sebagai berikut:

- **Penempatan produk yang lebih baik.** Sama seperti contoh roti dan mentega, toko online dapat menggabungkan produk bersama atau menyarankan produk tersebut sebagai 'Sering Dibeli Bersama' berdasarkan aturan asosiasi.
- **Promosi yang disesuaikan.** Jika ditemukan asosiasi yang kuat antara produk A dan B, pelanggan yang membeli A dapat diberikan diskon atau penawaran promosi untuk B, sehingga meningkatkan kemungkinan penjualan tambahan.
- **Manajemen inventaris.** Dengan memahami produk mana yang sering dibeli bersama, bisnis dapat mengelola inventaris mereka dengan lebih efisien, memastikan bahwa jika satu produk tersedia, produk pelengkap juga tersedia.
- **Pengalaman pengguna yang lebih baik.** Pelanggan menghargai pengalaman berbelanja yang lancar di mana mereka dapat dengan mudah menemukan produk terkait. Aturan asosiasi dapat mendukung rekomendasi tersebut, membuat perjalanan berbelanja menjadi intuitif dan efisien.

Konteks historis. Lahirnya penambangan aturan asosiasi sebagian besar dikaitkan dengan karya yang dilakukan oleh Rakesh Agrawal, Tomasz Imieliński, dan Arun Swami pada awal tahun 1990-an. Mereka memperkenalkan algoritma untuk menentukan keteraturan antara produk dalam data transaksi skala besar yang dicatat oleh sistem titik penjualan di supermarket. Karya mereka meletakkan dasar bagi banyak mesin rekomendasi dan alat pembahasan keranjang belanja yang digunakan saat ini.

Kesimpulannya, seiring dengan terus berkembang dan meluasnya pasar digital, pentingnya pembahasan keranjang belanja dan sistem rekomendasi semakin meningkat. Dengan memahami dan menerapkan prinsip-prinsip aturan asosiasi, bisnis tidak hanya dapat mendorong penjualan tetapi juga menawarkan pengalaman berbelanja yang tak tertandingi kepada pelanggan mereka.

Prinsip-prinsip Aturan Asosiasi

Pada intinya, aturan asosiasi adalah hubungan "Jika-Maka" antara dua set item. Misalnya, aturan {Bawang Bombay, Kentang} → {Burger} menunjukkan bahwa jika seseorang membeli bawang bombay dan kentang, kemungkinan besar mereka juga akan membeli burger.

Metrik Utama

- **Dukungan.** Mewakili proporsi transaksi dalam dataset yang berisi item tertentu atau kombinasi item. Ini membantu menyaring item atau kombinasi item yang jarang muncul.
- **Kepercayaan.** Menunjukkan kemungkinan bahwa item Y dibeli ketika item X dibeli. Ini mengukur keandalan inferensi.
- **Lift.** Ini menunjukkan peningkatan rasio penjualan item Y ketika item X terjual. Nilai lift lebih besar dari 1 menunjukkan bahwa item Y kemungkinan besar akan dibeli bersama dengan item X, sedangkan nilai kurang dari 1 menunjukkan bahwa item tersebut tidak mungkin dibeli bersama.

Algoritma Apriori dan Pembahasan Keranjang Belanja

Ini adalah algoritma populer yang digunakan untuk mengidentifikasi kumpulan item yang sering muncul dalam basis data dan merupakan dasar dari pembahasan keranjang belanja. Prinsipnya sederhana: jika suatu kumpulan item sering muncul, maka semua subsetnya juga harus sering muncul.

Langkah-langkah

1. Tentukan dukungan kumpulan item dalam basis data transaksional, dan pilih ambang batas dukungan minimum.
2. Hasilkan kumpulan item yang lebih besar menggunakan kumpulan item yang sering muncul yang diidentifikasi pada langkah sebelumnya.
3. Ulangi proses hingga tidak ada kumpulan item yang lebih besar yang dapat dibentuk.

Dengan menggunakan prinsip-prinsip seperti algoritma Apriori, pembahasan keranjang belanja bertujuan untuk menemukan hubungan antara produk yang dibeli bersama. Ini digunakan secara luas di bidang ritel untuk menginformasikan berbagai strategi, mulai dari desain tata letak toko hingga penggabungan promosi.

Tabel 5.5 memberikan cuplikan data keranjang belanja sampel, yang merinci item yang dibeli dalam transaksi individual. Ini menjadi dasar untuk menerapkan algoritma Apriori untuk menentukan kumpulan item yang sering muncul. Selanjutnya, Tabel 5.6 mengilustrasikan aturan asosiasi yang dihasilkan dari data, lengkap dengan metrik seperti support, confidence, dan lift. Tabel-tabel ini menunjukkan penerapan praktis algoritma Apriori dari eksplorasi data awal hingga ekstraksi aturan asosiasi yang bermakna, yang sangat berharga untuk menginformasikan strategi ritel.

Tabel 5.5 Contoh Data Keranjang Belanja.

ID Transaksi	Barang yang Dibeli
1	Roti, Susu
2	Roti, Popok, Bir, Telur

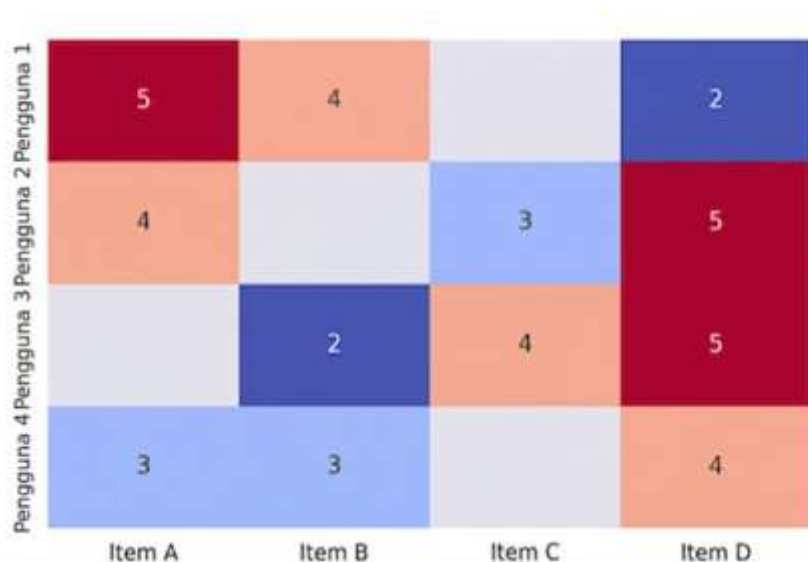
3	Susu, Popok, Bir, Coca-Cola
4	Roti, Susu, Popok, Bir
5	Roti, Susu, Popok, Coca-Cola

Tabel 5.6 Aturan Asosiasi yang Dihasilkan.

Aturan Asosiasi	Dukungan	Kepercayaan diri	Angkat
Roti -> Susu	0.6	0,75	1,25
Bir, Popok -> Susu	0.2	0,5	1,5
Susu, Roti -> Popok	0.4	0,67	1,2
Popok -> Bir	0.5	0,62	1,4

Penyaringan Kolaboratif dalam Sistem Rekomendasi

Penyaringan kolaboratif (CF) didasarkan pada gagasan bahwa pengguna yang telah sepakat di masa lalu cenderung sepakat di masa depan tentang preferensi untuk item tertentu. Ini melibatkan prediksi minat pengguna dengan mengumpulkan informasi preferensi atau selera dari banyak pengguna (berkolaborasi). Gambar 5.9 memberikan representasi visual dari matriks peringkat item pengguna, komponen fundamental dari penyaringan kolaboratif.



Gambar 5.9 Matriks Peringkat Item Pengguna untuk Pemfilteran Kolaboratif.

Jenis-jenis CF

- **Berbasis pengguna.** Ini menemukan pengguna yang mirip dengan pengguna target dan merekomendasikan item berdasarkan apa yang disukai pengguna serupa tersebut.
- **Berbasis item.** Ini merekomendasikan item dengan membandingkan konten item dan profil pengguna, dengan konten yang dijelaskan dalam beberapa deskriptor yang melekat pada item (misalnya, sebuah buku dapat dijelaskan oleh penulisnya, penerbitnya, dan lain-lain.).

Tantangan dalam Penyaringan Kolaboratif

- **Skalabilitas.** Seiring bertambahnya jumlah pengguna dan item, kompleksitas komputasi untuk menemukan pengguna atau item yang serupa juga meningkat.
- **Keterbatasan data.** Dalam banyak skenario dunia nyata, matriks item pengguna yang digunakan dalam CF bersifat jarang, artinya sebagian besar pengguna hanya berinteraksi dengan sebagian kecil dari keseluruhan item, sehingga sulit untuk menemukan kesamaan yang dapat diandalkan.
- **Cold start.** Pengguna atau item baru yang baru saja masuk ke sistem tidak akan memiliki data yang cukup bagi sistem untuk memberikan rekomendasi yang andal. Kekurangan data ini disebut masalah cold start.

Keuntungan Pemfilteran Kolaboratif

- **Personalisasi.** Rekomendasi disesuaikan dengan pengguna individu berdasarkan perilaku dan preferensi mereka di masa lalu.
- **Tidak perlu metadata item.** Tidak seperti pemfilteran berbasis konten, CF tidak bergantung pada atribut item; ia murni beroperasi berdasarkan perilaku pengguna, sehingga serbaguna di berbagai jenis item.
- **Keragaman.** Karena CF mempertimbangkan perilaku pengguna yang serupa, ia berpotensi untuk menyarankan item yang belum secara eksplisit ditunjukkan minatnya oleh pengguna target tetapi mungkin masih menarik.

Konsep cold start. Masalah cold start muncul ketika pengguna atau item baru bergabung dengan sistem. Bagi pengguna baru, karena sistem tidak memiliki pengetahuan sebelumnya tentang preferensi mereka, membuat rekomendasi yang andal menjadi sulit. Ini disebut cold start pengguna. Demikian pula, untuk item baru yang belum diinteraksi oleh pengguna, sistem kesulitan untuk menyarankan item tersebut kepada pengguna yang relevan, yang dikenal sebagai item cold start. Solusi untuk masalah cold start seringkali melibatkan model hibrida atau memanfaatkan penyaringan berbasis konten hingga data interaksi yang cukup terkumpul.

Sistem Rekomendasi Berbasis Konten dan Hibrida

Sistem ini menggunakan fitur item untuk merekomendasikan item tambahan yang serupa dengan apa yang disukai pengguna, berdasarkan tindakan sebelumnya atau umpan balik eksplisit. Misalnya, jika pengguna menyukai film yang termasuk dalam genre aksi, sistem akan merekomendasikan film lain dari genre yang sama. Sistem ini menggabungkan penyaringan kolaboratif dan berbasis konten untuk memberikan rekomendasi. Sistem hibrida dapat diimplementasikan dengan berbagai cara:

- Memisahkan model rekomendasi dan menggabungkannya
- Menggabungkan metode kolaboratif dan berbasis konten ke dalam satu model
- Menyatukan model ke dalam satu model

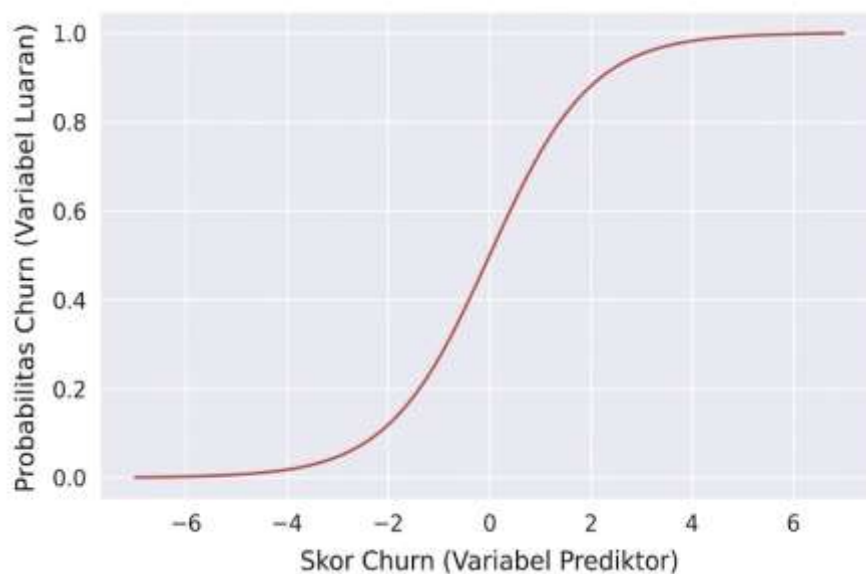
Model hibrida, dalam banyak kasus, memberikan rekomendasi yang lebih akurat daripada model kolaboratif atau berbasis konten murni karena mampu mengatasi keterbatasan yang melekat pada kedua model tersebut.

Kesimpulannya, pembahasan keranjang belanja dan sistem rekomendasi telah merevolusi industri ritel dan hiburan, di antara industri lainnya. Kemampuan mereka untuk

memberikan wawasan mendalam tentang preferensi dan perilaku pelanggan berarti bisnis dapat memenuhi kebutuhan individu secara lebih efektif, mendorong kepuasan dan pendapatan. Seiring dengan terus bertambahnya data dan semakin canggihnya algoritma, alat-alat ini akan memainkan peran yang lebih sentral dalam membentuk lanskap konsumen.

5.7 MEMPREDIKSI PENGURANGAN PELANGGAN DENGAN REGRESI LOGISTIK

Pengurangan pelanggan, juga dikenal sebagai pengurangan pelanggan, mengacu pada fenomena pelanggan meninggalkan layanan atau berhenti menggunakan produk. Di pasar yang kompetitif, memprediksi dan mencegah pengurangan pelanggan merupakan fokus utama bagi banyak bisnis. Mengidentifikasi pelanggan yang kemungkinan akan berhenti berlangganan secara akurat dapat memungkinkan bisnis untuk secara proaktif berinteraksi dengan pelanggan ini dan menerapkan strategi retensi.



Gambar 5.10 Prediksi Churn.

Regresi logistik adalah metode statistik yang umum digunakan untuk prediksi pengurangan pelanggan karena kemampuannya untuk menangani hasil biner, interpretasi, dan efisiensi komputasi (lihat Gambar 5.10). Model ini memodelkan hubungan antara variabel dependen biner (berhenti berlangganan atau tidak berhenti berlangganan) dan satu atau lebih variabel independen (misalnya, demografi pelanggan, pola penggunaan, skor kepuasan), dan menghasilkan probabilitas bahwa variabel dependen tersebut benar (yaitu, pelanggan akan berhenti berlangganan).

Proses tipikal untuk membangun model regresi logistik untuk prediksi berhenti berlangganan mungkin melibatkan langkah-langkah berikut:

1. **Pengumpulan dan pra-pemrosesan data.** Kumpulkan data historis tentang perilaku pelanggan, interaksi, dan peristiwa berhenti berlangganan. Bersihkan data, tangani nilai yang hilang, dan buat fitur yang relevan.

2. **Pelatihan model.** Bagi data menjadi set pelatihan dan set pengujian. Gunakan set pelatihan untuk menyesuaikan model regresi logistik, yaitu, perkiraan parameter model yang memaksimalkan kemungkinan data yang diamati.
3. **Evaluasi model.** Evaluasi kinerja model pada set pengujian menggunakan metrik yang sesuai, seperti akurasi, presisi, recall, atau area di bawah kurva karakteristik operasi penerima (AUC-ROC).
4. **Penerapan dan pemantauan model.** Terapkan model, gunakan untuk memberi skor pelanggan berdasarkan risiko churn mereka, dan pantau kinerjanya dari waktu ke waktu. Perbarui model sesuai kebutuhan ketika data baru tersedia atau ketika konteks bisnis berubah.

Meskipun regresi logistik dapat menjadi alat yang efektif untuk prediksi churn, ia memiliki keterbatasan. Ia mengasumsikan hubungan linier antara logit hasil dan prediktor, dan mungkin tidak berkinerja baik ketika asumsi ini dilanggar atau ketika ada interaksi kompleks antara prediktor.

Peramalan Penjualan dengan Model Deret Waktu

Peramalan penjualan adalah operasi bisnis penting yang memandu banyak keputusan strategis dan operasional, termasuk manajemen persediaan, perencanaan anggaran, dan alokasi sumber daya. Prakiraan penjualan yang akurat dapat menghasilkan penghematan biaya, peningkatan kepuasan pelanggan, dan kinerja bisnis secara keseluruhan yang lebih baik.

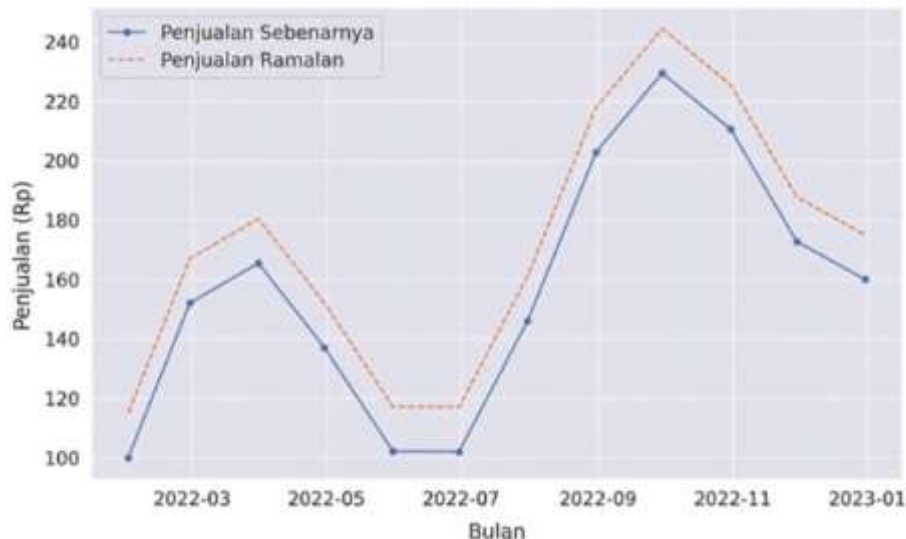
Model deret waktu umumnya digunakan untuk peramalan penjualan karena kemampuannya untuk menangkap pola dan tren temporal dalam data penjualan. Model-model ini membahas data historis untuk mengidentifikasi pola seperti musiman (pola berulang selama periode tetap), tren (pergerakan naik atau turun dari waktu ke waktu), dan siklus (fluktuasi di sekitar tren jangka panjang), dan menggunakan pola-pola ini untuk meramalkan penjualan di masa mendatang.

Salah satu model deret waktu yang paling umum digunakan dalam peramalan penjualan adalah model ARIMA (*autoregressive integrated moving average*). Model ARIMA dapat menangkap serangkaian struktur temporal yang berbeda, menjadikannya alat yang serbaguna untuk berbagai tugas peramalan. Model ini bekerja dengan menggunakan nilai masa lalu (autoregresi), perbedaan antara nilai masa lalu (integrasi), dan kesalahan peramalan masa lalu (rata-rata bergerak) untuk memprediksi penjualan di masa mendatang. Pendekatan populer lainnya adalah model perataan eksponensial, yang menghasilkan perkiraan dengan menerapkan rata-rata tertimbang dari pengamatan masa lalu, di mana bobotnya menurun secara eksponensial seiring bertambahnya usia pengamatan. Varian dari model ini, seperti metode Holt-Winters, juga dapat memperhitungkan tren dan musiman.

Proses tipikal untuk membangun model deret waktu untuk peramalan penjualan mungkin mencakup langkah-langkah berikut (lihat Gambar 5.11):

1. **Pengumpulan dan pra-pemrosesan data.** Kumpulkan data penjualan historis, tangani nilai yang hilang, dan mungkin transformasikan data (misalnya, transformasi log) untuk menstabilkan varians.

2. **Identifikasi model.** Pembahasan data, identifikasi potensi tren dan musiman, dan pilih model yang sesuai (misalnya, ARIMA, perataan eksponensial).
3. **Penyesuaian dan diagnostik model.** Perkiraan parameter model dan periksa kesesuaian model dan asumsi menggunakan plot dan tes diagnostik.
4. . Menghasilkan prakiraan penjualan di masa mendatang dan mengevaluasi kinerja prediktif model menggunakan teknik validasi di luar sampel dan ukuran akurasi yang sesuai (misalnya, kesalahan persentase absolut rata-rata, kesalahan kuadrat rata-rata).



Gambar 5.11 Peramalan Penjualan Menggunakan Pembahasan Deret Waktu.

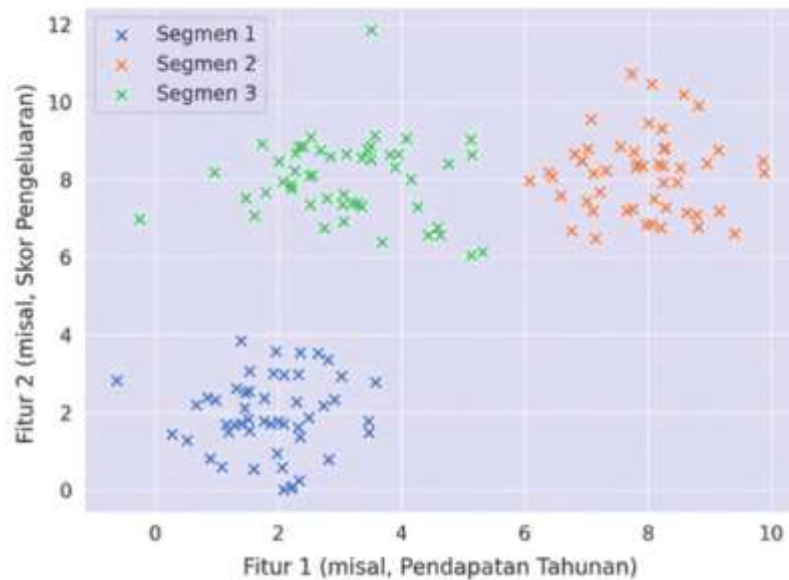
Meskipun model deret waktu dapat memberikan perkiraan yang akurat, model ini memiliki keterbatasan. Model ini mengasumsikan bahwa proses yang mendasarinya yang menghasilkan data tetap stabil dari waktu ke waktu, yang mungkin tidak berlaku dalam lingkungan pasar yang berubah dengan cepat. Selain itu, model-model ini umumnya tidak memasukkan faktor eksternal yang dapat memengaruhi penjualan, seperti indikator ekonomi atau aktivitas pemasaran.

Segmentasi Pelanggan dengan Pengelompokan

Dalam bidang pemasaran, memahami perilaku pelanggan sangat penting untuk membuat keputusan yang tepat. Salah satu cara untuk mendapatkan pemahaman tersebut adalah melalui segmentasi pelanggan, yang melibatkan pengelompokan pelanggan berdasarkan karakteristik yang sama. Karakteristik ini dapat berupa demografis (seperti usia, pendapatan, atau lokasi), perilaku (seperti pola pembelian atau penggunaan produk), psikografis (seperti gaya hidup atau sifat kepribadian), atau kombinasi dari semuanya.

Pengelompokan adalah bentuk pembelajaran mesin tanpa pengawasan yang sering digunakan untuk segmentasi pelanggan. Algoritma ini mengelompokkan titik data (dalam hal ini, pelanggan) sehingga titik-titik dalam kelompok (atau kluster) yang sama lebih mirip satu sama lain daripada dengan titik-titik di kluster lain. Kesamaan ini biasanya diukur berdasarkan jarak dalam ruang multidimensi.

Algoritma pengelompokan yang populer meliputi K-means, pengelompokan hierarkis, dan DBSCAN (pengelompokan spasial berbasis kepadatan aplikasi dengan noise). K-means dan pengelompokan hierarkis bekerja dengan baik pada kluster berbentuk bola dan sensitif terhadap skala variabel, sedangkan DBSCAN dapat mendeteksi kluster dengan bentuk sembarang dan kurang sensitif terhadap skala variabel.



Gambar 5.12 Segmentasi Pelanggan Melalui Pengelompokan.

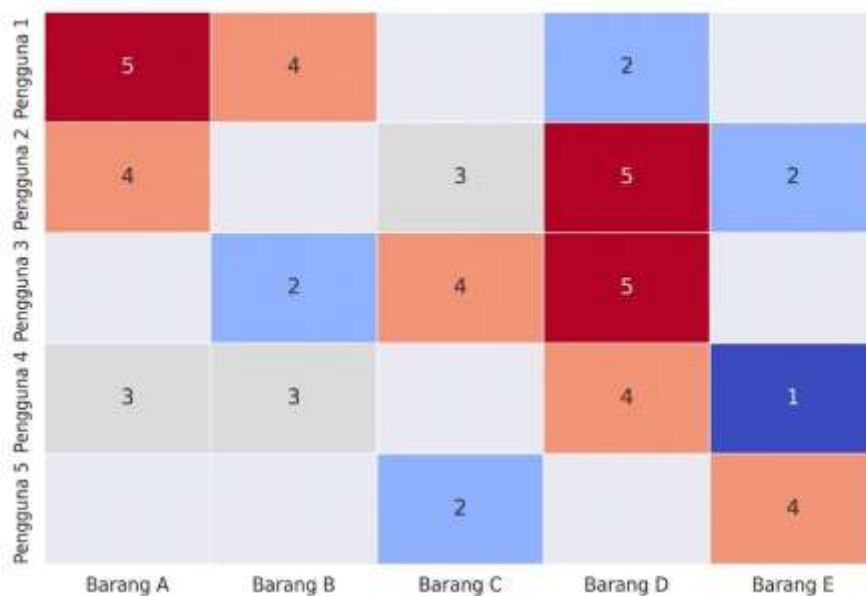
Proses segmentasi pelanggan menggunakan pengelompokan mungkin melibatkan langkah-langkah berikut (lihat Gambar 5.12):

1. **Pengumpulan dan pra-pemrosesan data.** Kumpulkan data tentang pelanggan, bersihkan dan pra-proses data, standarisasi variabel jika perlu, dan buat fitur yang relevan.
2. **Memilih metode pengelompokan.** Pilih metode pengelompokan berdasarkan data dan konteks bisnis. Misalnya, jika jumlah kluster diketahui sebelumnya, K-means mungkin merupakan pilihan yang baik.
3. **Pelatihan model.** Terapkan metode pengelompokan yang dipilih ke data untuk membentuk kluster.
4. **Interpretasi dan aplikasi.** Interpretasikan kluster dengan membahas karakteristik pelanggan dalam setiap kluster. Gunakan hasilnya untuk memandu strategi pemasaran, seperti pemasaran personalisasi, promosi yang ditargetkan, dan pengembangan produk.

Meskipun pengelompokan dapat menjadi alat yang ampuh untuk segmentasi pelanggan, ia memiliki keterbatasan. Kualitas hasil pengelompokan sangat bergantung pada pilihan algoritma pengelompokan dan parameternya, representasi fitur, dan ukuran jarak. Selain itu, pengelompokan tidak memberikan label eksplisit untuk kluster, sehingga menginterpretasikan kluster dan mendapatkan wawasan yang dapat ditindaklanjuti memerlukan pengetahuan domain dan pembahasan yang cermat.

Rekomendasi yang Dipersonalisasi dengan Penyaringan Kolaboratif

Di era kelebihan informasi, sistem rekomendasi yang dipersonalisasi telah menjadi alat penting bagi banyak bisnis untuk meningkatkan keterlibatan pelanggan, kepuasan, dan tingkat konversi. Sistem ini menyarankan item yang mungkin diminati pengguna, berdasarkan perilaku masa lalu mereka atau perilaku pengguna serupa lainnya. Sistem seperti itu banyak digunakan di berbagai domain, termasuk e-commerce (Amazon), konten multimedia (Netflix, Spotify), dan media sosial (Facebook, LinkedIn).



Gambar 5.13 Matriks Peringkat Item Pengguna untuk Pemfilteran Kolaboratif.

Penyaringan kolaboratif adalah salah satu teknik paling populer yang digunakan dalam sistem rekomendasi. Teknik ini memprediksi minat pengguna dengan mengumpulkan preferensi dari banyak pengguna. Asumsi yang mendasarinya adalah jika dua pengguna setuju pada satu masalah, mereka cenderung setuju pada masalah lain juga.

Ada dua jenis utama penyaringan kolaboratif: berbasis pengguna dan berbasis item. Penyaringan kolaboratif berbasis pengguna menemukan pengguna yang mirip dengan pengguna target dan merekomendasikan item yang disukai pengguna serupa tersebut di masa lalu. Namun, penyaringan kolaboratif berbasis item menentukan kesamaan antar item berdasarkan peringkat yang diterima dari pengguna. Untuk pengguna tertentu, ia merekomendasikan item yang mirip dengan item yang telah diberi peringkat tinggi oleh pengguna tersebut di masa lalu.

Proses pembuatan rekomendasi yang dipersonalisasi menggunakan pemfilteran kolaboratif mungkin melibatkan langkah-langkah berikut (lihat Gambar 5.13):

1. **Pengumpulan dan pra-pemrosesan data.** Kumpulkan data interaksi item pengguna, pra-proses data untuk menangani nilai yang hilang, dan ubah data menjadi matriks item pengguna.

2. **Perhitungan kesamaan.** Hitung kesamaan antara pengguna atau item. Ukuran kesamaan yang umum digunakan meliputi kesamaan kosinus, koefisien korelasi Pearson, dan koefisien kesamaan Jaccard.
3. **Pembuatan rekomendasi.** Untuk pengguna target, identifikasi pengguna atau item yang serupa, hitung peringkat prediksi untuk item yang belum dilihat oleh pengguna, dan rekomendasikan item dengan peringkat prediksi tertinggi.

Meskipun pemfilteran kolaboratif dapat memberikan rekomendasi yang dipersonalisasi, ia memiliki tantangannya sendiri. Misalnya, ia mengalami masalah cold start, yaitu, ia tidak dapat membuat rekomendasi untuk pengguna baru atau item baru yang memiliki sedikit data interaksi. Selain itu, cenderung merekomendasikan item populer dan mungkin tidak menangkap item ekor panjang yang kurang dikenal orang tetapi mungkin menarik bagi sebagian pengguna.

Penetapan Harga Dinamis dengan Pembelajaran Penguatan

Dalam lingkungan bisnis yang serba cepat dan kompetitif saat ini, menetapkan harga yang tepat untuk suatu produk atau layanan dapat menjadi tantangan. Strategi penetapan harga statis tradisional seringkali tidak memadai karena gagal memperhitungkan sifat dinamis dari permintaan pasar, penawaran, dan persaingan. Penetapan harga dinamis, yang melibatkan penyesuaian harga berdasarkan kondisi pasar saat ini, telah muncul sebagai strategi yang ampuh untuk memaksimalkan pendapatan atau keuntungan. Ini banyak digunakan di berbagai industri, seperti maskapai penerbangan, perhotelan, berbagi tumpangan, dan e-commerce. RL, cabang dari pembelajaran mesin, menawarkan pendekatan yang menjanjikan untuk penetapan harga dinamis. Tidak seperti pembelajaran terawasi, yang membutuhkan data pelatihan berlabel, RL belajar dari interaksi dengan lingkungan untuk membuat serangkaian keputusan yang memaksimalkan imbalan jangka panjang. Hal ini selaras dengan masalah penetapan harga dinamis, di mana tujuannya adalah untuk menetapkan harga dari waktu ke waktu untuk memaksimalkan total pendapatan atau keuntungan.

Proses implementasi penetapan harga dinamis dengan RL mungkin melibatkan langkah-langkah berikut:

1. **Perumusan masalah.** Rumuskan masalah penetapan harga dinamis sebagai proses keputusan Markov. Definisikan keadaan (misalnya, tingkat persediaan saat ini, waktu dalam sehari, hari dalam seminggu), tindakan (misalnya, tingkat harga), dan fungsi imbalan (misalnya, pendapatan atau keuntungan).
2. **Pemilihan model.** Pilih algoritma RL berdasarkan kompleksitas masalah dan data yang tersedia. Algoritma RL yang umum digunakan meliputi Q-learning, deep Q-network, dan optimasi kebijakan proksimal.
3. **Pelatihan model.** Latih model RL menggunakan data penjualan historis, simulasikan interaksi dengan lingkungan, atau kombinasi keduanya. Model mempelajari kebijakan yang menentukan tindakan apa yang harus diambil (yaitu, harga apa yang harus ditetapkan) berdasarkan keadaan saat ini.

4. **Implementasi kebijakan.** Terapkan kebijakan yang dipelajari dalam sistem dunia nyata. Pantau terus kinerja dan latih ulang model sesuai kebutuhan untuk beradaptasi dengan perubahan kondisi pasar.

Terlepas dari potensinya, penetapan harga dinamis berbasis RL memiliki tantangan tersendiri. Sebagai contoh, dibutuhkan data yang cukup dan berkualitas untuk pelatihan. Model RL mungkin berkinerja buruk jika datanya bising, jarang, atau tidak stasioner. Selain itu, RL biasanya melibatkan proses pembelajaran coba-coba, yang dapat menyebabkan keputusan yang kurang optimal selama fase pembelajaran.

5.8 PEMBELAJARAN MESIN DALAM PEMASARAN

Pembahasan prediktif dan pembelajaran mesin mewakili teknologi transformatif dalam domain pemasaran, menjembatani aspek historis dan masa depan operasi bisnis. Bab ini menjelaskan bagaimana metodologi ini memberdayakan bisnis untuk melangkah lebih jauh dari sekadar memahami apa yang telah terjadi dan mulai memprediksi kejadian di masa depan. Kita memulai perjalanan kita dengan menguraikan konsep pembahasan prediktif, menunjukkan kehebatannya dalam membahas data untuk meramalkan masa depan. Berdasarkan prinsip-prinsipnya, pembelajaran mesin memperkuat kemampuan ini dengan memfasilitasi komputer untuk belajar dari data tanpa pemrograman eksplisit, menawarkan algoritma yang penting untuk prediksi.

Ketika diterapkan dalam pemasaran, metode ini berpotensi untuk menyempurnakan pengalaman pelanggan, mengoptimalkan strategi pemasaran, dan memprediksi perilaku pelanggan dengan akurasi yang signifikan. Kita telah membahas secara mendalam berbagai teknik dalam pembahasan prediktif dan pembelajaran mesin. Ini berkisar dari teknik dasar seperti regresi linier dan logistik, yang sering menjadi titik awal banyak pembahasan pemasaran, hingga seluk-beluk peramalan deret waktu, yang sangat penting di bidang seperti proyeksi penjualan. Saat kami menjelajahi ranah pembelajaran mesin, perbedaan antara pembelajaran terawasi, tidak terawasi, dan penguatan menjadi jelas. Masing-masing memiliki tujuan yang berbeda, baik itu memprediksi hasil tertentu, mendeteksi pola dalam data, atau mempelajari urutan keputusan optimal dari lingkungan interaktif.

Lebih lanjut, melalui contoh praktis, penerapan teknik-teknik ini dalam skenario pemasaran dunia nyata dicontohkan. Kemampuan untuk memprediksi pelanggan yang akan berhenti berlangganan menggunakan regresi logistik, memperkirakan penjualan melalui model deret waktu, melakukan segmentasi pelanggan melalui pengelompokan, dan bahkan adaptasi harga secara dinamis menggunakan pembelajaran penguatan, menunjukkan penerapan yang luas dan potensi transformatif dari pembelajaran prediktif dan pembelajaran mesin. Namun, seperti semua alat dan teknik, pembelajaran prediktif dan pembelajaran mesin memiliki tantangan tersendiri. Memastikan kualitas data, memahami asumsi yang mendasarinya, mengatasi masalah cold-start dalam sistem rekomendasi, dan menavigasi kompleksitas lingkungan dinamis dalam pembelajaran penguatan dapat menghadirkan hambatan. Namun, dengan pemahaman yang mendalam dan penerapan yang cermat, mereka

menawarkan wawasan yang sangat berharga, menjadikannya tak tergantikan dalam perangkat kerja pemasar modern.

Sebagai penutup, seiring berkembangnya lanskap pemasaran, memiliki pengetahuan tentang pembelajaran prediktif dan pembelajaran mesin bukan hanya bermanfaat tetapi juga sangat penting. Mereka menawarkan keunggulan kompetitif, memungkinkan bisnis untuk bersikap proaktif daripada reaktif, dan memposisikan diri secara strategis di pasar yang dinamis. Saat kita beralih ke bab-bab selanjutnya, kita akan mengeksplorasi aspek-aspek yang lebih rumit dari interaksi antara data, pemasaran, dan prediksi, melanjutkan perjalanan kita menuju penguasaan ilmu data pemasaran.

LATIHAN 5.1:

MODEL PREDIKSI CHURN

Tujuan:

Menggunakan data **churn_data** untuk melatih model regresi logistik yang memprediksi churn pelanggan.

Tugas:

1. Pisahkan dataset "**churn_data.csv**" menjadi set pelatihan dan validasi.
2. Latih model regresi logistik untuk memprediksi variabel dependen biner churn.
3. Buat prediksi dan evaluasi model.

Langkah-langkah:

1. Mengimpor Pustaka yang Diperlukan:

```
1. import pandas as pd
2. from sklearn.model_selection import train_test_split
3. from sklearn.linear_model import LogisticRegression
4. from sklearn.metrics import classification_report
```

- **Pandas** digunakan untuk manipulasi dan pembahasan data.
- **train_test_split** dari **sklearn.model_selection** adalah utilitas untuk membagi dataset menjadi set pelatihan dan set pengujian.
- **LogisticRegression** dari **sklearn.linear_model** adalah model pembelajaran mesin untuk tugas klasifikasi.
- **classification_report** dari **sklearn.metrics** menyediakan cara untuk mengevaluasi kualitas prediksi dari algoritma klasifikasi.

2. Memuat Dataset:

```
5. churn_data = pd.read_csv('/data/churn_data.csv')
```

- Kami memuat dataset churn dari file CSV ke dalam DataFrame pandas. Dataset tersebut berisi fitur yang menggambarkan perilaku pelanggan dan variabel target yang menunjukkan apakah pelanggan telah berhenti berlangganan atau tidak.

3. Mendefinisikan Fitur dan Target:

```
6. X = churn_data.drop('churn', axis=1)
7. y = churn_data['churn']
```

- Kita memisahkan fitur (X) dan target (Y). Fitur mencakup semua kolom kecuali kolom target 'churn', yang ingin kita prediksi. Targetnya adalah kolom 'churn', yang akan dipelajari model kita untuk diprediksi.

4. Memisahkan Data:

```
8. X_train, X_test, y_train, y_test = train_test_split(X,y, test_size=0.2,  
    random_state=42)
```

- Dataset dibagi menjadi set pelatihan (80%) dan set pengujian (20%) menggunakan `'train_test_split'`. Parameter `'test_size'` menentukan proporsi dataset yang akan dimasukkan dalam pembagian pengujian. Parameter `'random_state'` memastikan bahwa pembagian tersebut dapat direproduksi; seed acak yang sama berarti pembagian akan sama setiap kali kode dijalankan.

5. Menginisialisasi Model 'LogisticRegression':

```
9. logreg = LogisticRegression()
```

- Sebuah contoh model 'LogisticRegression' dibuat. 'LogisticRegression' dipilih karena merupakan model umum untuk tugas klasifikasi biner, mirip dengan memprediksi pelanggan berhenti berlangganan (ya atau tidak).

6. Melatih Model:

```
10. logreg.fit(X_train, y_train)
```

- Model 'LogisticRegression' dilatih pada data pelatihan (**X_train** dan **y_train**). Metode `fit` menyesuaikan bobot model untuk menemukan batas linier terbaik yang memisahkan kelas.

7. Membuat Prediksi:

```
11. y_pred = logreg.predict(X_test)
```

- Model yang telah dilatih digunakan untuk membuat prediksi pada data uji (**X_test**). Metode prediksi menerapkan bobot yang dipelajari selama pelatihan ke data uji untuk memprediksi hasil churn.

8. Mengevaluasi Model:

```
12. classification_report_output=classification_report(y_test, y_pred)
```

```
13. print(classification_report_output)
```

- Terakhir, fungsi **classification_report** digunakan untuk mengevaluasi prediksi. Fungsi ini membandingkan hasil prediksi pelanggan yang berhenti berlangganan (**y_pred**) dengan hasil aktual dari set data uji (**y_test**).
- Laporan ini menyediakan metrik seperti presisi, recall, dan skor F1 yang membantu memahami kinerja model di berbagai kelas (berhenti berlangganan atau tidak berhenti berlangganan).
- Output dicetak untuk memberikan gambaran yang jelas tentang kinerja model.

Seluruh proses ini merupakan alur kerja dasar untuk melatih dan mengevaluasi model klasifikasi biner dalam pembelajaran mesin. Setiap langkah sangat penting untuk memahami bagaimana model dibangun dan seberapa baik kinerjanya pada data yang belum pernah dilihat sebelumnya.

LATIHAN 5.2

MEMPREDIKSI PENJUALAN MINGGUAN

Tujuan

Membangun model regresi linier untuk memprediksi penjualan mingguan berdasarkan pengeluaran pemasaran dan fitur toko lainnya.

Tugas:

1. Pisahkan dataset “**regression_data.csv**” menjadi set pelatihan dan validasi.
2. Latih model regresi linier untuk memprediksi variabel dependen **weekly_sales**.
3. Buat prediksi dan evaluasi model.

Langkah-langkah:

1. Mengimpor Pustaka:

```
1. import pandas as pd
2. from sklearn.model_selection import train_test_split
3. from sklearn.linear_model import LinearRegression
4. from sklearn.metrics import mean_squared_error
```

- **Pandas** digunakan untuk manipulasi data.
- ``train_test_split`` akan membantu membagi data menjadi set pelatihan dan pengujian.
- ``LinearRegression`` adalah model yang akan kita gunakan untuk prediksi.
- ``mean_squared_error`` akan digunakan untuk mengevaluasi kinerja model.

2. Memuat Dataset:

```
5. regression_data = pd.read_csv('/data/regression_data.csv')
```

- Data regresi dimuat ke dalam DataFrame pandas dari file CSV.

3. Mendefinisikan Fitur dan Target:

```
6. X = regression_data.drop('weekly_sales', axis=1)
7. y = regression_data['weekly_sales']
```

- **X** berisi variabel independen (fitur), yang merupakan semua kolom kecuali ‘weekly_sales’.
- **y** adalah variabel dependen (target), yang ingin kita prediksi—dalam hal ini, ‘weekly_sales’.

4. Membagi Data:

```
8. X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=42)
```

- Dataset dibagi menjadi set pelatihan (80%) dan pengujian (20%), dengan **random_state** diatur untuk reproduksibilitas.

5. Menginisialisasi dan Melatih Model:

```
9. linreg = LinearRegression()
10. linreg.fit(X_train, y_train)
```

- Instance model `LinearRegression` dibuat dan kemudian dilatih menggunakan data pelatihan.

6. Membuat Prediksi:

```
11. y_pred = linreg.predict(X_test)
```

- Model tersebut membuat prediksi (**y_pred**) pada data uji (**X_test**).

7. Mengevaluasi Model:

```
12. mse = mean_squared_error(y_test, y_pred)
```

- Kesalahan kuadrat rata-rata (MSE) dihitung antara nilai aktual (**y_test**) dan nilai prediksi (**y_pred**).

BAB 6

PEMROSESAN BAHASA ALAMI DALAM PEMASARAN

6.1 PENGANTAR NLP UNTUK PEMASAR

Sebelum menyelami dunia pemrosesan bahasa alami (NLP) dalam pemasaran, mari kita luangkan waktu sejenak untuk memahami beberapa konsep dasar. Jika Anda sudah familiar dengan NLP dan kecerdasan buatan (AI), silakan lewati bagian ini.

Apa Itu Pemrosesan Bahasa Alami, dan Mengapa Pemasar Harus Peduli?

Bayangkan mengajari robot untuk memahami bahasa manusia bagaimana kita mengobrol, berdebat, bercanda, dan bahkan melakukan pembelian. NLP adalah teknologi yang membantu komputer memahami, menanggapi, dan memproses bahasa kita. Bagi pemasar, ini berarti lebih memahami umpan balik pelanggan, menyesuaikan iklan, atau bahkan memiliki chatbot yang membantu pelanggan 24/7.

Analogi Sederhana: Resep Bahasa

Anggap bahasa sebagai resep. Setiap bahan (kata) memiliki perannya masing-masing, dan ketika dicampur dengan cara yang tepat, mereka menciptakan hidangan lezat (kalimat yang bermakna). NLP seperti koki ahli yang mengenal setiap bahan dengan baik dan bahkan dapat mengubah resep agar sesuai dengan selera yang berbeda (konteks).

Pemrosesan Bahasa Alami dalam Pemasaran Sehari-hari

- **Media sosial.** Pernahkah Anda bertanya-tanya bagaimana beberapa alat dapat mengetahui apakah tweet tentang suatu produk bersifat positif atau negatif? Itulah NLP yang sedang beraksi, membahas sentimen.
- **Chatbot.** Ketika Anda mengunjungi situs web dan jendela obrolan muncul menawarkan bantuan, NLP mendukung interaksi tersebut, membantu bot memahami dan menanggapi pertanyaan Anda.
- **Asisten suara.** Perangkat seperti Alexa dari Amazon atau Google Home menggunakan NLP untuk memahami dan menjalankan perintah suara, mengubah cara kita mencari produk atau informasi.

Mari Kita Selami Lebih Dalam

Sekarang setelah kita mengupas permukaannya, mari kita menyelami lautan NLP dan menjelajahi bagaimana NLP membentuk kembali dunia pemasaran. Dari memahami komponen-komponennya hingga melihatnya beraksi, bab ini akan memberikan pengetahuan dasar dan wawasan praktis.

6.2 GAMBARAN UMUM PEMROSESAN BAHASA ALAMI

NLP, sebuah disiplin ilmu yang berada di bawah payung AI yang lebih luas, didedikasikan untuk interaksi antara komputer dan bahasa manusia. Tujuan utama NLP adalah untuk memungkinkan komputer memahami, menafsirkan, dan menghasilkan bahasa manusia dengan cara yang bermakna dan berharga. Ini melibatkan pengajaran mesin untuk memahami kompleksitas bahasa manusia, termasuk semantik, sintaksis, dan konteksnya, di antara hal-hal

lainnya. Pada tahun-tahun awalnya, NLP sangat berbasis aturan, dan para ahli bahasa secara manual menulis serangkaian aturan yang kompleks untuk pemrosesan bahasa. Namun, dengan munculnya pembelajaran mesin dan terutama pembelajaran mendalam, pendekatannya bergeser ke arah metode statistik dan berbasis data. Metode-metode ini bergantung pada sejumlah besar data bahasa, atau korpus, dan belajar memproses bahasa dengan mengidentifikasi pola dalam data.

NLP memiliki berbagai macam aplikasi, termasuk terjemahan mesin, pengenalan ucapan, pembahasan sentimen, ekstraksi informasi, dan ringkasan teks, untuk menyebutkan beberapa saja. Dalam konteks pemasaran, NLP dapat digunakan untuk membahas sentimen pelanggan, mempersonalisasi konten iklan, mengotomatiskan layanan pelanggan, dan mendapatkan wawasan dari sejumlah besar data teks tidak terstruktur. Dengan memanfaatkan kekuatan NLP, pemasar dapat lebih memahami pelanggan mereka, meningkatkan strategi pemasaran mereka, dan pada akhirnya mendorong pertumbuhan bisnis.

Pentingnya Pemrosesan Bahasa Alami dalam Pemasaran

Pertumbuhan pesat platform digital telah mengakibatkan peningkatan eksponensial dalam data teks tidak terstruktur, seperti ulasan pelanggan, komentar media sosial, dan diskusi online. Diperkirakan sekitar 80% data dunia tidak terstruktur, dan sebagian besar di antaranya adalah data teks. Namun, metode pembahasan data tradisional kurang cocok untuk menangani jenis data ini. Di sinilah NLP berperan. NLP memungkinkan bisnis untuk membahas volume besar data teks tidak terstruktur, memperoleh wawasan yang bermakna, dan membuat keputusan pemasaran berbasis data. Misalnya, NLP dapat digunakan untuk membahas sentimen pelanggan dari ulasan online atau unggahan media sosial, membantu bisnis untuk memahami bagaimana pelanggan mereka memandang produk, layanan, atau merek mereka. Informasi ini dapat sangat berharga untuk memandu strategi pemasaran dan meningkatkan kepuasan pelanggan.

Selain itu, NLP dapat meningkatkan efektivitas komunikasi pemasaran. Dengan membahas bahasa yang digunakan oleh pelanggan, NLP dapat membantu bisnis untuk menyesuaikan pesan pemasaran mereka dengan preferensi dan sentimen pelanggan individu, sehingga menghasilkan komunikasi pemasaran yang lebih personal dan menarik. NLP juga dapat meningkatkan layanan pelanggan, aspek penting lainnya dalam pemasaran. Misalnya, chatbot yang didukung NLP dapat menyediakan layanan pelanggan instan 24/7, menjawab pertanyaan pelanggan, dan memberikan rekomendasi produk dalam bahasa alami. Hal ini tidak hanya meningkatkan pengalaman pelanggan tetapi juga mengurangi beban pada perwakilan layanan pelanggan dan memangkas biaya.

Singkatnya, NLP memainkan peran penting dalam pemasaran modern, membantu bisnis untuk lebih memahami pelanggan mereka, meningkatkan komunikasi pemasaran mereka, dan meningkatkan layanan pelanggan. Dengan memanfaatkan kekuatan NLP, bisnis dapat memperoleh keunggulan kompetitif dalam lanskap bisnis yang semakin digital dan berbasis data (lihat Tabel 6.1).

Tabel 6.1 Berbagai Masalah Pemasaran dan Bagaimana Pemrosesan Bahasa Alami Memberikan Solusi.

Tantangan Pemasaran	Solusi NLP
Memahami sentimen pelanggan	Pembahasan sentimen
Mengotomatiskan dukungan pelanggan	Chatbot dan asisten virtual
Periklanan yang ditargetkan	Optimasi konten iklan menggunakan NLP
Rekomendasi konten	Penyaringan berbasis konten menggunakan NLP
Pemantauan merek di media sosial	Pengambilan data media sosial dan pembahasan sentimen

Komponen Pemrosesan Bahasa Alami: Sintaksis, Semantik, dan Pragmatik

Sintaksis

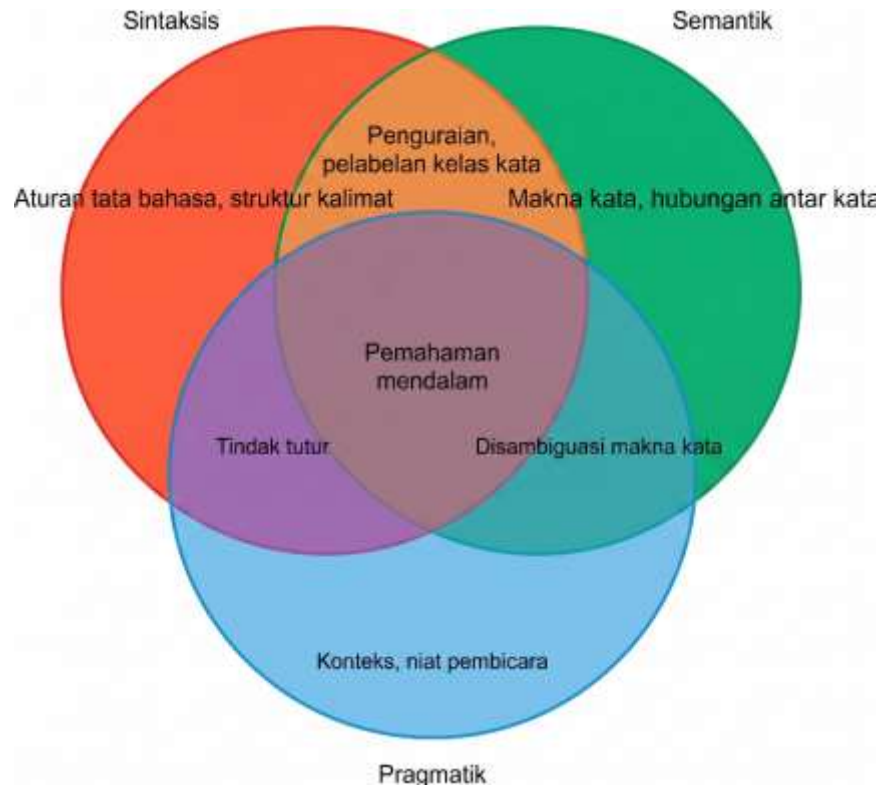
Pada intinya, sintaksis mengacu pada aturan yang menentukan struktur kalimat dalam suatu bahasa. Dengan kata lain, sintaksis berkaitan dengan bagaimana kata-kata bergabung untuk membentuk frasa, klausa, dan kalimat. Untuk NLP, pembahasan sintaksis sering melibatkan tugas-tugas seperti parsing (memecah kalimat menjadi bagian-bagian penyusunnya dan memahami hubungannya) dan penandaan bagian ucapan (mengidentifikasi kategori tata bahasa dari kata-kata individual, seperti kata benda, kata kerja, kata sifat, dan lain-lain.). Memahami sintaksis sangat penting karena bahkan sedikit perubahan dalam urutan kata atau struktur dapat secara drastis mengubah makna kalimat.

Semantik

Melampaui struktur, semantik menyelami makna. Semantik berkaitan dengan interpretasi tanda atau simbol dalam sistem komunikasi, baik itu kata-kata, tanda, simbol, atau isyarat. Dalam NLP, pembahasan semantik digunakan untuk memahami makna kata-kata individual dalam konteks, menyelesaikan ambiguitas (misalnya, menentukan arti kata bank dalam konteks tepi sungai versus bank tabungan), dan mengekstrak informasi terstruktur dari teks yang tidak terstruktur. Ontologi dan grafik pengetahuan, yang menangkap informasi terstruktur tentang dunia dan hubungan antar entitas, memainkan peran penting dalam pemahaman semantik.

Pragmatik

Pragmatik menyelidiki konteks di mana komunikasi terjadi, menjawab pertanyaan-pertanyaan seperti: Siapa yang berbicara? Kepada siapa? Dalam keadaan apa? Dan dengan maksud apa? Ini tentang memahami bahasa dalam konteks, menangkap makna tersirat, memahami komunikasi tidak langsung, dan memahami norma dan aturan sosial yang memandu komunikasi. Dalam ranah NLP, pragmatik dapat membantu dalam tugas-tugas seperti pembahasan sentimen (di mana kata yang sama dapat memiliki konotasi yang berbeda berdasarkan konteks) atau sistem dialog (di mana pemahaman maksud pengguna dan konteks percakapan sangat penting) (lihat Gambar 6.1).



Gambar 6.1 Tumpang Tindih dan Perbedaan Antara Sintaksis, Semantik, dan Pragmatik dalam Pemrosesan Bahasa Alami.

Tantangan dalam Pemrosesan Bahasa Alami

Ambiguitas

Tantangan utama dalam NLP adalah ambiguitas, yang muncul ketika sebuah kata atau kalimat dapat memiliki banyak interpretasi. Misalnya, "Saya melihatnya menghindari" dapat berarti mengamati burung atau seseorang yang menghindari. Menyelesaikan ambiguitas semacam itu membutuhkan pemahaman kontekstual, yang dapat menjadi kompleks bagi sistem komputasi.

Idiom dan Frasa Kata Kerja

Bahasa dipenuhi dengan ungkapan idiomatik dan frasa kata kerja yang tidak dapat diterjemahkan secara langsung kata demi kata. Misalnya, "menendang ember" bukan tentang menendang atau ember tetapi menunjukkan kematian seseorang. Mengenali dan menafsirkan ungkapan semacam itu merupakan tantangan bagi sistem NLP.

Sarkasme dan Ironi

Manusia sering menggunakan bahasa dengan cara yang berarti kebalikan dari interpretasi literal, seperti sarkasme atau ironi. Mendeteksi dan memahami nuansa tersebut dalam teks tertulis atau lisan dapat menjadi tantangan yang cukup besar.

Nuansa Budaya dan Kontekstual

Bahasa berkembang dalam konteks budaya. Kata, frasa, atau konstruksi mungkin memiliki konotasi budaya spesifik atau referensi historis yang mungkin tidak jelas bagi orang luar. Memastikan sistem NLP mengenali dan memahami nuansa ini merupakan tantangan tetapi sangat penting untuk interpretasi yang akurat.

Evolusi Bahasa

Bahasa adalah entitas yang hidup dan berevolusi. Kata-kata baru muncul, makna bergeser, dan kata-kata lama menjadi usang. Sistem NLP perlu terus beradaptasi dan belajar agar tetap relevan.

Tantangan Multibahasa

Karena bisnis dan layanan beroperasi secara global, ada kebutuhan akan sistem NLP untuk memahami, menafsirkan, dan menghasilkan berbagai bahasa. Menangani nuansa, idiom, dan struktur lintas bahasa, terutama dalam tugas-tugas seperti penerjemahan, tetap menjadi tantangan yang signifikan.

Kekhawatiran Etis

Karena model NLP dilatih pada sejumlah besar data, terkadang model tersebut dapat mencerminkan dan melanggengkan bias yang ada dalam data. Mengatasi bias ini dan memastikan keadilan dalam aplikasi NLP merupakan tantangan dan tanggung jawab.

Singkatnya, meskipun NLP menawarkan potensi transformatif untuk berbagai aplikasi, NLP juga memiliki tantangannya sendiri. Mengatasi tantangan ini membutuhkan perpaduan keahlian linguistik, teknik komputasi tingkat lanjut, dan pemahaman mendalam tentang sifat bahasa manusia yang terus berkembang.

6.3 TEKNIK NLP DALAM PEMASARAN

Di lautan luas data pemasaran, informasi tekstual merupakan tambang emas wawasan. Namun, sebelum pembahasan yang bermakna dapat dilakukan, teks mentah ini harus melalui serangkaian transformasi agar dapat dipahami dan ditindaklanjuti. Bagian ini membahas teknik inti pra-pemrosesan teks dan ekstraksi fitur, langkah-langkah penting dalam mengubah data tekstual yang tidak terstruktur menjadi format terstruktur yang sesuai untuk pembahasan tingkat lanjut. Dari tokenisasi dan stemming hingga seluk-beluk vektorisasi dan penyematan kata, kita akan mengeksplorasi bagaimana teknik-teknik ini meletakkan dasar untuk aplikasi NLP yang efektif dalam pemasaran, memastikan bahwa wawasan yang diperoleh akurat dan relevan.

Tokenisasi dan Stemming

Tokenisasi adalah proses mendasar untuk mengubah teks menjadi token, yang merupakan potongan atau kata yang lebih kecil. Proses ini membantu menyederhanakan jenis penguraian selanjutnya dan memungkinkan pemetaan makna yang lebih mudah dari bahasa manusia. Namun, stemming adalah proses mengurangi kata-kata yang diinfleksikan (atau terkadang diturunkan) menjadi bentuk dasar atau akar kata umumnya bentuk kata tertulis. Misalnya, akar kata jumps mungkin adalah jump. Stemming banyak digunakan dalam mesin pencari dan sistem pengambilan informasi untuk memastikan berbagai bentuk kata cocok selama pencarian, seperti ketika pengguna mencari marketing, hasil dengan marketed atau marketer juga muncul.

Penghapusan Kata Hentian

Kata hentian adalah kata-kata yang umum digunakan dalam suatu bahasa seperti and, the, is, dan sebagainya, yang mungkin tidak mengandung informasi penting saat membahas

data teks. Dalam NLP dan penambangan teks, kata-kata ini sering disaring sebelum atau setelah proses pemrosesan teks. Alasan di balik penghapusan kata-kata ini adalah untuk fokus pada kata-kata yang bermakna, yang dapat meningkatkan efisiensi proses selanjutnya seperti klasifikasi teks atau pembahasan sentimen.

Vektorisasi: Bag of Words dan TF-IDF

Setelah data teks di-tokenisasi dan dibersihkan, data tersebut dikonversi ke format yang mudah dipahami oleh algoritma pembelajaran mesin. Salah satu metode tersebut adalah model *bag of words* (BoW). BoW merepresentasikan setiap dokumen atau teks sebagai vektor numerik, di mana keberadaan atau ketidakhadiran suatu kata ditunjukkan oleh 1 atau 0.

Mari kita pertimbangkan dokumen teks sampel berikut:

1. **Dokumen 0.** "AI dengan pemrosesan bahasa alami mengubah dunia."
2. **Dokumen 1.** "AI dan robotika adalah masa depan teknologi."
3. **Dokumen 2.** "Pemasaran dengan AI dan data adalah kunci kesuksesan."

Dari dokumen sampel ini, kita memperoleh representasi BoW dan TF-IDF (lihat Tabel 6.2).

TF-IDF (frekuensi istilah-frekuensi dokumen terbalik) adalah metode lain yang tidak hanya mempertimbangkan frekuensi suatu kata dalam dokumen tertentu tetapi juga seberapa sering kata tersebut muncul di semua dokumen. Ide di balik ini adalah untuk memberikan bobot yang lebih tinggi pada kata-kata yang unik untuk dokumen tertentu. Metode ini membantu menekan kata-kata yang sering muncul di semua dokumen yang mungkin tidak mengandung banyak informasi.

Tabel 6.2 Representasi BoW.

	AI	dan	adalah	berubah	data	masa depan	kunci	bahasa	pemasaran	...
Dok 0	1	0	1	1	0	0	0	1	0	...
Dok 1	1	1	1	0	0	1	0	0	0	...
Dok 2	1	1	1	0	1	0	1	0	1	...

	alami	pemrosesan	robotika	kesuksesan	teknologi	(the)	ke/untuk	dengan	dunia
Dok 0	1	1	0	0	0	0	0	1	1
Dok 1	0	0	1	0	1	1	0	0	0
Dok 2	0	0	0	1	0	1	1	1	0

Untuk menentukan representasi TF-IDF, kita menghitung frekuensi istilah untuk setiap istilah dalam setiap dokumen dan kemudian menyesuaikannya dengan frekuensi dokumen terbalik untuk istilah tersebut di semua dokumen. Dengan tiga dokumen sampel kita, mari kita uraikan perhitungan TF-IDF untuk dokumen sampel:

Langkah 1. Hitung frekuensi istilah (TF).

Frekuensi istilah (TF) suatu kata adalah frekuensi suatu kata dalam suatu dokumen dibagi dengan jumlah total kata dalam dokumen tersebut. Misalnya, untuk kata 'AI' dalam Dokumen 0: $TF('AI', \text{Dokumen 0}) = 1 / 9 = 0,111$.

Langkah 2. Hitung frekuensi dokumen terbalik (IDF).

Frekuensi dokumen terbalik (IDF) suatu kata adalah logaritma dari jumlah total dokumen dibagi dengan jumlah dokumen yang mengandung kata tersebut. Misalnya, untuk kata 'AI': $IDF('AI') = \log(3 / 3) = 0$ (karena 'AI' muncul di ketiga dokumen).

Langkah 3. Hitung TF-IDF.

TF-IDF adalah hasil perkalian TF dan IDF. Misalnya, untuk kata 'AI' dalam Dokumen 0:
 $TF-IDF('AI', \text{Dokumen 0}) = TF('AI', \text{Dokumen 0}) * IDF('AI') = 0,111 * 0 = 0$.

Dengan menggunakan proses ini, kita dapat menghitung nilai TF-IDF untuk semua kata di semua dokumen. Tabel 6.3 menunjukkan nilai TF-IDF ilustratif. (Catatan: Untuk kesederhanaan, kami telah membulatkan nilai ke tiga desimal dan mempertimbangkan logaritma natural.) Perlu diingat bahwa nilai-nilai ini didasarkan pada perhitungan sederhana untuk tujuan ilustrasi. Dalam implementasi sebenarnya, langkah-langkah pra-pemrosesan dan penyesuaian tambahan mungkin diterapkan.

Tabel 6.3 Representasi Frekuensi Istilah-Frekuensi Dokumen Terbalik (TF-IDF).

	AI	dan	adalah	berubah	data	masa depan	kunci	bahasa	pemasaran	...
Dok 0	0	0	0	0.356	0	0	0	0.356	0	...
Dok 1	0	0	0	0	0	0.356	0	0	0	...
Dok 2	0	0	0	0	0.356	0	0.356	0	0.356	...

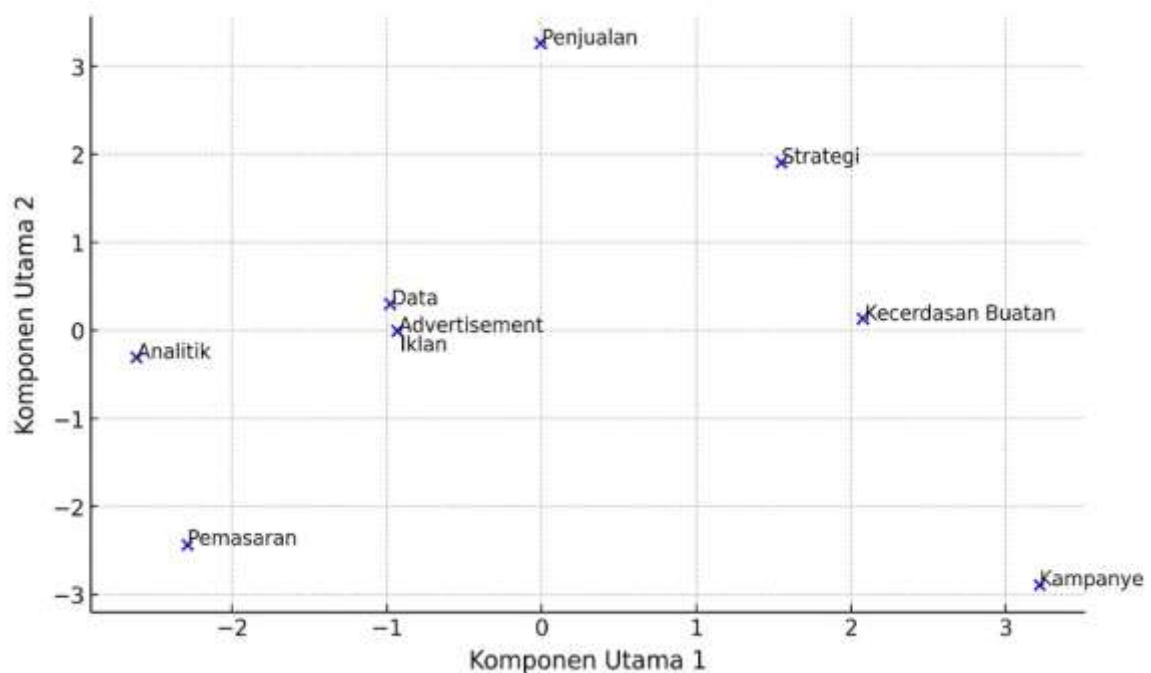
	alami	pemrosesan	robotika	kesuksesan	teknologi	itu	untuk	dengan	dunia
Dok 0	0.356	0.356	0	0	0	0	0	0	0.356
Dok 1	0	0	0.356	0	0.356	0	0	0	0
Dok 2	0	0	0	0.356	0	0	0.356	0	0

Embedding Kata: Word2Vec, GloVe

Embedding kata adalah cara modern untuk merepresentasikan kata sebagai vektor dalam ruang padat di mana semantik kata ditangkap oleh posisi dalam ruang tersebut. Dua model paling terkenal untuk membuat embedding tersebut adalah Word2Vec dan GloVe.

Word2Vec, yang diperkenalkan oleh Mikolov dan rekan-rekan. pada tahun 2013, menangkap makna semantik kata berdasarkan konteksnya dalam data. Terdapat dua algoritma pelatihan utama untuk Word2Vec: *continuous bag of words* (CBOW) dan Skip-Gram.

GloVe (*global vectors for word representation*), yang diperkenalkan oleh Pennington dan rekan-rekan. pada tahun 2014, adalah metode lain untuk mendapatkan representasi vektor untuk kata-kata. Metode ini menggabungkan manfaat dari dua metode utama di bidang ini: metode berbasis hitungan dan metode prediktif. Hal ini dilakukan dengan memfaktorkan matriks kemunculan bersama kata. Word2Vec dan GloVe telah banyak digunakan dalam berbagai aplikasi NLP, termasuk pembahasan sentimen, terjemahan mesin, dan banyak lagi (lihat Gambar 6.2).



Gambar 6.2 Penyematan Kata dalam Ruang 2D.

6.4 TEKNIK PEMROSESAN BAHASA ALAMI UTAMA UNTUK PEMASARAN

Saat kita mempelajari lebih dalam aplikasi praktis NLP di sektor pemasaran, sangat penting untuk memahami teknik inti yang memungkinkan aplikasi ini. Pada bagian ini, kita akan mengeksplorasi beberapa teknik NLP utama ini dan bagaimana teknik tersebut dapat dimanfaatkan untuk mengoptimalkan upaya pemasaran. Teknik-teknik ini meliputi pembahasan sentimen, pemodelan topik, pengenalan entitas bernama (NER), dan klasifikasi teks, yang masing-masing memiliki peran berbeda dalam mengubah data teks tidak terstruktur menjadi wawasan yang dapat ditindaklanjuti.

Dengan evolusi algoritma pembelajaran mesin dan kemajuan berkelanjutan dalam kemampuan komputasi, teknik-teknik ini menjadi lebih efisien dan akurat. Teknik-teknik ini sekarang memberikan kerangka kerja yang kuat bagi pemasar untuk memahami opini pelanggan, mendeteksi tren yang muncul, mempersonalisasi komunikasi, dan banyak lagi. Bagian selanjutnya akan membahas setiap teknik ini, memberikan gambaran komprehensif tentang cara kerjanya, pentingnya, dan aplikasi praktisnya dalam pemasaran.

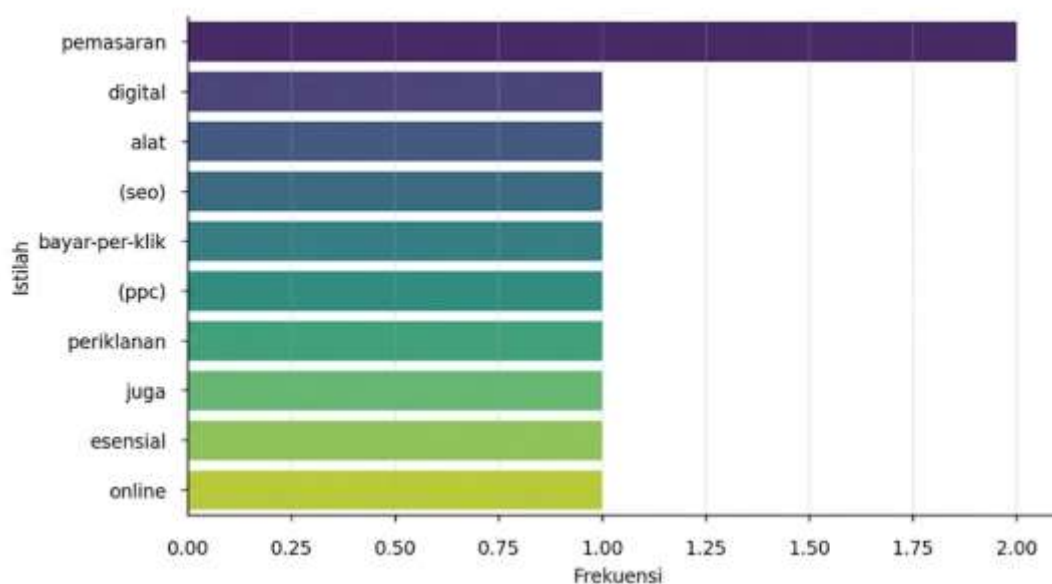
Pembahasan Teks

Pembahasan teks, juga dikenal sebagai penambangan teks, adalah proses untuk mendapatkan informasi berkualitas tinggi dari data teks menggunakan berbagai teknik NLP, pengambilan informasi, dan pembelajaran mesin. Proses ini melibatkan penataan teks masukan, pengambilan pola dalam data terstruktur, dan akhirnya, evaluasi dan interpretasi hasil. Dalam pemasaran, pembahasan teks dapat sangat bermanfaat dalam memahami suara pelanggan, karena memungkinkan pemasar untuk membahas sejumlah besar data teks tidak terstruktur seperti ulasan pelanggan, unggahan media sosial, dan tiket dukungan pelanggan. Dengan demikian, pemasar dapat mengidentifikasi tema umum, mendeteksi sentimen, memahami kebutuhan dan preferensi pelanggan, dan memperoleh wawasan berharga yang

dapat menjadi dasar strategi pemasaran. Sebagai contoh, pembahasan teks dapat digunakan untuk membahas ulasan pelanggan dan mengidentifikasi atribut produk utama yang sering disebutkan pelanggan, dan apakah sentimen terhadap atribut ini umumnya positif atau negatif. Hal ini dapat membantu pemasar untuk memahami kekuatan dan kelemahan suatu produk, sebagaimana yang dirasakan oleh pelanggan, dan melakukan penyesuaian yang diperlukan pada produk atau strategi pemasarannya.

Aplikasi lain dari pembelajaran teks dalam pemasaran adalah dalam pemantauan media sosial. Dengan membahas unggahan media sosial, pemasar dapat mendeteksi tren yang muncul, memantau reputasi online merek, dan mendapatkan wawasan tentang sikap pelanggan terhadap merek atau pesaingnya. Hal ini dapat memandu pengembangan kampanye pemasaran dan membantu pemasar untuk bereaksi cepat terhadap perubahan lingkungan pasar.

Dalam ranah pemantauan media sosial, pembelajaran teks berfungsi sebagai alat yang ampuh untuk mengekstrak wawasan berharga dari sejumlah besar data yang tidak terstruktur. Gambar 6.3 mengilustrasikan aplikasi ini dengan jelas, menyajikan bagan batang yang menampilkan istilah yang paling sering muncul dalam sampel teks media sosial. Visualisasi ini merangkum inti dari pembelajaran teks dalam pemasaran, menunjukkan bagaimana frasa dan kata kunci umum dapat menandakan tren, reputasi merek, dan sentimen pelanggan.



Gambar 6.3 Istilah yang Paling Sering Muncul dalam Contoh Teks Menggunakan Pembahasan Teks.

Singkatnya, pembahasan teks menyediakan alat yang ampuh bagi pemasar untuk memahami sejumlah besar data teks yang tidak terstruktur dan memperoleh wawasan pelanggan yang berharga, sehingga memungkinkan pengambilan keputusan berbasis data dalam pemasaran.

Pembahasan Sentimen

Pembahasan sentimen, terkadang juga disebut sebagai penambangan opini, adalah subbidang NLP yang melibatkan penggunaan pemrosesan bahasa alami, pembahasan teks, dan linguistik komputasional untuk mengidentifikasi, mengekstrak, mengukur, dan mempelajari keadaan afektif dan informasi subjektif dari materi sumber. Dalam konteks pemasaran, hal ini umumnya digunakan untuk memahami sentimen pelanggan terhadap produk, layanan, atau merek berdasarkan ulasan tertulis (atau lisan), umpan balik, atau percakapan media sosial.

Salah satu penggunaan utama pembahasan sentimen dalam pemasaran adalah dalam pemantauan media sosial. Pemasar dapat membahas tweet, unggahan Facebook, atau konten media sosial lainnya untuk mengukur sentimen publik terhadap merek, produk, atau kampanye pemasaran tertentu mereka. Wawasan yang diperoleh dari pembahasan tersebut dapat membantu pemasar untuk merespons tren positif atau negatif dengan cepat, menyesuaikan komunikasi mereka untuk mengatasi kekhawatiran pelanggan atau memanfaatkan umpan balik positif, dan secara umum membuat keputusan berdasarkan data yang dapat meningkatkan persepsi merek dan kepuasan pelanggan.

Namun, pembahasan sentimen bukannya tanpa tantangan. Mendeteksi nuansa dalam bahasa manusia, terutama sarkasme dan ambiguitas, bisa sangat sulit bagi sistem otomatis. Misalnya, pernyataan seperti “Oh bagus, tepat seperti yang saya butuhkan!” bisa benar-benar positif atau negatif secara sarkastik, tergantung pada konteksnya. Mengidentifikasi dan menafsirkan nuansa tersebut secara akurat sangat penting untuk mendapatkan gambaran sentimen yang sebenarnya, dan ini tetap menjadi tantangan yang berkelanjutan dalam bidang pembahasan sentimen.

Untuk mengatasi hal ini, beberapa kemajuan dan solusi sedang dieksplorasi:

- **Pembahasan kontekstual.** Algoritma dirancang untuk mempertimbangkan informasi kontekstual yang lebih luas daripada membahas kalimat secara terisolasi. Ini membantu dalam menentukan suasana hati atau maksud di balik pernyataan berdasarkan teks di sekitarnya.
- **Teknik pembelajaran mendalam.** Jaringan saraf, terutama jaringan saraf berulang (RNN) dan transformer, digunakan untuk memodelkan urutan dalam teks, yang membantu mendeteksi nuansa linguistik dengan lebih baik.
- **Pembahasan emoji dan tanda baca.** Sistem modern menggabungkan emoji dan tanda baca sebagai indikator sentimen, mengingat penggunaannya yang umum dalam mengekspresikan emosi secara online.
- **Human-in-the-loop (HITL).** Beberapa perusahaan mengintegrasikan peninjau manusia ke dalam proses pembahasan sentimen mereka. Peninjau ini membantu melatih model dengan mengoreksi kesalahannya, sehingga meningkatkan akurasi dari waktu ke waktu.

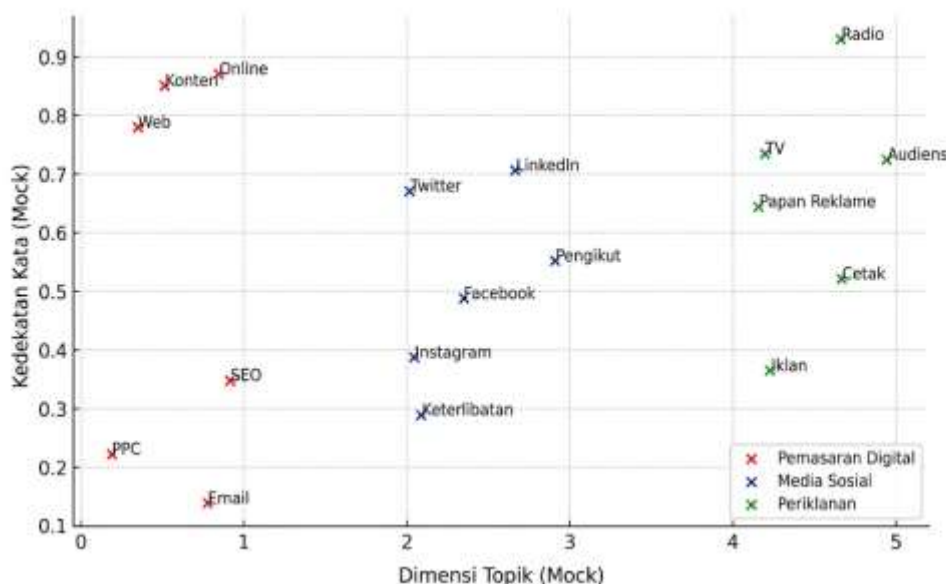
Pembahasan sentimen juga dapat digunakan untuk menilai ulasan dan peringkat pelanggan di platform e-commerce atau situs ulasan. Dengan membahas sentimen yang diungkapkan dalam ulasan tertulis secara otomatis, pemasar dapat memperoleh wawasan tentang pengalaman

pelanggan, kualitas produk atau layanan, dan area potensial untuk perbaikan, seringkali secara real-time. Hal ini dapat membantu perusahaan untuk mengatasi masalah secara proaktif dan meningkatkan penawaran produk atau layanan mereka berdasarkan umpan balik pelanggan.

Selain itu, pembahasan sentimen juga dapat digunakan dalam pembahasan kompetitif. Dengan membahas sentimen terhadap merek atau produk pesaing, perusahaan dapat mengidentifikasi peluang dan ancaman di pasar dan menyusun strategi yang sesuai. Melalui pembahasan sentimen, pemasar memperoleh metode yang kuat untuk memahami dan mengatasi sikap dan opini pelanggan, sehingga meningkatkan kemampuan mereka untuk memenuhi kebutuhan pelanggan dan menjalin hubungan merek yang lebih dalam.

Pemodelan Topik

Pemodelan topik adalah jenis model statistik yang digunakan untuk menemukan topik abstrak yang muncul dalam kumpulan dokumen, seperti ulasan pelanggan atau unggahan media sosial. Alokasi Dirichlet laten adalah salah satu metode yang paling umum digunakan untuk pemodelan topik. Dalam pemasaran, pemodelan topik dapat digunakan untuk secara otomatis mengidentifikasi tema atau topik umum dalam sejumlah besar data teks yang tidak terstruktur. Misalnya, dengan menerapkan pemodelan topik pada ulasan pelanggan, pemasar dapat mengidentifikasi topik-topik utama yang sering dibahas pelanggan, seperti fitur produk, layanan pelanggan, harga, dan sebagainya. Hal ini dapat memberikan wawasan berharga mengenai aspek-aspek produk atau layanan apa yang paling penting bagi pelanggan, yang dapat menjadi dasar pengembangan produk dan strategi pemasaran.



Gambar 6.4 Hasil Pemodelan Topik yang Menunjukkan Klaster Kata untuk Topik yang Berbeda.

Pemodelan topik juga dapat digunakan dalam pembahasan media sosial. Misalnya, dengan menerapkan pemodelan topik pada tweet yang menyebutkan suatu merek, pemasar dapat mengidentifikasi topik utama percakapan seputar merek tersebut. Hal ini dapat membantu

pemasar untuk memahami persepsi merek, mengidentifikasi tren yang muncul, dan memantau dampak kampanye pemasaran.

Selain itu, pemodelan topik dapat digunakan untuk pembahasan konten dalam pemasaran konten. Dengan menerapkan pemodelan topik pada kumpulan posting blog atau artikel, pemasar dapat mengidentifikasi topik utama yang dibahas dan bagaimana topik-topik ini saling terkait. Hal ini dapat menginformasikan pengembangan konten di masa mendatang dan membantu pemasar untuk membuat konten yang sesuai dengan target audiens mereka. Melalui pemodelan topik, pemasar memiliki instrumen yang efektif untuk secara otomatis membedakan topik utama dalam sejumlah besar data teks, menawarkan wawasan berharga yang dapat membentuk strategi pemasaran.

Pengenalan Entitas Bernama

Pengenalan entitas bernama (NER) adalah sub-tugas ekstraksi informasi yang mengidentifikasi dan mengklasifikasikan entitas bernama dalam teks ke dalam kategori yang telah ditentukan sebelumnya seperti nama orang, organisasi, lokasi, kode medis, ekspresi waktu, kuantitas, nilai moneter, persentase, dan sebagainya. Di bidang pemasaran, NER menawarkan banyak aplikasi untuk menyaring wawasan dari teks yang tidak terstruktur. Berikut beberapa aplikasi ilustratif:

1. Pembahasan media sosial:

- **Contoh.** Pertimbangkan sebuah tweet, “Baru saja mencoba Frappuccino baru dari Starbucks, dan rasanya luar biasa! #coffeeLover.” NER dapat mengekstrak entitas ‘Frappuccino’ (produk) dan ‘Starbucks’ (organisasi). Hal ini memungkinkan pemasar untuk memantau dan membahas penyebutan produk tertentu dan memahami sentimen tentangnya.
- Dengan mengekstrak entitas dari media sosial, pemasar dapat membedakan subjek utama seperti merek, produk, atau tokoh. NER membantu dalam pelacakan penyebutan merek, pembahasan sentimen terhadap produk tertentu, atau identifikasi influencer di ranah digital merek.

2. Pembahasan kompetitif:

- **Contoh.** Dari sebuah artikel yang menyatakan, “Apple akan meluncurkan iPhone barunya pada bulan September, membuat Samsung waspada,” NER dapat mengidentifikasi ‘Apple’ dan ‘Samsung’ sebagai organisasi dan ‘iPhone’ sebagai produk. Ekstraksi semacam itu memungkinkan bisnis untuk memantau aktivitas pesaing dan menyusun strategi yang sesuai.
- Melalui NER dalam laporan industri, berita, atau media sosial pesaing, perusahaan dapat memperoleh wawasan tentang pesaing, tren pasar, dan potensi peluang atau ancaman.

3. Pemasaran konten:

- **Contoh.** Dalam sebuah postingan blog berjudul “Destinasi Terbaik Eropa di Musim Panas,” NER dapat secara otomatis menandai entitas seperti ‘Paris’, ‘Roma’, atau ‘Amsterdam’. Ketika pengguna membaca tentang ‘Paris’ dan berinteraksi secara

positif, sistem dapat merekomendasikan konten lain yang terkait dengan 'Paris' untuk dibaca selanjutnya.

- NER membantu secara otomatis mendeteksi dan memberi tag pada entitas dalam konten, meningkatkan kemampuan penemuan konten dan personalisasi. Hal ini memastikan pengguna dapat dengan mudah menemukan konten atau menerima rekomendasi berdasarkan interaksi entitas sebelumnya.

Dengan memanfaatkan NER, pemasar dapat dengan mahir mengekstrak dan membahas informasi penting dari data teks tidak terstruktur yang luas, memperoleh wawasan penting untuk menyempurnakan strategi pemasaran.

Klasifikasi Teks

Klasifikasi teks, juga disebut kategorisasi teks, menetapkan kategori atau label yang telah ditentukan sebelumnya pada dokumen tekstual berdasarkan kontennya. Proses ini mendukung berbagai operasi NLP dan pengambilan informasi seperti deteksi spam, pembahasan sentimen, dan pengorganisasian dokumen. Secara teknologi, klasifikasi teks difasilitasi oleh model pembelajaran mesin yang dilatih pada kumpulan data berlabel. Berikut adalah gambaran singkat tentang algoritma dan teknik yang terlibat:

- **Praproses.** Sebelum klasifikasi, teks harus diproses. Langkah-langkah umum meliputi tokenisasi, stemming atau lemmatisasi, dan penghapusan kata-kata penghenti (stop words). Ini mengubah teks mentah menjadi format terstruktur yang sesuai untuk pembelajaran mesin.
- **Ekstraksi fitur.** Teknik seperti BoW, TF-IDF, dan penyematan kata (seperti Word2Vec atau GloVe) mengubah data tekstual menjadi vektor numerik yang menangkap esensi konten.
- **Pelatihan model.** Berbagai algoritma dapat digunakan untuk klasifikasi teks:
 - **Naive Bayes.** Algoritma probabilistik ini populer karena kesederhanaan dan efisiensinya dalam tugas klasifikasi teks, terutama deteksi spam.
 - **Support vector machines (SVM).** Ini dapat membuat hyperplane yang memisahkan data ke dalam kelas yang berbeda, sehingga cocok untuk data teks berdimensi tinggi.
 - **Jaringan saraf:** Model pembelajaran mendalam, khususnya RNN dan transformer, telah mencapai hasil terbaik dalam klasifikasi teks dengan menangkap ketergantungan sekuensial dalam teks.
- **Evaluasi dan optimasi.** Setelah dilatih, model dievaluasi menggunakan metrik seperti akurasi, presisi, recall, dan skor F1. Kemudian, model dapat dioptimalkan menggunakan teknik seperti penyetelan hyperparameter.

Dalam bidang pemasaran, klasifikasi teks dapat memiliki berbagai aplikasi. Klasifikasi teks dapat digunakan untuk pembahasan sentimen, di mana ulasan pelanggan, komentar, atau unggahan media sosial diklasifikasikan ke dalam sentimen positif, negatif, atau netral. Hal ini memungkinkan pemasar untuk mendapatkan wawasan tentang opini dan perasaan pelanggan terhadap produk atau layanan mereka.

Klasifikasi teks juga dapat digunakan untuk deteksi spam dalam pemasaran email. Dengan mengklasifikasikan email ke dalam spam atau bukan spam, pemasar dapat

memastikan email pemasaran mereka tidak salah ditandai sebagai spam dan sampai ke penerima yang dituju. Dalam pemasaran konten, klasifikasi teks dapat digunakan untuk penandaan atau kategorisasi otomatis artikel atau unggahan blog. Dengan menetapkan tag atau kategori yang relevan untuk setiap konten, pemasar dapat meningkatkan kemampuan penemuan dan rekomendasi konten, yang mengarah pada pengalaman pengguna yang lebih baik. Selain itu, klasifikasi teks dapat digunakan untuk otomatisasi layanan pelanggan. Dengan mengklasifikasikan pertanyaan atau keluhan pelanggan ke dalam berbagai kategori, perusahaan dapat secara otomatis mengarahkan setiap pertanyaan ke departemen yang tepat atau menghasilkan respons otomatis, sehingga menghasilkan layanan pelanggan yang lebih efisien. Klasifikasi teks membekali pemasar dengan mekanisme yang kuat untuk membahas dan mengkategorikan sejumlah besar data teks, menghasilkan wawasan dan efisiensi berharga yang membentuk strategi pemasaran.

6.5 KOMUNIKASI INSTAN MELALUI BOT DAN ASISTEN VIRTUAL

Dalam lanskap digital kontemporer, komunikasi instan dan interaksi yang berpusat pada pengguna telah menjadi norma. Perusahaan dan merek semakin beralih ke solusi berbasis teknologi, seperti chatbot dan asisten suara, untuk mendorong interaksi waktu nyata dan meningkatkan pengalaman pelanggan. Teknologi ini tidak hanya memfasilitasi perjalanan pelanggan yang lancar tetapi juga memberikan data dan wawasan yang tak ternilai kepada pemasar untuk menyesuaikan strategi mereka. Pada bagian ini, kita akan mengeksplorasi evolusi, praktik terbaik, dan implikasi chatbot dan asisten suara dalam pemasaran.

Evolusi dan Pentingnya Chatbot

Evolusi. Chatbot, meskipun sekarang ada di mana-mana, memiliki asal-usul yang sederhana. Berawal sebagai sistem berbasis aturan di mana interaksi terbatas pada input dan output yang telah ditentukan sebelumnya, chatbot modern sekarang menggunakan algoritma NLP canggih yang memungkinkan percakapan yang lebih lancar dan mirip manusia. Dengan munculnya platform seperti Facebook Messenger dan WhatsApp, integrasi chatbot ke dalam platform ini telah mendorong pertumbuhan eksponensial dalam adopsinya.

Pentingnya. Chatbot merevolusi layanan pelanggan dan keterlibatan dalam beberapa cara. Mereka menyediakan dukungan pelanggan 24/7, menangani banyak pertanyaan secara bersamaan, dan menawarkan respons instan, yang mengarah pada peningkatan kepuasan pelanggan. Selain itu, data yang mereka kumpulkan membantu pemasar dalam memahami perilaku pengguna, preferensi, dan masalah yang dihadapi, sehingga memungkinkan kampanye pemasaran yang lebih terarah dan personal.

Merancang Chatbot yang Efektif: Praktik Terbaik

- **Desain berpusat pada pengguna.** Desain harus mencerminkan kebutuhan utama audiens Anda. Chatbot harus dirancang dengan mempertimbangkan maksud pengguna, memastikan bahwa pertanyaan yang paling umum ditangani secara efisien.
- **Pengalihan yang mulus ke manusia.** Chatbot harus mengenali kapan pertanyaan pengguna terlalu kompleks dan secara mulus mengalihkan percakapan ke perwakilan manusia.

- **Siklus umpan balik iteratif.** Kumpulkan umpan balik pengguna secara teratur dan gunakan data ini untuk menyempurnakan dan meningkatkan respons dan fungsionalitas chatbot.
- **Pemahaman bahasa alami.** Gabungkan teknik NLP tingkat lanjut untuk memastikan chatbot memahami masukan pengguna dengan lebih akurat dan memberikan keluaran yang relevan.
- **Interaksi multimodal.** Selain teks, pertimbangkan untuk mengintegrasikan konten multimedia, seperti gambar atau video, untuk memperkaya interaksi pengguna.

Untuk lebih memperjelas praktik terbaik desain chatbot, Tabel 6.4 memberikan contoh konkret untuk setiap prinsip, yang menggambarkan bagaimana prinsip-prinsip tersebut dapat diimplementasikan secara efektif untuk meningkatkan interaksi dan kepuasan pengguna.

Tabel 6.4 Praktik Terbaik untuk Desain Chatbot dengan Contoh.

Praktik Terbaik	Contoh
Tetapkan tujuan yang jelas untuk chatbot	Chatbot yang dirancang untuk membantu menjawab pertanyaan produk di situs e-commerce
Rancang untuk audiens target	Chatbot untuk situs web game yang menggunakan jargon game
Pastikan peralihan yang lancar ke agen manusia	Ketika chatbot tidak dapat menjawab pertanyaan, ia mengarahkan pengguna ke perwakilan layanan pelanggan.
Pilih nada percakapan	Alih-alih "Pertanyaan tidak dipahami," chatbot mengatakan "Ups, saya tidak menangkapnya. Bisakah Anda mengulangnya?"
Terapkan mekanisme umpan balik	Setelah sesi obrolan, bot bertanya, "Apakah saya membantu menjawab pertanyaan Anda?"

Asisten Suara dan Pencarian Suara

Asisten suara, seperti Alexa dari Amazon, Google Assistant, dan Siri dari Apple, telah menandai pergeseran paradigma dalam perilaku pengguna dan interaksi merek. Asisten digital ini semakin menjadi antarmuka utama bagi pengguna untuk mengumpulkan informasi, melakukan tugas, dan terhubung dengan merek.

Adopsi speaker pintar, sistem suara dalam mobil, dan perangkat wearable yang diaktifkan suara telah mendorong peningkatan pencarian suara. Seiring semakin banyak rumah tangga mengintegrasikan perangkat ini ke dalam rutinitas harian mereka, pencarian suara siap melampaui kueri teks tradisional. Bagi pemasar, ini tidak hanya menandakan perubahan dalam media tetapi juga perubahan mendasar dalam cara konsumen mengartikulasikan kebutuhan dan harapan mereka.

Karakteristik Pencarian Suara

- **Nada percakapan.** Kueri suara pada dasarnya lebih bersifat percakapan dan alami, sering dirumuskan sebagai pertanyaan. Misalnya, "Di mana saya dapat menemukan gaun musim panas biru?" alih-alih "toko gaun musim panas biru di dekat saya."

- **Kata kunci ekor panjang.** Karena sifatnya yang seperti percakapan, pencarian suara cenderung lebih panjang dan lebih spesifik, menekankan pentingnya optimasi kata kunci ekor panjang dalam konten.
- **Berorientasi lokal.** Banyak pencarian suara memiliki tujuan lokal, seperti “Di mana toko pakaian terdekat?” Ini menggarisbawahi pentingnya SEO lokal untuk bisnis.

Tantangan dan Pertimbangan bagi Pemasar

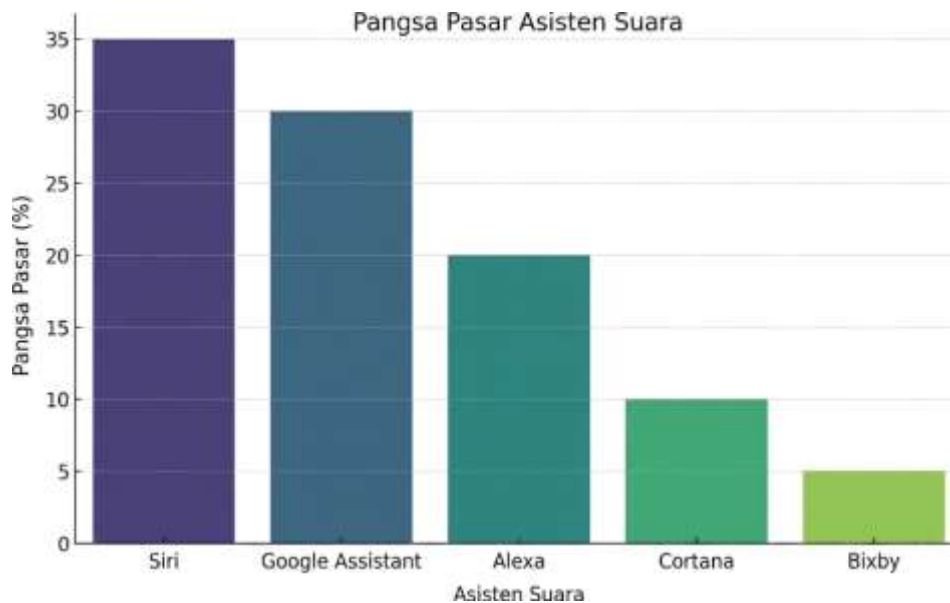
- **Dinamika SEO yang berubah.** Praktik SEO tradisional berfokus pada pola pencarian berbasis teks. Pencarian suara menuntut perombakan strategi ini, dengan memprioritaskan bahasa alami, kueri berbasis pertanyaan, dan cuplikan unggulan yang mungkin dibacakan oleh asisten suara.
- **Memastikan respons yang cepat dan relevan.** Pengguna pencarian suara mengharapkan jawaban yang cepat dan akurat. Konten harus terstruktur untuk memberikan respons langsung terhadap potensi pertanyaan suara.
- **Adaptasi terhadap berbagai platform.** Setiap asisten suara memiliki algoritma dan preferensi yang unik. Pemasar perlu memahami nuansa setiap platform untuk mengoptimalkan secara efektif.
- **Kehilangan ruang layar.** Tidak seperti pencarian berbasis teks yang menampilkan banyak hasil, asisten suara biasanya memberikan satu jawaban, sehingga persaingan untuk posisi teratas yang didambakan menjadi semakin ketat.
- **Kekhawatiran privasi.** Sifat beberapa perangkat yang selalu mendengarkan telah menimbulkan kekhawatiran privasi di kalangan pengguna. Pemasar harus berhati-hati, memastikan praktik pengumpulan dan penggunaan data yang transparan.
- **Memonetisasi pencarian suara.** Iklan online tradisional tidak dapat dengan mudah diterjemahkan ke interaksi suara. Merek harus menemukan cara inovatif untuk mengintegrasikan diri ke dalam hasil pencarian suara tanpa mengganggu pengalaman pengguna.

Pendekatan Inovatif dalam Pemasaran Pencarian Suara

- **Membuat aplikasi atau keterampilan suara.** Merek dapat mengembangkan aplikasi khusus untuk asisten suara, seperti Alexa Skills atau Google Actions, menawarkan pengguna pengalaman yang lebih interaktif dan bermerek.
- **FAQ dan konten yang kaya.** Membuat bagian FAQ dan memberikan jawaban terperinci untuk pertanyaan yang sering diajukan dapat memposisikan merek sebagai otoritas dan meningkatkan peluang untuk menjadi jawaban unggulan dalam pencarian suara.
- **Terlibat dalam pemasaran percakapan.** Merangkul chatbot dan asisten yang diaktifkan suara untuk keterlibatan konsumen langsung, menjawab pertanyaan, menawarkan saran, dan bahkan memfasilitasi pembelian dapat mengubah pengalaman berbelanja.

Dalam lanskap interaksi berbasis suara yang terus berkembang, pemasar perlu tetap gesit, terus beradaptasi dengan norma dan harapan konsumen yang berubah. Merek yang dapat secara efektif mengintegrasikan pencarian suara ke dalam strategi digital mereka akan mendapatkan keunggulan signifikan di arena yang berkembang pesat ini. Saat pemasar menavigasi lanskap yang mengutamakan suara, mengenali distribusi pangsa pasar di antara

berbagai asisten suara sangatlah penting. Gambar 6.5 menyajikan grafik batang yang menggambarkan perkiraan pangsa pasar, menawarkan tolok ukur visual dari bidang kompetitif di mana merek-merek bersaing untuk membangun kehadiran pencarian suara mereka.



Gambar 6.5 Estimasi Pangsa Pasar untuk Berbagai Asisten Suara.

Implikasi dan Tantangan Pemasaran yang Diaktifkan Suara

Implikasi

- **Pergeseran strategi SEO.** Karena pencarian suara cenderung ke arah nada percakapan, pemasar perlu memprioritaskan frasa panjang dan kalimat lengkap daripada kata kunci tradisional. Ini berarti mengoptimalkan untuk kueri bahasa alami seperti “Apa tabir surya terbaik untuk kulit sensitif?” daripada “tabir surya terbaik.”
- **Kepuasan instan.** Pengguna pencarian suara tidak memiliki kesabaran untuk menyaring banyak hasil pencarian. Mereka mengharapkan jawaban langsung dan akurat. Hal ini mewajibkan merek untuk membuat konten yang tidak hanya relevan tetapi juga ringkas dan tepat sasaran.
- **Peningkatan personalisasi.** Asisten suara, yang memiliki akses ke data pengguna dan interaksi sebelumnya, dapat menawarkan respons yang disesuaikan atau saran produk. Ini membuka jalan bagi strategi pemasaran yang sangat personal, di mana penawaran dan rekomendasi disesuaikan dengan perilaku dan preferensi pengguna individu.

Tantangan

Privasi data:

- **Kekhawatiran.** Asisten suara mengumpulkan data pengguna yang ekstensif, mulai dari rutinitas harian hingga preferensi pribadi. Akumulasi data ini telah meningkatkan kekhawatiran tentang privasi pengguna dan potensi penyalahgunaan informasi pribadi.
- **Mengatasi tantangan ini.** Merek harus memprioritaskan kebijakan pengumpulan dan penggunaan data yang transparan. Mereka juga harus memberi pengguna kendali atas data mereka, memungkinkan mereka untuk memilih ikut serta atau tidak dalam proses

pengumpulan data, memastikan kepercayaan dan kepatuhan terhadap peraturan privasi.

Monetisasi:

- **Kekhawatiran.** Banyak perangkat yang diaktifkan suara, seperti speaker pintar, tidak memiliki layar. Ini berarti iklan visual tradisional, seperti banner atau iklan video, tidak memungkinkan.
- **Mengatasi tantangan ini.** Pemasar harus berpikir di luar kotak. Strategi potensial termasuk konten bersponsor atau aplikasi suara bermerek. Pendekatan lain adalah bermitra dengan penyedia platform suara untuk menjadi solusi "pilihan utama", memastikan merek mereka adalah pilihan pertama yang direkomendasikan ketika pertanyaan yang relevan diajukan.

Interoperabilitas:

- **Kekhawatiran.** Dengan beberapa platform suara di pasaran (Alexa, Google Assistant, Siri, dan lain-lain.), tidak ada satu solusi yang cocok untuk semua. Setiap platform memiliki algoritma dan perilaku pengguna yang unik.
- **Mengatasi tantangan ini.** Merek harus berupaya untuk menyediakan konten dan solusi yang tidak bergantung pada platform jika memungkinkan. Hal ini memastikan pengalaman merek yang konsisten di berbagai perangkat. Selain itu, memahami nuansa dan kekuatan setiap platform dapat membantu merek menyesuaikan strategi mereka untuk mendapatkan hasil terbaik di setiap platform.

Singkatnya, meningkatnya penggunaan chatbot dan asisten suara dalam kehidupan sehari-hari kita sedang membentuk kembali lanskap pemasaran. Teknologi ini, meskipun menawarkan jalan untuk keterlibatan dan personalisasi yang lebih dalam, membawa serta serangkaian tantangan baru. Pemasar yang dapat secara efektif memanfaatkan potensi suara sambil dengan terampil mengatasi tantangannya akan memposisikan diri mereka di garis depan revolusi yang mengutamakan suara ini.

6.6 PEMBAHASAN SENTIMEN MEDIA SOSIAL

Pembahasan sentimen media sosial adalah penggunaan NLP, pembahasan teks, dan linguistik komputasional untuk mengidentifikasi dan mengekstrak informasi subjektif dari materi sumber di platform media sosial. Hal ini membantu menentukan sikap pembicara atau penulis terhadap suatu topik atau polaritas kontekstual keseluruhan dari sebuah dokumen.

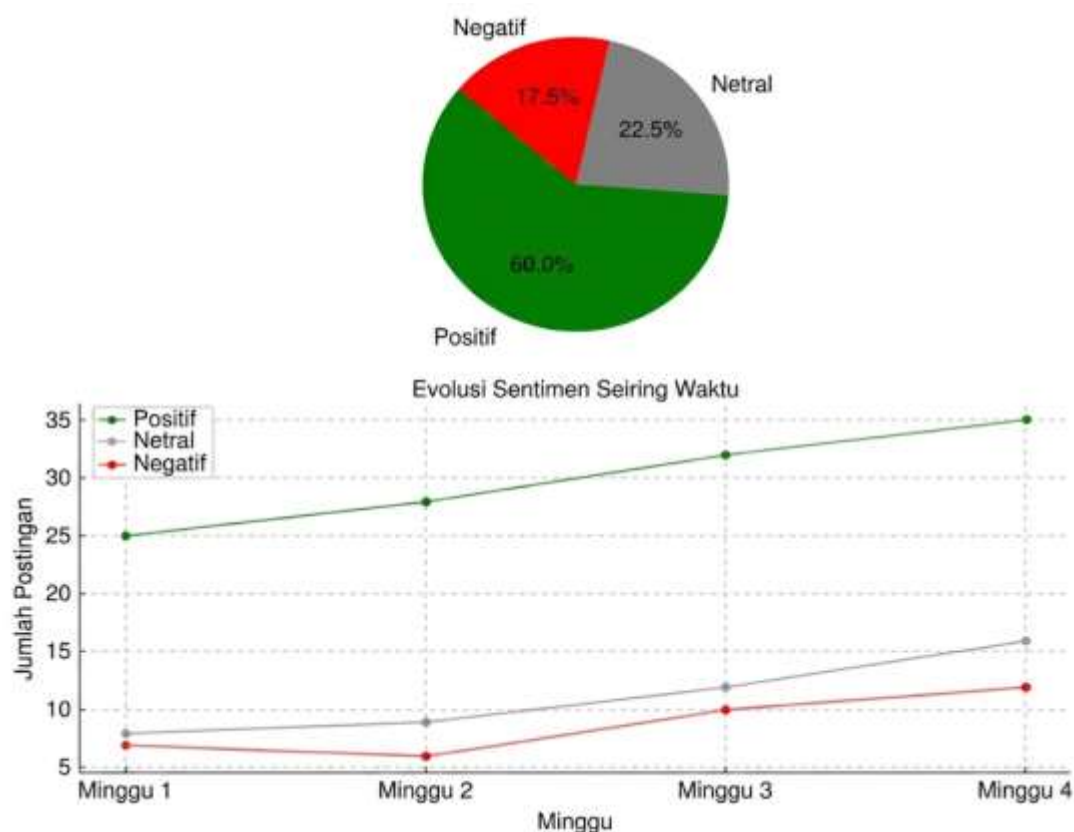
Dalam konteks pemasaran, pembahasan sentimen media sosial memungkinkan bisnis untuk mengidentifikasi sentimen konsumen terhadap produk, merek, atau layanan dalam percakapan dan umpan balik online. Hasil pembahasan dapat dimanfaatkan untuk memahami emosi pelanggan terhadap suatu merek, untuk mengukur efektivitas kampanye pemasaran, dan untuk mengidentifikasi potensi krisis sebelum krisis tersebut meningkat.

Misalnya, sebuah perusahaan dapat memantau tweet tentang merek mereka, mengkategorikannya sebagai positif, negatif, atau netral. Hal ini akan memungkinkan perusahaan untuk segera mengidentifikasi ketidakpuasan pelanggan dan menangani situasi tersebut sebelum merusak citra merek mereka.

Lebih lanjut, pembahasan sentimen dapat memandu proses pembuatan konten. Dengan memahami apa yang dihargai atau tidak disukai pelanggan tentang suatu produk atau layanan, bisnis dapat membuat pesan yang membahas masalah ini dan beresonansi dengan audiens mereka.

Terakhir, pembahasan sentimen juga dapat membantu dalam pembahasan kompetitif. Dengan membandingkan sentimen terhadap merek mereka dengan sentimen pesaing, bisnis dapat mengidentifikasi area di mana mereka perlu meningkatkan dan menemukan peluang untuk diferensiasi.

Untuk mengilustrasikan penerapan praktis pembahasan sentimen media sosial dalam pemasaran, Gambar 6.6 menyediakan tangkapan layar dasbor yang menunjukkan hasil pembahasan tersebut yang berasal dari konten media sosial, menawarkan visualisasi yang jelas tentang sentimen pelanggan yang dikategorikan sebagai positif, negatif, atau netral.



Gambar 6.6 Tangkapan Layar Dasbor yang Menunjukkan Hasil Pembahasan Sentimen dari Media Sosial.

Kesimpulannya, pembahasan sentimen media sosial menyediakan alat yang ampuh bagi pemasar untuk memahami dan bereaksi terhadap sentimen pelanggan secara real-time, memberikan wawasan berharga yang dapat menginformasikan strategi pemasaran.

Chatbot untuk Layanan Pelanggan

Seperti yang dibahas di Bagian 6.4, dalam pemasaran, chatbot berfungsi sebagai alat inovatif untuk melibatkan pelanggan, menjawab pertanyaan, dan memberikan rekomendasi yang dipersonalisasi secara real-time, kapan saja. NLP merupakan komponen kunci dalam

desain dan fungsi chatbot, yang memungkinkan mereka untuk memahami dan menanggapi masukan pengguna dengan cara seperti manusia.



Gambar 6.7 Alur Percakapan yang Menggambarkan Bagaimana Chatbot Menangani Interaksi Layanan Pelanggan.

Misalnya, pelanggan dapat berinteraksi dengan chatbot pengecer pakaian untuk memeriksa ketersediaan barang tertentu atau mencari informasi tentang koleksi mode terbaru. Chatbot, menggunakan NLP, dapat memahami maksud pengguna dan memberikan respons atau tindakan yang sesuai. Hal ini membantu mengurangi waktu tunggu dan meningkatkan pengalaman pelanggan, yang dapat meningkatkan retensi dan loyalitas pelanggan.

Untuk memberikan contoh nyata bagaimana chatbot meningkatkan layanan pelanggan dalam pemasaran, Gambar 6.7 menyajikan diagram alur percakapan, yang menggambarkan interaksi tipikal antara pelanggan dan chatbot. Visual ini mengilustrasikan kemampuan chatbot untuk memahami dan menanggapi pertanyaan pelanggan, menunjukkan efisiensi dan efektivitas asisten digital ini dalam keterlibatan pelanggan secara real-time.

Selain itu, chatbot dapat mengumpulkan dan membahas data pelanggan secara real-time, memberikan wawasan berharga kepada bisnis tentang perilaku dan preferensi pelanggan. Data ini kemudian dapat digunakan untuk mempersonalisasi kampanye pemasaran dan meningkatkan penawaran produk. Namun, sangat penting bagi perusahaan untuk menangani implementasi chatbot dengan bijak, memastikan keseimbangan antara otomatisasi

dan sentuhan manusia. Meskipun chatbot dapat menangani pertanyaan rutin secara efektif, intervensi manusia mungkin masih diperlukan untuk masalah yang kompleks atau sensitif.

Dengan bantuan NLP, chatbot merevolusi layanan pelanggan dalam pemasaran, memberikan bisnis instrumen yang berharga untuk meningkatkan keterlibatan pelanggan dan memperoleh wawasan.

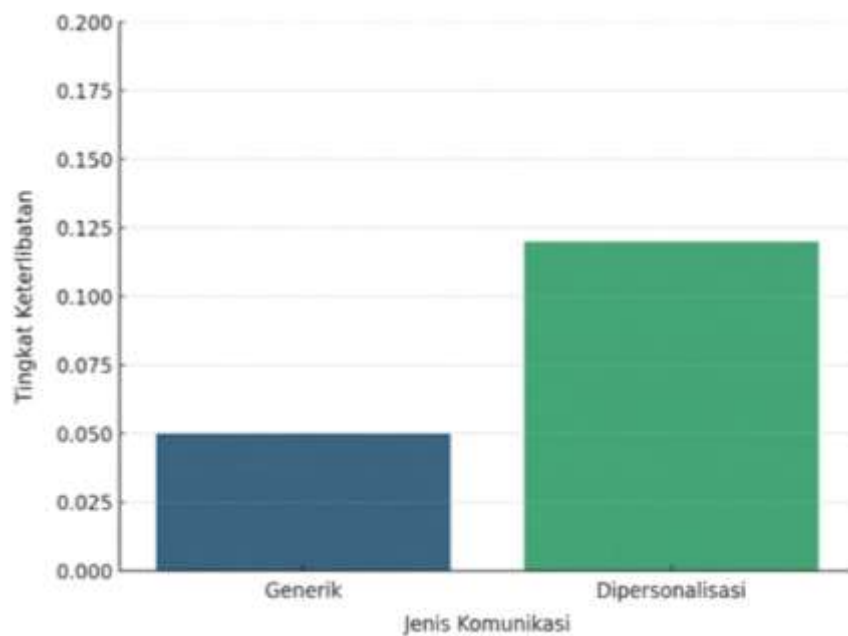
Komunikasi Pemasaran yang Dipersonalisasi

Personalisasi dalam komunikasi pemasaran telah mendapatkan perhatian yang cukup besar dalam beberapa tahun terakhir sebagai alat yang ampuh untuk meningkatkan hubungan pelanggan dan mendorong penjualan. Ini melibatkan penyesuaian pesan kepada pelanggan individu berdasarkan preferensi, perilaku, dan informasi demografis mereka. Tingkat kustomisasi ini dimungkinkan oleh teknik pembahasan data tingkat lanjut, termasuk NLP. NLP menyediakan cara untuk memahami dan menghasilkan bahasa manusia dengan cara yang bermakna dan personal. Teknologi ini dapat digunakan untuk pembahasaninteraksi pelanggan, seperti email, unggahan media sosial, dan ulasan pelanggan, untuk mendapatkan wawasan tentang preferensi dan sentimen pelanggan individu.

Misalnya, perusahaan e-commerce dapat menggunakan NLP untuk pembahasanulasan pelanggan dan mengidentifikasi tema atau sentimen umum. Informasi ini kemudian dapat digunakan untuk mempersonalisasi kampanye pemasaran email, menyarankan produk berdasarkan minat dan sentimen yang diungkapkan pelanggan.

Selain itu, NLP dapat membantu menghasilkan konten yang dipersonalisasi untuk setiap pelanggan. Misalnya, pembuatan konten dinamis dapat menciptakan deskripsi produk atau pesan promosi unik yang menarik secara khusus bagi penerima berdasarkan interaksi dan preferensi mereka sebelumnya. Untuk mendemonstrasikan secara visual dampak personalisasi dalam komunikasi pemasaran, Gambar 6.8 memberikan perbandingan sebelum dan sesudah, yang menggambarkan transformasi dari pesan generik menjadi komunikasi yang dipersonalisasi berdasarkan data dan wawasan pelanggan.

Namun, penting untuk dicatat bahwa meskipun personalisasi dapat meningkatkan keterlibatan pelanggan dan tingkat konversi, hal itu harus dilakukan dengan pemahaman yang mendalam tentang pelanggan dan menghormati privasi mereka. Pesan personalisasi yang terlalu mengganggu atau tidak relevan dapat memiliki efek sebaliknya, mengasingkan pelanggan dan merusak hubungan merek-pelanggan.



Gambar 6.8 Perbandingan Sebelum dan Sesudah Komunikasi Pemasaran Generik Versus Personalisasi.

Kesimpulannya, NLP menawarkan peluang signifikan untuk mempersonalisasi komunikasi pemasaran, sehingga meningkatkan hubungan pelanggan dan meningkatkan efektivitas pemasaran. Namun, penerapannya harus diimbangi dengan pemahaman mendalam tentang pelanggan dan penghormatan terhadap privasi mereka.

6.6 KESIMPULAN

Dalam bab ini, kita telah menjelajahi ranah transformatif NLP dan implikasinya yang mendalam bagi pemasaran modern. Saat kita berada di persimpangan teknologi dan komunikasi manusia, peluang yang ditawarkan NLP kepada pemasar sangat luas dan revolusioner. Dari memahami aspek-aspek dasar NLP hingga menyelami teknik-teknik spesifik seperti pembahasan sentimen dan penggunaan chatbot, jelas bahwa jembatan antara bahasa manusia dan pemrosesan komputasi semakin menguat. Jembatan ini tidak hanya memungkinkan bisnis untuk menavigasi lautan data tidak terstruktur yang luas, tetapi juga untuk menguraikan sentimen, keinginan, dan kebutuhan audiens mereka.

Munculnya chatbot dan asisten suara merupakan contoh konvergensi utilitas dan pengalaman. Alat-alat ini, yang didukung oleh NLP, menyediakan interaksi personal secara real-time, membentuk kembali ekspektasi konsumen digital dan menetapkan tolok ukur baru untuk pengalaman pelanggan. Namun, seperti semua teknologi, sangat penting untuk menggunakannya dengan bijaksana, mengenali kekuatan dan keterbatasannya. Tantangan sarkasme, ambiguitas, dan nuansa budaya dalam bahasa menggarisbawahi kompleksitas komunikasi manusia dan mengingatkan kita bahwa meskipun teknologi dapat meningkatkan upaya kita, sentuhan manusialah yang seringkali membuat perbedaan.

Aplikasi NLP dalam pemasaran—baik dalam pembahasan sentimen media sosial, personalisasi konten, atau otomatisasi layanan pelanggan—bukan hanya alat untuk efisiensi tetapi juga instrumen wawasan. Aplikasi ini memungkinkan merek untuk mendengarkan dengan saksama, merespons dengan tepat, dan mengantisipasi kebutuhan, menjalin jalinan kepercayaan dan loyalitas dengan audiens mereka. Saat kita mengakhiri bab ini, penting untuk diingat bahwa lanskap pemasaran dan teknologi terus berkembang. Meskipun NLP menyediakan berbagai kemampuan saat ini, masa depan pasti akan membawa kemajuan, tantangan, dan peluang baru. Bagi para pemasar, seruan untuk bertindak sudah jelas: tetaplah ingin tahu, tetaplah mudah beradaptasi, dan selalu berusaha untuk meningkatkan dialog dengan audiens Anda. Dalam simfoni bisnis dan teknologi, semoga merek Anda selalu menemukan nada yang tepat.

LATIHAN 6.1

PEMBAHASAN SENTIMEN

Tujuan: Menulis skrip Python untuk melakukan pembahasan sentimen pada postingan media sosial yang diberikan.

Tugas:

1. Muat file “**sentiment_analysis_data.csv**” ke dalam program Python.
2. Gunakan pustaka pembahasan sentimen **TextBlob** untuk pembahasansentimen setiap postingan.
3. Kategorikan setiap postingan sebagai ‘Positif’, ‘Negatif’, atau ‘Netral’ berdasarkan skor sentimen.
4. Keluarkan hasil pembahasan sentimen dalam format yang mudah dibaca.

Langkah-langkah:

1. Muat Data:

- Pertama, impor pustaka yang diperlukan dan baca file CSV yang berisi data sentimen.

```
1. import pandas as pd
2. sentiment_df = pd.read_csv('path_to_csv_file')
```

2. Instal dan Impor TextBlob:

- Instal TextBlob, jika belum terinstal, menggunakan perintah `!pip install textblob`.
- Impor TextBlob untuk pembahasan sentimen.

```
3. !pip install textblob
4. from textblob import TextBlob
```

3. Lakukan Pembahasan Sentimen:

- Definisikan fungsi untuk pembahasan sentimen menggunakan TextBlob.
- Terapkan fungsi ini pada setiap postingan dalam dataset.

```
5. def analyze_sentiment(post):
6.     analysis = TextBlob(post)
7.     return 'Positive' if analysis.sentiment.polarity > 0 else 'Negative'
   if analysis.sentiment.polarity < 0 else 'Neutral'
```

```
8. sentiment_df['Analyzed_Sentiment']=sentiment_df['Post'].apply(analyze_sentiment)
```

4. Tampilkan Hasil:

- Tampilkan beberapa baris pertama dari DataFrame yang berisi postingan asli dan sentimen yang telah dipembahasan.

```
9. sentiment_df.head()
```

Kode ini menyediakan pendekatan langkah demi langkah untuk melakukan pembahasan sentimen pada data teks menggunakan Python dan pustaka TextBlob. Sentimen setiap postingan dipembahasan dan dikategorikan sebagai 'Positif', 'Negatif', atau 'Netral' berdasarkan kontennya.

LATIHAN 6.2

KLASIFIKASI TEKS

Tujuan

Mengembangkan program Python untuk mengklasifikasikan ulasan pelanggan ke dalam kategori seperti elektronik, pakaian, atau makanan.

Tugas:

1. Muat data “**text_classification_data.csv**”.
2. Implementasikan model klasifikasi teks menggunakan pustaka pembelajaran mesin **scikit-learn**.
3. Latih model pada sebagian data dan uji pada data yang tersisa.
4. Evaluasi kinerja model menggunakan metrik seperti akurasi dan matriks kebingungan.

Langkah-langkah:

1. Muat Data:

- Impor pandas untuk membaca file CSV.
- Muat data klasifikasi teks.

```
1. import pandas as pd
2. classification_df=pd.read_csv('path_to_classification_csv_file')
```

2. Siapkan Data:

- Pisahkan data menjadi fitur (ulasan) dan label (kategori).

```
3. X = classification_df['Review']
4. y = classification_df['Category']
```

3. Preproses Teks dan Ekstraksi Fitur:

- Gunakan TfidfVectorizer untuk mengubah data teks menjadi vektor numerik.

```
5. from sklearn.feature_extraction.text import TfidfVectorizer
6. vectorizer = TfidfVectorizer()
7. X = vectorizer.fit_transform(X)
```

4. Pisahkan Dataset:

- Bagi dataset menjadi set pelatihan dan pengujian.

```
8. from sklearn.model_selection import train_test_split
```



```
9. X_train, X_test, y_train, y_test = train_test_split(X, y,  
    test_size=0.2)
```

5. Pelatihan Model:

- Pilih model klasifikasi (misalnya, Naive Bayes multinomial).
- Latih model pada set pelatihan.

```
10. from sklearn.naive_bayes import MultinomialNB  
11. model = MultinomialNB()  
12. model.fit(X_train, y_train)
```

6. Evaluasi Model:

- Prediksi kategori pada set pengujian.
- Evaluasi kinerja model menggunakan metrik seperti akurasi, presisi, dan recall.

```
13. from sklearn.metrics import classification_report  
14. y_pred = model.predict(X_test)  
15. print(classification_report(y_test, y_pred))
```

Proses ini melibatkan langkah-langkah tipikal dalam alur kerja pembelajaran mesin: persiapan data, ekstraksi fitur, pelatihan model, dan evaluasi.

BAB 7

MEMBACA PERILAKU PASAR DI MEDIA SOSIAL DAN SITUS WEB

7.1 PENDAHULUAN

Revolusi digital telah sangat memengaruhi lanskap pemasaran. Seiring dengan perpindahan bisnis dan konsumen ke dunia online, platform media sosial dan situs web telah menjadi titik kontak penting bagi merek. Platform-platform ini menghasilkan banyak data, memberikan wawasan unik tentang perilaku, preferensi, dan sentimen konsumen. Kemampuan untuk memahami dan menggunakan data ini sangat penting bagi pemasar di lingkungan yang mengutamakan digital saat ini.

Pembahasan media sosial dan pembahasan web adalah dua metodologi utama untuk memanfaatkan data ini. Pembahasan media sosial melibatkan pengumpulan dan pembahasan data dari platform media sosial untuk menginformasikan keputusan pemasaran. Ini dapat membantu pemasar memahami audiens mereka, mengukur sentimen terhadap merek mereka, dan melacak efektivitas kampanye media sosial mereka. Namun, pembahasan web berfokus pada pembahasan data yang dihasilkan oleh aktivitas pengunjung di situs web. Ini dapat mencakup metrik seperti tampilan halaman, rasio pantulan, durasi sesi, dan rasio konversi. Pembahasan web dapat memberikan wawasan tentang perilaku pengguna, memungkinkan pemasar untuk mengoptimalkan situs web mereka untuk pengalaman pengguna (UX) yang lebih baik dan peningkatan konversi.

Dalam bab ini, kami akan memberikan gambaran umum yang komprehensif tentang pembahasan media sosial dan pembahasan web, mengeksplorasi konsep-konsep kunci, teknik, dan aplikasi praktis. Kami juga akan membahas metodologi spesifik dalam domain ini, termasuk pembahasan jaringan sosial, pemantauan dan pelacakan media sosial, dan optimasi tingkat konversi. Saat kita menelusuri bab ini, kita akan memahami bagaimana teknik-teknik ini dapat digunakan secara efektif untuk mendorong keberhasilan pemasaran di era digital.

7.2 JARINGAN SOSIAL (SNA)

Sebelum menyelami lebih dalam alat pembahasan, penting untuk memahami kerangka kerja yang mendasari pemeriksaan kita. Dunia media sosial secara intrinsik terhubung, menyerupai jaring luas dari simpul-simpul yang saling terkait.

Gambaran Umum Pembahasan Jaringan Sosial

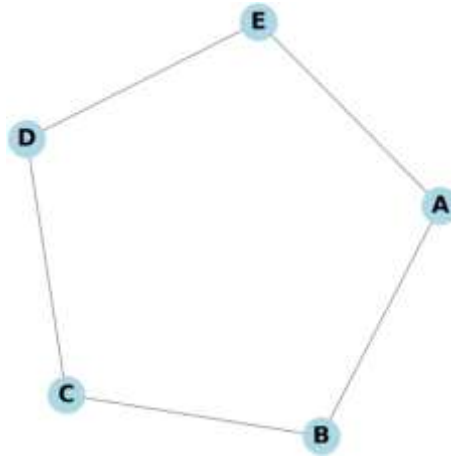
Pembahasan jaringan sosial (SNA) adalah teknik pembelajaran yang digunakan untuk memvisualisasikan dan mengpembahasan hubungan antara berbagai entitas. Dalam konteks pemasaran, entitas-entitas ini seringkali berupa individu atau organisasi yang berinteraksi di platform media sosial. SNA menyediakan cara untuk memetakan dan mengukur hubungan yang kompleks, dan terkadang tersembunyi, yang seringkali sulit divisualisasikan atau dipahami dengan cara tradisional. SNA mencakup berbagai metrik dan konsep, termasuk simpul (aktor individu dalam jaringan), ikatan (hubungan antar aktor), sentralitas (ukuran pentingnya suatu simpul dalam jaringan), dan kepadatan (tingkat konektivitas umum dalam

suatu jaringan). Melalui metrik-metrik ini, SNA dapat membantu pemasar memahami struktur dan dinamika audiens media sosial mereka. Misalnya, metrik sentralitas dapat mengidentifikasi influencer atau pemimpin opini kunci dalam suatu jaringan, sedangkan metrik kepadatan dapat memberikan wawasan tentang keterlibatan keseluruhan audiens. Gambar 7.1 menyajikan grafik berbentuk segi lima dengan simpul berlabel A, B, C, D, dan E, masing-masing mewakili aktor individu dalam jaringan.

Simpul-simpul ini dihubungkan oleh garis penghubung yang menunjukkan hubungan di antara mereka. Susunan geometris dalam segi lima menunjukkan bahwa setiap simpul terhubung ke dua simpul lainnya, menggambarkan lingkaran tertutup di mana informasi atau pengaruh dapat mengalir secara melingkar. Visualisasi ini berfungsi sebagai contoh dasar bagaimana individu atau entitas saling terhubung dalam jaringan sosial. Penempatan simpul dan garis penghubungnya (garis penghubung) memberikan wawasan tentang struktur jaringan. Misalnya, tidak adanya garis langsung antara simpul yang tidak berdekatan, seperti A dan C atau B dan D, menyoroti kurangnya jalur komunikasi langsung antara aktor-aktor tertentu. Struktur ini mungkin menunjukkan tingkat hierarki atau penjaga gerbang, di mana informasi harus melewati simpul yang berdekatan untuk mencapai simpul lain dalam jaringan.

Bentuk simetris pentagon menunjukkan jaringan dengan distribusi hubungan yang sama di antara simpul-simpulnya. Setiap simpul memiliki jumlah koneksi yang sama, menunjukkan tingkat pengaruh atau aksesibilitas yang seragam dalam model jaringan tertentu ini. Dalam konteks pemasaran, ini mungkin menyiratkan bahwa pesan memiliki peluang yang sama untuk disebarkan melalui jaringan dari titik awal mana pun, dengan asumsi bahwa kekuatan dan pengaruh setiap koneksi sama. Namun, penting untuk dicatat bahwa jaringan sosial di dunia nyata seringkali lebih kompleks, dengan berbagai tingkat konektivitas dan sentralitas. Dengan demikian, meskipun Gambar 7.1 memberikan pandangan yang disederhanakan, ia membentuk dasar untuk memahami jalur potensial yang dapat dilalui informasi dan pengaruh dalam jaringan sosial. Dengan meng pembahasan struktur tersebut, pemasar dapat merancang strategi untuk mengoptimalkan komunikasi dan pengaruh dalam audiens target mereka.

Dengan memberikan pemahaman yang lebih bernuansa tentang interaksi media sosial, SNA dapat memberikan informasi untuk berbagai keputusan pemasaran. Ini dapat membantu merek mengidentifikasi influencer potensial untuk kampanye pemasaran, memahami penyebaran informasi atau sentimen dalam jaringan, atau bahkan memprediksi tren masa depan berdasarkan dinamika jaringan.



Gambar 7.1 Jaringan Sosial Sederhana, Mengilustrasikan Node dan Edge.

Dasar-Dasar Teori Jaringan

Untuk memahami pola-pola ini, pertama-tama kita perlu membangun pemahaman dasar tentang teori jaringan, yang mencakup komponen, jenis, dan strukturnya. Teori jaringan adalah bidang multidisiplin yang mempelajari jaringan untuk memahami struktur, perilaku, evolusi, dan interkoneksi elemen-elemen penyusunnya. Dari biologi hingga sosiologi dan ilmu komputer, jaringan muncul di mana-mana. Dalam konteks pembelajaran media sosial, teori jaringan menyediakan lensa untuk melihat dan memahami jalinan hubungan dan interaksi yang rumit di platform.

Komponen Jaringan

- Node. Entitas individual dalam jaringan, yang dapat mewakili pengguna, halaman web, atau entitas lainnya.
- Edge. Koneksi atau hubungan antar node. Untuk media sosial, ini bisa berupa pertemanan, pengikut, atau interaksi apa pun yang menghubungkan dua node.

Jenis Jaringan

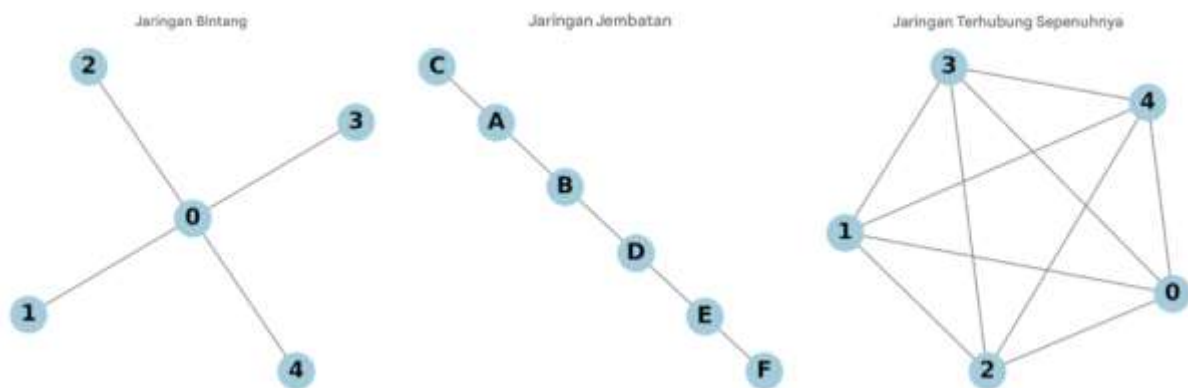
- Jaringan tak berarah. Di mana edge tidak memiliki arah, misalnya, pertemanan Facebook.
- Jaringan berarah. Di mana edge memiliki arah; misalnya, X (sebelumnya Twitter) mengikuti di mana satu pengguna mengikuti pengguna lain tidak selalu berarti hubungan timbal balik.

Struktur Jaringan

- Jaringan reguler. Di mana setiap node memiliki jumlah koneksi yang sama.
- Jaringan acak. Di mana koneksi terbentuk secara acak.
- Jaringan bebas skala. Di mana beberapa node (sering disebut hub) memiliki koneksi yang jauh lebih banyak daripada yang lain.
- Jaringan bintang. Struktur jaringan di mana node pusat terhubung ke semua node lain, tetapi node perifer tidak terhubung satu sama lain. Ini membentuk pola yang menyerupai bintang.

- Jaringan jembatan. Struktur di mana dua atau lebih kelompok node dihubungkan oleh beberapa node atau tautan perantara. Node atau tautan perantara ini bertindak sebagai jembatan antara kelompok-kelompok tersebut.
- Jaringan yang terhubung sepenuhnya. Setiap node terhubung langsung ke setiap node lain dalam jaringan. Jenis jaringan ini memastikan redundansi dan toleransi kesalahan maksimum karena ada banyak jalur yang dapat dilalui data, tetapi kompleks dan mahal untuk diimplementasikan seiring pertumbuhan jaringan.

Untuk lebih menggambarkan beragam jenis struktur jaringan, Gambar 7.2 menyajikan diagram yang secara visual menggambarkan konfigurasi jaringan dasar, termasuk jaringan bintang, jembatan, dan yang terhubung sepenuhnya, sehingga memberikan pemahaman yang jelas tentang pola dan koneksi yang berbeda.



Gambar 7.2 Struktur Jaringan Dasar (Bintang, Jembatan, dan Terhubung Sepenuhnya).

Metrik: Sentralitas, Kepadatan, dan Kluster

Melampaui dasar-dasar tersebut, metrik spesifik membantu kita mengukur sifat dan kekuatan koneksi dalam suatu jaringan, yang memandu strategi pemasaran kita. Sentralitas mengukur pentingnya node individual dalam jaringan:

- Sentralitas derajat. Menghitung jumlah tepi yang dimiliki suatu node. Pengguna dengan banyak teman atau pengikut akan memiliki sentralitas derajat yang lebih tinggi.
- Sentralitas kedekatan. Mengukur seberapa dekat suatu node dengan semua node lain dalam jaringan. Ini dihitung sebagai kebalikan dari jumlah jalur terpendek dari node ke semua node lain. Sentralitas kedekatan yang tinggi menunjukkan bahwa suatu node dapat menyebarkan informasi dengan cepat ke semua node lain.
- Sentralitas antara. Mengukur seberapa sering suatu node muncul pada jalur terpendek antara node dalam jaringan. Sentralitas antara yang tinggi menunjukkan pengaruh terhadap aliran informasi.
- Sentralitas vektor eigen. Mempertimbangkan kualitas koneksi. Terhubung dengan node bersentralitas tinggi lainnya meningkatkan sentralitas vektor eigen suatu node.

Kepadatan adalah proporsi ikatan langsung dalam suatu jaringan relatif terhadap jumlah total yang mungkin. Kepadatan tinggi menunjukkan komunitas yang terjalin erat. Kluster adalah himpunan bagian dari node yang saling terhubung dalam jaringan yang lebih besar. Koefisien

pengelompokan mengukur sejauh mana tetangga dari node tertentu juga merupakan tetangga satu sama lain.

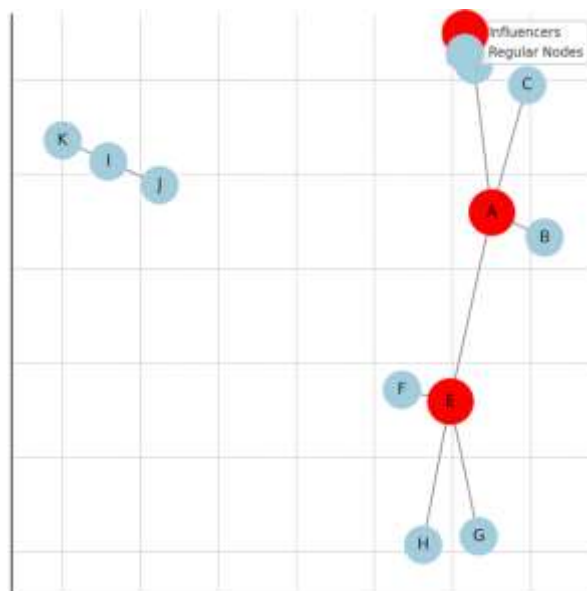
Untuk menawarkan pemahaman komprehensif tentang berbagai ukuran sentralitas dan relevansinya dalam pembahasan jaringan, Tabel 7.1 menyajikan perbandingan yang mencakup definisi singkat dan aplikasi potensial untuk setiap ukuran, menjelaskan bagaimana ukuran tersebut dapat diterapkan dalam berbagai konteks untuk mengidentifikasi node dan pola yang berpengaruh dalam suatu jaringan.

Tabel 7.1 Berbagai Ukuran Sentralitas dengan Definisi Singkat dan Aplikasi Potensial.

Ukuran	Definisi	Aplikasi
Sentralitas derajat	Mengukur jumlah sisi yang terhubung ke suatu simpul	Mengidentifikasi node populer
Sentralitas kedekatan	Mengukur panjang jalur terpendek rata-rata dari suatu simpul ke semua simpul lainnya	Mengidentifikasi node yang dapat menyebarkan informasi secara efisien
Sentralitas antara	Mengukur jumlah jalur terpendek yang melewati suatu simpul	Mengidentifikasi node yang bertindak sebagai jembatan atau penjaga gerbang
Sentralitas vektor eigen	Mempertimbangkan pentingnya tetangga suatu simpul	Mengidentifikasi node berpengaruh dalam jaringan

Strategi Identifikasi dan Keterlibatan Influencer

Dengan memanfaatkan kekuatan metrik ini, pemasar dapat menentukan influencer kunci dan merancang pendekatan keterlibatan yang ditargetkan. Mengingat luasnya jaringan sosial, mengidentifikasi pemain kunci atau influencer menjadi sangat penting. Skor sentralitas tinggi (derajat, antara, vektor eigen) sering menunjukkan node yang berpengaruh.

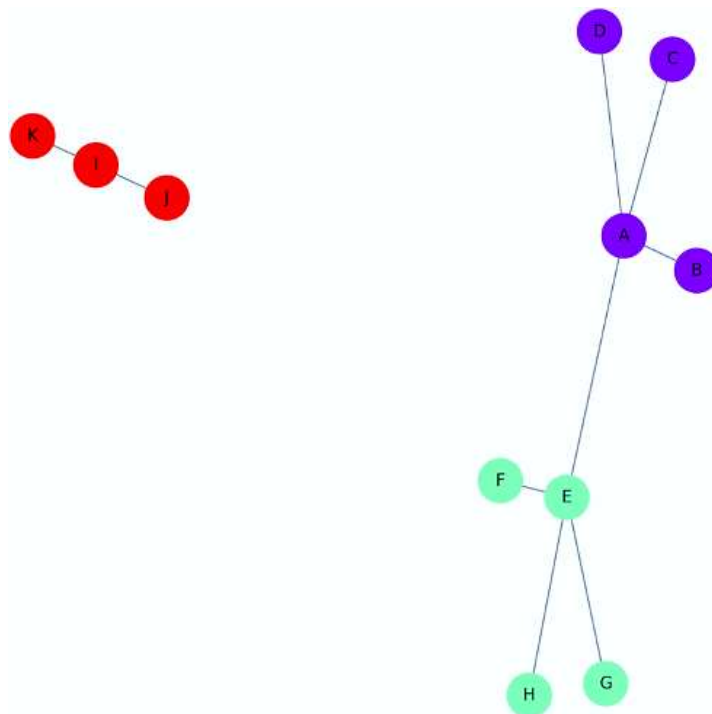


Gambar 7.3 Node Influencer (Abu-abu Gelap) Versus Node Reguler (Abu-abu Muda) dalam Hal Ukuran Sentralitasnya.

Strategi

- Keterlibatan. Terlibat langsung dengan influencer melalui kolaborasi atau kemitraan.
- Amplifikasi konten. Gunakan influencer untuk memperkuat jangkauan dan keterlibatan konten.
- Umpan balik. Influencer dapat menjadi sumber umpan balik yang berharga, mengingat hubungan mereka yang mendalam dengan komunitas.
- Perluasan jaringan. Bermitra dengan influencer dapat membantu merek menembus lebih dalam atau ke jaringan yang lebih baru.

Untuk secara visual menggarisbawahi perbedaan antara node influencer dan node biasa dalam suatu jaringan, Gambar 7.3 menyajikan grafik yang membandingkan kedua jenis node ini (diwakili dengan warna abu-abu gelap untuk influencer dan abu-abu terang untuk node biasa) dalam hal ukuran sentralitas masing-masing, menyoroti bagaimana influencer biasanya menunjukkan skor sentralitas yang lebih tinggi dalam jaringan sosial.



Gambar 7.4 Deteksi Komunitas Dalam Jaringan Sosial yang Lebih Besar.

Deteksi Komunitas dalam Jaringan Sosial

Di dalam jaringan yang luas ini, terdapat subkelompok atau komunitas. Mendeteksi kelompok-kelompok ini dapat menawarkan lapisan tambahan wawasan strategis. Komunitas adalah kelompok yang terjalin erat dalam jaringan yang lebih besar. Mengidentifikasi komunitas-komunitas ini dapat memberikan wawasan tentang perilaku pengguna, preferensi, dan segmen pemasaran potensial. Untuk menjelaskan konsep deteksi komunitas dalam jaringan sosial, Gambar 7.4 memberikan representasi visual, yang dengan jelas menggambarkan berbagai komunitas dalam jaringan yang lebih besar. Visualisasi ini

membantu dalam memahami bagaimana komunitas-komunitas ini terbentuk dan saling terhubung, menawarkan wawasan berharga untuk strategi pemasaran yang ditargetkan.

Modularitas

Modularitas adalah metrik yang mengukur kekuatan pembagian jaringan menjadi kluster. Modularitas tinggi menunjukkan koneksi padat antar node dalam kluster dan koneksi jarang antar node dalam kluster yang berbeda.

Algoritma

- Algoritma Girvan-Newman. Algoritma ini secara iteratif menghilangkan tepi yang paling sentral (betweenness tertinggi) dari jaringan untuk mendeteksi komunitas.
- Metode Louvain. Metode ini memaksimalkan modularitas untuk mendeteksi komunitas pada skala yang berbeda.

Deteksi komunitas membantu dalam kampanye pemasaran yang ditargetkan, memahami dinamika pengguna, dan menciptakan konten yang lebih beresonansi berdasarkan preferensi dan perilaku komunitas.

Konsep Kunci dalam Pembahasan Jaringan Sosial

Setelah kita menjelajahi dunia jaringan sosial, penting untuk menyaring konsep-konsep inti yang akan berperan penting dalam aplikasi dunia nyata. SNA menggunakan berbagai konsep kunci yang sangat penting untuk memahami dinamika kompleks interaksi sosial. Mari kita telusuri lebih dalam beberapa konsep ini:

- **Node dan ikatan.** Dalam SNA, node mewakili entitas dalam jaringan, seperti individu, kelompok, atau bahkan perusahaan, dan ikatan mewakili hubungan atau interaksi antara entitas-entitas ini. Sifat hubungan ini dapat bervariasi, mulai dari persahabatan, mengikuti, atau menyukai di platform media sosial hingga interaksi yang lebih kompleks seperti me-retweet atau berbagi konten.
- **Sentralitas.** Ukuran sentralitas membantu mengidentifikasi node yang paling berpengaruh dalam suatu jaringan. Ukuran ini menentukan entitas mana yang sentral atau berpengaruh berdasarkan posisinya dalam struktur jaringan. Ada beberapa ukuran sentralitas, seperti sentralitas derajat (jumlah hubungan langsung yang dimiliki suatu node), sentralitas antara (jumlah kali suatu node bertindak sebagai jembatan di sepanjang jalur terpendek antara dua node lain), dan sentralitas eigenvektor (ukuran pengaruh suatu node dalam jaringan).
- **Kepadatan.** Kepadatan adalah ukuran keterkaitan jaringan. Kepadatan dihitung sebagai proporsi dari total kemungkinan hubungan yang merupakan hubungan aktual. Jaringan dengan kepadatan tinggi menunjukkan bahwa anggota jaringan terhubung dengan baik, yang dapat memengaruhi kecepatan dan luasnya penyebaran informasi atau pengaruh.
- **Koefisien pengelompokan.** Ini adalah ukuran sejauh mana node dalam jaringan cenderung berkelompok bersama. Dalam konteks media sosial, koefisien pengelompokan yang tinggi dapat menunjukkan bahwa pengguna cenderung membentuk kelompok yang erat dan ditandai dengan tingkat interaksi yang relatif tinggi.

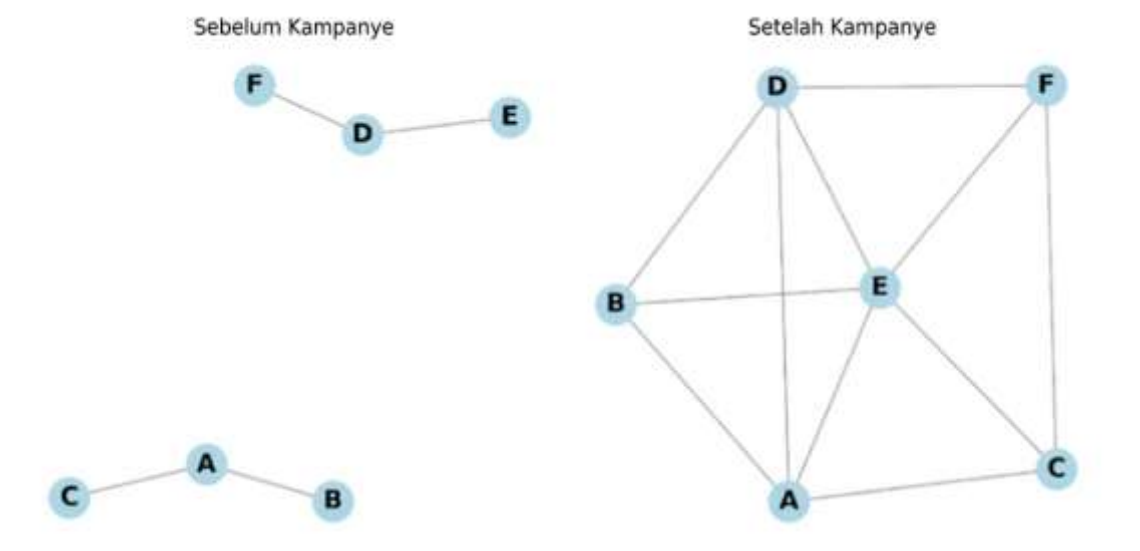
Memahami konsep-konsep ini dapat memberikan wawasan yang lebih dalam kepada para profesional pemasaran tentang perilaku audiens mereka, sehingga memungkinkan strategi pemasaran yang lebih tepat sasaran dan efektif.

Pembahasan Jaringan Sosial dalam Pemasaran Influencer

SNA telah menjadi alat penting dalam pemasaran influencer, memungkinkan pemasar untuk mengidentifikasi dan menargetkan influencer kunci di industri masing-masing. Untuk menjembatani teori dan praktik, mari kita periksa contoh nyata bagaimana SNA membantu dalam pemasaran influencer. Bayangkan sebuah merek fesyen yang ingin meluncurkan lini produk baru. Tim pemasaran memutuskan untuk memanfaatkan pemasaran influencer untuk menciptakan kehebohan untuk produk tersebut. Tetapi dengan jutaan influencer fesyen di media sosial, bagaimana mereka memutuskan dengan siapa mereka akan bermitra? Di sinilah SNA berperan. Dengan membuat grafik jaringan influencer dan pengikut mereka, tim dapat mengidentifikasi influencer kunci yang berada di pusat jaringan (sentralitas derajat tinggi), influencer yang bertindak sebagai jembatan antara komunitas yang berbeda (sentralitas antara yang tinggi), dan influencer yang pengikutnya juga berpengaruh (sentralitas eigenvektor tinggi). Misalnya, seorang influencer dengan sentralitas antara yang tinggi mungkin tidak memiliki jumlah pengikut terbesar tetapi dapat menjangkau berbagai komunitas atau demografi, menjadikannya mitra yang berharga untuk kampanye tersebut. Namun, seorang influencer dengan sentralitas derajat tinggi memiliki koneksi yang baik dan dapat menyebarkan informasi dengan cepat melalui banyak koneksi mereka.

Selain itu, dengan mengpembahasan kepadatan dan koefisien pengelompokan jaringan, tim dapat memahami seberapa erat komunitas influencer mode tersebut dan seberapa cepat informasi dapat menyebar di dalam jaringan. Untuk secara visual menunjukkan efektivitas pemasaran influencer dalam mengubah dinamika jaringan, Gambar 7.5 menyajikan grafik jaringan sebelum dan sesudah, yang menggambarkan dampak keterlibatan influencer pada konektivitas dan struktur jaringan sosial.

Pendekatan strategis berbasis data ini memungkinkan merek untuk memaksimalkan upaya pemasarannya, memastikan pesan yang tepat sampai kepada audiens yang tepat pada waktu yang tepat. Pendekatan ini juga memberikan wawasan tentang dinamika sosial komunitas influencer, yang dapat sangat berharga dalam merencanakan kampanye di masa mendatang.



Gambar 7.5 Dampak Sebelum dan Sesudah Pemasaran Influencer pada Konektivitas Jaringan.

7.3 INDIKATOR UTAMA KINERJA SITUS WEB

Mengalihkan pandangan kita dari jejaring sosial ke web yang lebih luas, kita terjun ke ranah pembelajaran web, mempersenjatai diri dengan alat dan metrik yang menguraikan interaksi pengguna di situs web.

Gambaran Umum Alat dan Metrik Pembelajaran Web

Pembelajaran web, pada dasarnya, adalah pengukuran, pengumpulan, pembahasan, dan pelaporan data web untuk tujuan memahami dan mengoptimalkan penggunaan web. Ini menyediakan alat dan metrik yang memungkinkan pemasar untuk mengukur efektivitas strategi pemasaran mereka, mengidentifikasi potensi peningkatan, dan membuat keputusan yang tepat tentang kehadiran online mereka. Alat pembelajaran web dapat diklasifikasikan secara luas menjadi dua kategori:

- Pembelajaran web di situs. Alat-alat ini mengukur perilaku pengunjung setelah mereka berada di situs web Anda. Ini termasuk metrik seperti tampilan halaman, pengunjung unik, rasio pentalan, rasio konversi, dan waktu rata-rata yang dihabiskan di situs. Metrik ini memberikan wawasan tentang seberapa baik kinerja situs web Anda dan bagaimana pengguna berinteraksi dengannya.
- Pembahasan web di luar situs. Alat-alat ini mengukur potensi audiens situs web Anda (peluang), pangsa suara (visibilitas), dan perbincangan (komentar) yang terjadi di internet secara keseluruhan.

Beberapa alat pembahasan web populer termasuk Google Analytics, Adobe Analytics, dan IBM Digital Analytics. Alat-alat ini menyediakan banyak data yang dapat dipembahasan untuk mendapatkan wawasan tentang perilaku pengguna, kinerja situs web, dan efektivitas pemasaran.

Untuk memberikan perspektif yang jelas tentang lanskap persaingan alat pembahasan web, Gambar 7.6 menyajikan diagram lingkaran yang menggambarkan perkiraan pangsa pasar

alat pembahasan web populer. Misalnya, Google Analytics menyediakan data tentang dari mana pengunjung Anda berasal (perujuk), halaman mana yang paling populer, berapa lama pengunjung tinggal di situs Anda, dan kata kunci apa yang mengarahkan orang ke situs Anda. Ini juga menyediakan informasi demografis tentang pengunjung Anda, seperti lokasi, usia, dan minat mereka, yang dapat sangat berguna untuk kampanye pemasaran yang ditargetkan.

Metrik Utama: Tampilan Halaman, Rasio Pentalan, dan Rasio Konversi

Tampilan Halaman

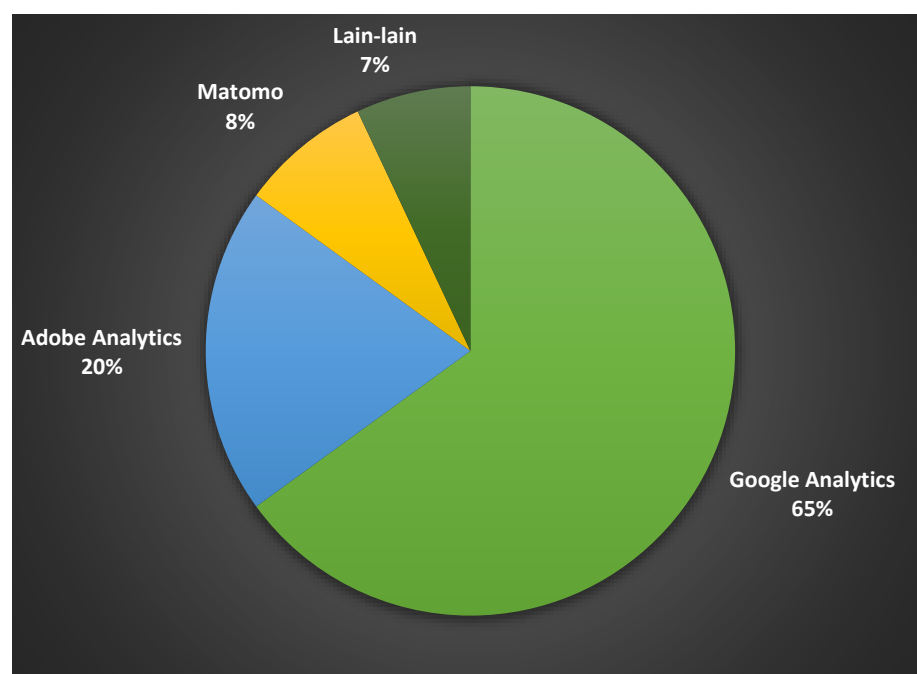
Tampilan halaman mewakili satu kali pengguna melihat halaman web. Ini menunjukkan jumlah total berapa kali halaman telah dimuat atau dimuat ulang. Meskipun memberikan ukuran dasar aktivitas halaman web, ini tidak membedakan antara pengguna unik atau kualitas interaksi. Tampilan halaman memungkinkan pemantauan tren lalu lintas, pemahaman konten populer, dan pelacakan efektivitas iklan.

Rasio Pentalan

Rasio pentalan adalah persentase sesi di mana pengguna memuat situs web dan kemudian pergi tanpa melakukan tindakan lain, seperti mengklik tautan atau mengisi formulir. Rasio pentalan yang tinggi mungkin menunjukkan konten yang tidak menarik atau bahwa pengguna tidak menemukan apa yang mereka cari. Rasio pentalan memungkinkan evaluasi efektivitas halaman arahan, mengukur keterlibatan pengguna, dan mengidentifikasi potensi masalah situs.

Rasio Konversi

Ini adalah persentase pengunjung situs web yang melakukan tindakan yang diinginkan, seperti melakukan pembelian, mendaftar untuk buletin, atau mengisi formulir kontak. Ini adalah metrik penting untuk mengukur efektivitas situs web dalam membimbing pengguna menuju tujuannya. Rasio konversi memungkinkan penilaian efektivitas saluran penjualan, pengoptimalan halaman arahan, dan pengukuran keberhasilan kampanye pemasaran.



Gambar 7.6 Perkiraan Pangsa Pasar Alat Pembahasan Web Populer.

Untuk menawarkan pemahaman komprehensif tentang berbagai metrik yang digunakan dalam pembahasan web, Tabel 7.2 mencantumkan metrik utama seperti tampilan halaman, rasio pentalan, dan rasio konversi, disertai dengan deskripsi singkat dan signifikansinya dalam mengukur dan mengoptimalkan kinerja situs web.

Tabel 7.2 Metrik Utama dengan Deskripsi Singkat dan Pentingnya.

Metrik	Deskripsi	Signifikansi
Tayangan halaman	Jumlah total halaman yang dilihat	Menilai popularitas konten
Sesi	Kelompok interaksi dalam jangka waktu tertentu	Mengukur keterlibatan situs
Pengguna	Individu yang telah mengunjungi situs setidaknya sekali	Memperkirakan ukuran audiens dan potensi jangkauan
Rasio pentalan	Persentase kunjungan satu halaman	Mengevaluasi keterlibatan pengguna dan relevansi konten
Durasi sesi rata-rata	Rata-rata waktu yang dihabiskan pengguna di situs per sesi	Mengpembahasan kedalaman keterlibatan pengguna
Rasio konversi	Persentase kunjungan yang menghasilkan konversi	Mengukur efektivitas dalam mendorong tindakan yang diinginkan
Rasio keluar	Persentase keluar dari halaman	Mengidentifikasi area masalah potensial dalam perjalanan pengguna
Nilai halaman	Nilai rata-rata halaman yang dikunjungi	Menentukan kontribusi halaman terhadap pendapatan situs
Rasio klik-tayang	Persentase pengguna yang mengklik tautan tertentu	Mengevaluasi efektivitas iklan online atau ajakan bertindak

Metrik Lanjutan: Pembahasan Corong dan Pembahasan Kohort

Di luar metrik dasar terdapat pembahasan yang lebih bernuansa, yang menawarkan pandangan terperinci tentang perilaku pengguna melalui metode pembelajaran tingkat lanjut.

Pembahasan Corong

Corong mewakili perjalanan yang ditempuh pengguna untuk menyelesaikan tindakan tertentu di situs web, seperti melakukan pembelian. Pembahasan corong memecah perjalanan ini menjadi langkah-langkah individual dan menunjukkan di mana pengguna berhenti di setiap tahap. Dengan mengidentifikasi titik-titik berhenti ini, pemasar dapat mengoptimalkan perjalanan pengguna untuk meningkatkan konversi.

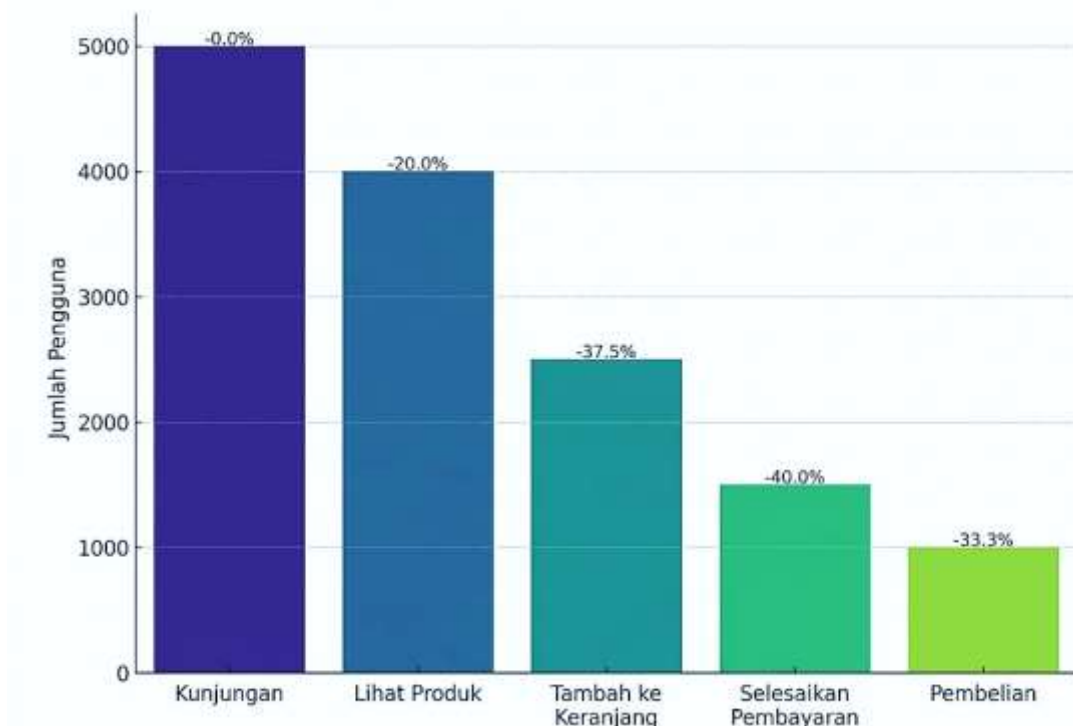
Strategi optimasi untuk pembahasan corong:

- Segmentasi. Bagi pengguna ke dalam segmen yang berbeda berdasarkan kriteria seperti sumber lalu lintas, jenis perangkat, atau informasi demografis. Hal ini

memungkinkan bisnis untuk mengidentifikasi segmen tertentu yang mungkin menghadapi masalah dan menargetkan optimasi yang sesuai.

- Pengujian A/B. Jika terjadi penurunan pengguna pada tahap tertentu, uji variasi langkah tersebut untuk melihat versi mana yang mempertahankan lebih banyak pengguna. Ini bisa berupa ajakan bertindak (CTA) yang berbeda, desain halaman, atau konten.
- Umpan balik pengguna. Gabungkan data funnel dengan umpan balik pengguna untuk memahami mengapa pengguna mungkin berhenti menggunakan situs. Survei atau permintaan umpan balik singkat dapat memberikan konteks pada data yang diamati.
- Optimasi kinerja. Halaman yang lambat dimuat atau gangguan teknis dapat menjadi alasan pengguna berhenti menggunakan situs. Memastikan halaman dimuat dengan cepat dan bebas dari kesalahan dapat membantu kelancaran progres pengguna melalui funnel.
- Aplikasi. Selain mengoptimalkan proses checkout, meningkatkan alur onboarding untuk produk perangkat lunak, dan menyempurnakan funnel pembuatan prospek, pembahasan funnel dapat memandu strategi konten, membentuk desain UX, dan banyak lagi.

Untuk secara visual menunjukkan efektivitas pembahasan funnel dalam mengidentifikasi area yang perlu ditingkatkan, Gambar 7.7 menyajikan bagan batang yang menunjukkan alur pengguna dan tingkat berhenti menggunakan situs di setiap tahap funnel situs web, memberikan wawasan yang jelas tentang di mana potensi optimasi dapat dilakukan.



Gambar 7.7 Alur Pengguna dan Tingkat Penurunan Pengguna pada Tahap yang Berbeda.

Pembahasan Kohort

Pembahasan kohort adalah jenis pembelajaran selang waktu yang membagi basis pengguna ke dalam kelompok-kelompok terkait dari waktu ke waktu. Kelompok-kelompok ini, atau kohort, memiliki karakteristik atau pengalaman umum dalam periode yang ditentukan. Dengan mempelajari bagaimana kohort tertentu berperilaku dari waktu ke waktu, bisnis dapat memperoleh wawasan yang lebih dalam tentang pola dan tren siklus hidup.

Aspek-aspek utama pembahasan kohort:

- **Kohort berbasis waktu.** Jenis kohort ini mengelompokkan pengguna berdasarkan kerangka waktu tertentu, seperti kohort bulanan atau mingguan pengguna baru. Ini dapat membantu dalam memahami bagaimana perubahan produk atau kampanye pemasaran tertentu memengaruhi perilaku pengguna dari waktu ke waktu.
- **Kohort berbasis perilaku.** Kohort ini mengelompokkan pengguna berdasarkan tindakan spesifik yang mereka lakukan, seperti melakukan pembelian pertama, meningkatkan produk, atau berinteraksi dengan fitur tertentu.
- **Ukuran dan durasi.** Kohort dapat bervariasi dalam ukuran dan durasi. Misalnya, kohort harian mungkin lebih kecil dan lebih spesifik, sedangkan kohort bulanan atau triwulanan menawarkan pandangan yang lebih luas.

Strategi optimasi untuk pembahasan kohort:

- **Pemasaran yang disesuaikan.** Dengan memahami perilaku kohort tertentu, bisnis dapat merancang kampanye pemasaran yang sesuai dengan kebutuhan dan preferensi unik setiap kelompok.
- **Pengembangan produk.** Pengembangan produk mengidentifikasi fitur atau aspek produk yang paling sesuai dengan kohort tertentu, memandu pengembangan atau penyempurnaan di masa mendatang.
- **Strategi retensi.** Strategi retensi mengenali pola kapan dan mengapa kohort tertentu berhenti menggunakan produk atau layanan, memungkinkan bisnis untuk mengatasi masalah ini dan meningkatkan retensi.
- **Prediksi siklus hidup.** Prediksi siklus hidup membantu memahami siklus hidup rata-rata pengguna atau pelanggan, meningkatkan peramalan dan perencanaan jangka panjang.
- **Aplikasi.** Pembahasan kohort melampaui pelacakan keberhasilan kampanye pemasaran. Ini juga sangat penting dalam memahami loyalitas pelanggan, memprediksi churn, mengevaluasi perubahan produk, dan membentuk strategi bisnis jangka panjang.

Integrasi dengan Sumber Data Lain

Untuk mencapai pemahaman holistik tentang perilaku pengguna, seringkali bermanfaat untuk mengintegrasikan data analitik web dengan sumber data lain.

Sistem Manajemen Hubungan Pelanggan

Dengan mengintegrasikan sistem manajemen hubungan pelanggan (CRM), bisnis dapat menjembatani kesenjangan antara perilaku online dan penjualan aktual atau interaksi

pelanggan. Integrasi ini memungkinkan segmentasi pengguna yang lebih detail dan kampanye pemasaran yang dipersonalisasi.

Analitik Media Sosial

Menggabungkan analitik web dengan metrik media sosial dapat memberikan wawasan tentang bagaimana interaksi sosial mendorong perilaku situs web. Misalnya, hal ini dapat menunjukkan bagaimana tweet viral memengaruhi lalu lintas situs web atau konversi.

Data Penjualan Offline

Bagi bisnis dengan saluran penjualan online dan offline, mengintegrasikan data penjualan offline dapat memberikan pandangan komprehensif tentang perjalanan pelanggan. Hal ini dapat menunjukkan bagaimana kampanye pemasaran online memengaruhi penjualan offline atau sebaliknya.

Survei dan Formulir Umpan Balik

Penggabungan data kualitatif dari survei dengan analitik web kuantitatif dapat memberikan konteks pada perilaku pengguna. Misalnya, jika pengguna meninggalkan situs web pada tahap tertentu, data survei dapat mengungkapkan mengapa mereka melakukannya.

Untuk mengilustrasikan bagaimana data analitik web dapat diintegrasikan secara efektif dengan sistem CRM untuk pemahaman yang lebih komprehensif tentang perilaku pengguna, Tabel 7.3 memberikan contoh yang menggabungkan metrik seperti sesi dan tampilan halaman dengan informasi spesifik pelanggan seperti total pembelian dan nilai seumur hidup.

Tabel 7.3 Contoh yang Menunjukkan Integrasi Data Analitik Web dengan Sistem CRM.

ID Pengguna	Sesi	Tampilan Halaman	Konversi	Nama	Total Pembelian	Nilai Seumur Hidup (CLV)
1	5	25	2	Alice	5	300
2	3	15	1	Bob	3	200
...	...	...	...	...	...	...

Komplikasi dan Jebakan dalam Mengintegrasikan Sumber Data

Meskipun mengintegrasikan berbagai sumber data menjanjikan perspektif yang lebih kaya dan komprehensif, hal ini tidak lepas dari tantangan. Sangat penting untuk menyadari potensi jebakan ini untuk memastikan bahwa integrasi data menghasilkan wawasan yang dapat ditindaklanjuti tanpa mengorbankan akurasi atau privasi.

- **Masalah konsistensi data.** Sistem yang berbeda seringkali memiliki cara yang berbeda dalam menangkap dan menyimpan data. Ketidaksesuaian dalam format data, satuan, atau zona waktu dapat menyebabkan interpretasi yang menyimpang.
- **Kekhawatiran privasi.** Mengintegrasikan data pribadi dari berbagai sumber dapat melanggar peraturan privasi seperti GDPR atau CCPA. Bisnis harus berhati-hati tentang bagaimana mereka menangani, menyimpan, dan memproses data terintegrasi.

- **Tumpang tindih data.** Jika data yang sama ditangkap dalam beberapa sistem, ada risiko penghitungan ganda. Ini dapat secara artifisial meningkatkan metrik dan menyebabkan keputusan yang salah arah.
- **Kompleksitas integrasi sistem.** Tidak semua sistem dirancang untuk terintegrasi dengan mulus. Integrasi mungkin memerlukan solusi khusus atau middleware, yang menambah biaya dan potensi titik kegagalan.
- **Sinkronisasi data tidak lengkap.** Dalam integrasi waktu nyata, ada risiko data tidak tersinkronisasi dengan benar atau lengkap antar sistem, yang menyebabkan kesenjangan data.
- **Kesalahpahaman.** Lebih banyak data tidak selalu berarti wawasan yang lebih baik. Tanpa konteks atau pemahaman yang tepat, data yang terintegrasi dapat menyebabkan kesimpulan yang salah.
- **Peningkatan biaya pengelolaan data.** Menggabungkan data dari berbagai sumber dapat secara eksponensial meningkatkan jumlah data yang harus dikelola, disimpan, dan diproses, sehingga memerlukan sistem yang lebih kuat dan mungkin menimbulkan biaya yang lebih tinggi.

Meskipun menggabungkan beberapa sumber data menawarkan bisnis pemahaman yang lebih mendalam tentang pengguna mereka dan meningkatkan efektivitas strategi digital mereka, penting untuk menavigasi kompleksitas dan jebakan integrasi. Dengan demikian, organisasi dapat memastikan bahwa pengambilan keputusan mereka tidak hanya lebih informatif tetapi juga akurat dan sesuai dengan praktik terbaik dan peraturan.

Konsep Kunci dalam Pembahasan Web

Ada beberapa konsep dan metrik kunci yang penting untuk memahami data yang disediakan oleh alat pembahasan web. Berikut adalah gambaran umum beberapa konsep penting ini:

- **Tampilan halaman.** Ini adalah jumlah total halaman yang dilihat oleh semua pengunjung. Tampilan berulang dari satu halaman juga dihitung.
- **Pengunjung unik.** Ini adalah individu yang telah mengunjungi situs web setidaknya sekali dalam periode waktu tertentu.
- **Rasio pentalan.** Ini adalah persentase pengunjung yang meninggalkan situs setelah hanya melihat satu halaman.
- **Rasio keluar.** Ini adalah persentase pengunjung yang meninggalkan situs Anda dari halaman tertentu berdasarkan jumlah kunjungan ke halaman tersebut.
- **Durasi sesi rata-rata.** Ini adalah rata-rata lama waktu yang dihabiskan pengunjung di situs Anda selama satu kunjungan.
- **Rasio konversi.** Ini adalah persentase pengunjung yang menyelesaikan tindakan yang diinginkan di situs web, seperti mengisi formulir, mendaftar untuk buletin, atau melakukan pembelian.
- **Jalur klik.** Ini adalah urutan halaman yang dilihat oleh pengunjung selama kunjungan situs web.

- **Sumber lalu lintas.** Ini menunjukkan dari mana pengunjung Anda berasal — kunjungan langsung, mesin pencari, atau rujukan dari situs web lain.
- **Biaya per klik.** Ini adalah jumlah uang yang dibayarkan pengiklan kepada penerbit untuk setiap klik dalam kampanye iklan bayar per klik.
- **Pengembalian investasi (ROI).** Ini adalah ukuran profitabilitas suatu investasi. Dalam konteks analitik web, ini mengacu pada pendapatan yang dihasilkan dari kampanye pemasaran digital dibandingkan dengan biaya kampanye tersebut.

Memahami konsep-konsep ini dapat membantu pemasar membuat keputusan yang tepat tentang strategi online mereka, mengidentifikasi area peningkatan, dan memaksimalkan pengembalian investasi pemasaran mereka.

Tantangan dalam Menginterpretasikan Data Analitik Web

Meskipun analitik web dapat sangat mencerahkan, ia tidak lepas dari kompleksitasnya. Seringkali, bisnis mendapati diri mereka terjebak dalam jaringan kesalahpahaman atau kesalahan saat mencoba untuk mendapatkan makna dari data. Mari kita uraikan beberapa tantangan paling umum yang dihadapi dan berikan solusi untuk mengatasinya:

- **Kelebihan data.** Dengan banyaknya metrik yang tersedia, pemasar dapat merasa kewalahan, tidak yakin metrik mana yang paling sesuai dengan tujuan bisnis mereka. Kuncinya adalah fokus pada KPI yang secara langsung terkait dengan tujuan, daripada mencoba pembahasan semuanya.
- **Kesalahpahaman metrik.** Tanpa pemahaman yang jelas, metrik seperti rasio pentalan atau waktu di situs dapat menyesatkan. Misalnya, rasio pentalan yang tinggi mungkin tidak selalu menunjukkan ketidakpuasan; itu bisa berarti pengguna menemukan apa yang mereka butuhkan dengan cepat.
- **Data yang tidak konsisten di berbagai platform.** Alat analitik yang berbeda mungkin menghasilkan hasil yang sedikit berbeda karena metodologi pelacakan yang unik. Sangat penting untuk memahami nuansa setiap alat dan, jika memungkinkan, mengandalkan sumber utama untuk pengambilan keputusan.
- **Tidak memperhitungkan faktor eksternal.** Tren musiman, kampanye pemasaran offline, atau bahkan peristiwa global dapat memengaruhi lalu lintas web dan perilaku pengguna. Pemasar perlu mengkontekstualisasikan data mereka dalam gambaran yang lebih besar.
- **Ketergantungan pada metrik kesombongan.** Meskipun metrik seperti tampilan halaman atau total pengikut dapat meningkatkan ego, metrik tersebut mungkin tidak berkontribusi pada tujuan bisnis. Pemasar harus membedakan antara metrik kesombongan dan metrik yang dapat ditindaklanjuti.
- **Kurangnya tujuan yang jelas.** Tanpa tujuan yang ditetapkan untuk situs web atau kampanye, interpretasi data menjadi tanpa arah. Ini seperti memiliki peta tanpa tujuan. Pastikan tujuan didefinisikan, terukur, dan selaras dengan sasaran bisnis.

Contoh Praktis: Menggunakan Google Analytics untuk Wawasan Perilaku Pelanggan

Google Analytics adalah alat analitik web yang banyak digunakan yang menyediakan data mendalam tentang pengguna situs web, memungkinkan bisnis untuk lebih memahami

perilaku pelanggan. Mari kita jelajahi ranah praktis analitik web dengan pembahasan bagaimana Google Analytics menjelaskan pola perilaku pelanggan.

Mari kita pertimbangkan contoh hipotetis dari pengecer fesyen online. Tujuan utama pengecer adalah untuk meningkatkan penjualan dan mengurangi jumlah pembatalan keranjang belanja. Dengan menggunakan Google Analytics, tim pemasaran dapat melacak berbagai metrik seperti tampilan halaman, rasio pentalan, durasi sesi rata-rata, dan rasio konversi. Mereka juga dapat mengidentifikasi halaman dengan tingkat keluar tertinggi, yang berpotensi menunjukkan masalah dengan halaman-halaman ini yang mendorong pelanggan untuk pergi.

Misalnya, tim menemukan bahwa banyak pengguna berhenti di halaman pembayaran. Setelah menyelidiki lebih dalam, mereka menemukan bahwa biaya pengiriman, yang hanya terungkap pada langkah terakhir, menyebabkan pembatalan ini. Berdasarkan wawasan ini, pengecer mungkin memutuskan untuk menawarkan pengiriman gratis untuk pesanan di atas jumlah tertentu atau lebih transparan tentang biaya pengiriman lebih awal dalam proses untuk mengurangi kejutan saat pembayaran.

Selain itu, peritel dapat menggunakan Google Analytics untuk mensegmentasikan audiensnya dan memahami perilaku pengguna yang berbeda. Misalnya, mereka mungkin menemukan bahwa pengguna seluler memiliki tingkat pentalan (bounce rate) yang lebih tinggi dibandingkan pengguna desktop. Hal ini dapat mengindikasikan masalah dengan versi seluler situs web, yang mendorong peninjauan UX seluler.

Mereka juga dapat melacak efektivitas kampanye pemasaran mereka dengan memantau sumber lalu lintas. Misalkan sebagian besar lalu lintas mereka berasal dari pencarian organik, tetapi kampanye media sosial berbayar menghasilkan konversi terbanyak. Dalam hal ini, peritel mungkin memutuskan untuk mengalokasikan lebih banyak anggaran untuk iklan media sosial.

Singkatnya, Google Analytics dapat memberikan banyak wawasan tentang perilaku pelanggan, yang memandu pengambilan keputusan dan strategi peritel. Hal ini memungkinkan peritel untuk memfokuskan upaya mereka pada area yang kemungkinan akan menghasilkan manfaat paling signifikan, meningkatkan efektivitas keseluruhan upaya pemasaran mereka.

7.4 MENDENGARKAN DAN MELACAK MEDIA SOSIAL

Mendengarkan sama pentingnya dengan berbicara di dunia digital. Dengan mengikuti percakapan di media sosial, pemasar dapat memperoleh banyak informasi.

Gambaran Umum Mendengarkan dan Melacak Media Sosial

Mendengarkan media sosial, juga dikenal sebagai pemantauan media sosial, melibatkan pelacakan penyebutan merek Anda, pesaing, produk, atau kata kunci yang relevan di seluruh platform media sosial dan web. Proses ini memungkinkan perusahaan untuk mendapatkan wawasan tentang opini pelanggan, tren yang muncul, dan persepsi keseluruhan merek mereka.

Salah satu aspek utama dari mendengarkan media sosial adalah pembahasan sentimen, yang melibatkan interpretasi dan pengkategorian emosi yang diungkapkan dalam postingan media sosial. Ini dapat membantu perusahaan memahami sentimen umum terhadap merek mereka atau produk tertentu dan mengidentifikasi potensi masalah sejak dini.

Namun, pelacakan media sosial melibatkan pengukuran kinerja kampanye dan konten media sosial Anda. Indikator kinerja utama (KPI) mungkin termasuk suka, berbagi, komentar, rasio klik-tayang (CTR), dan jangkauan keseluruhan postingan Anda. Memantau metrik ini dari waktu ke waktu dapat membantu perusahaan memahami jenis konten apa yang paling sesuai dengan audiens mereka dan memberikan informasi untuk strategi konten di masa mendatang (lihat Tabel 7.4).

Tabel 7.4 Gambaran Perbandingan Alat Pemantauan dan Pelacakan Media Sosial Populer.

Alat	Keunggulan	Kelemahan	Kasus Penggunaan
Brandwatch	Menawarkan pembahasan berbasis AI yang canggih, pelacakan data secara real-time, dan kumpulan data yang luas; kueri yang sangat dapat disesuaikan memungkinkan pencarian yang sangat spesifik	Mungkin terlalu rumit bagi pemula karena fitur-fiturnya yang ekstensif	Paling cocok untuk perusahaan besar atau industri yang membutuhkan pembahasan mendalam, perbandingan kinerja pesaing, dan prediksi tren
Hootsuite Insights	Antarmuka yang ramah pengguna, integrasi dengan berbagai platform media sosial, dan menyediakan data secara real-time	Kemampuan pembahasan mendalam yang lebih terbatas dibandingkan dengan alat khusus	Cocok untuk bisnis kecil hingga menengah atau agensi yang mengelola banyak akun media sosial dan menginginkan platform gabungan untuk posting dan pembahasan
Sprout Social	Dasbor analitik komprehensif, fitur CRM bawaan, dan alat kolaborasi tim	Lebih mahal untuk bisnis kecil dan memiliki kurva pembelajaran yang lebih curam	Ideal untuk bisnis yang berfokus pada manajemen komunitas, layanan pelanggan, dan kolaborasi tim bersamaan dengan pemantauan media sosial

Sebutkan	Pemantauan real-time, pembahasan kompetitif, dan integrasi dengan influencer, sehingga memudahkan untuk menjangkau dan berkolaborasi	Beberapa keterbatasan dalam akses data historis dan kedalaman pembahasan sentimen	Sangat cocok untuk bisnis yang menginginkan gambaran cepat tentang penyebutan merek mereka, perbandingan kinerja pesaing, dan jangkauan influencer

Pada intinya, pemantauan dan pelacakan media sosial merupakan aspek penting dari kehadiran online suatu perusahaan. Hal ini memberikan wawasan berharga tentang preferensi dan perilaku pelanggan, memungkinkan perusahaan untuk mengoptimalkan strategi pemasaran mereka, meningkatkan layanan pelanggan, dan membangun hubungan yang lebih kuat dengan pelanggan mereka.

Pentingnya Pemantauan Media Sosial dalam Pemasaran Modern

Di era digital yang sangat terhubung saat ini, di mana percakapan dan interaksi konsumen terus berlangsung di berbagai platform, memahami dialog-dialog ini menjadi sangat penting. Berikut alasan mengapa pemantauan media sosial sangat penting untuk pemasaran modern:

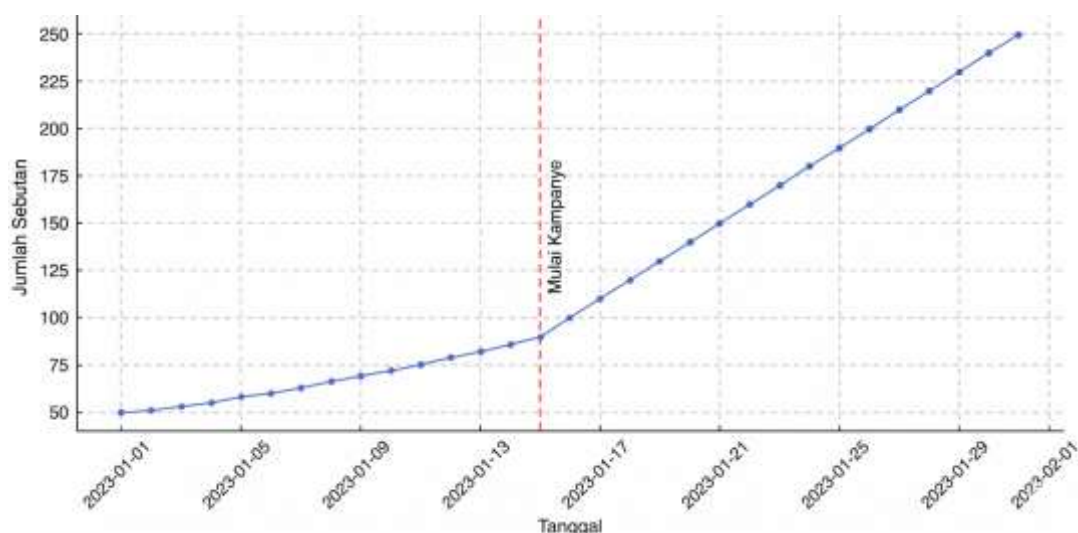
- **Umpan balik waktu nyata.** Pemantauan media sosial menawarkan umpan balik langsung. Baik itu peluncuran produk, kampanye pemasaran, atau aktivitas merek apa pun, pemasar dapat mengukur reaksi waktu nyata dan menyesuaikan strategi sesuai dengan itu.
- **Manajemen reputasi.** Komentar atau ulasan negatif dapat menyebar dengan cepat di ranah digital. Pemantauan media sosial memungkinkan merek untuk memantau sentimen, mendeteksi potensi krisis PR sejak dini, dan menyusun strategi pengendalian kerusakan atau mekanisme respons dengan cepat.
- **Memahami audiens.** Selain memantau penyebutan merek, alat mendengarkan media sosial dapat memberikan wawasan yang lebih dalam tentang apa yang dibicarakan audiens target dan masalah, keinginan, atau tren yang mereka ikuti, yang mengarah pada pembuatan profil dan segmentasi audiens yang lebih baik.
- **Pembahasan pesaing.** Pemasar dapat menggunakan mendengarkan media sosial untuk mengawasi pesaing. Memahami bagaimana pesaing dipersepsikan, kampanye sukses mereka, atau area di mana mereka gagal dapat menawarkan wawasan strategis yang sangat berharga.
- **Inovasi dan pengembangan produk.** Umpan balik waktu nyata dari pengguna sebenarnya memberikan informasi berharga untuk pengembangan produk. Merek dapat mengidentifikasi fitur yang diinginkan pengguna, masalah yang mereka hadapi, dan sesuai dengan itu melakukan iterasi pada penawaran mereka.
- **Keterlibatan dan pembangunan komunitas.** Mendengarkan media sosial bukan hanya tentang pemantauan; ini tentang keterlibatan. Dengan menanggapi penyebutan

pengguna—baik itu pertanyaan, keluhan, atau pujian—merek dapat membina komunitas, loyalitas, dan meningkatkan citra merek mereka.

- **Keputusan berbasis data.** Dengan data konkret tentang bagaimana kampanye atau jenis konten tertentu berkinerja di antara audiens, merek dapat membuat keputusan yang lebih tepat tentang di mana harus menginvestasikan sumber daya mereka dan bagaimana menyusun strategi inisiatif pemasaran masa depan mereka.
- **Menemukan influencer dan pendukung merek.** Seringkali, alat social listening dapat membantu mengidentifikasi tidak hanya volume penyebutan tetapi juga sumber yang memiliki pengaruh besar di komunitas mereka. Bermitra dengan influencer tersebut dapat memperkuat jangkauan dan kredibilitas merek.
- **Memanfaatkan peluang pemasaran secara real-time.** Memanfaatkan diskusi dan tren terkini dapat membuat merek menjadi bagian dari percakapan yang lebih besar. Misalnya, jika topik yang relevan dengan merek mulai menjadi tren, mereka dapat ikut serta dalam dialog tersebut, membuat upaya pemasaran mereka lebih kontemporer dan relevan.
- **Memperbaiki strategi SEO dan konten.** Memahami bahasa, frasa, dan topik yang dibahas audiens dapat menginformasikan strategi SEO dan pembuatan konten, memastikan bahwa merek beresonansi dengan apa yang secara aktif dicari audiens.

Untuk mencontohkan efektivitas social listening, khususnya dalam melacak pengaruh upaya pemasaran, Gambar 7.8 menampilkan grafik yang menunjukkan peningkatan penyebutan merek dari waktu ke waktu sebagai hasil dari kampanye tertentu, menyoroti bagaimana inisiatif strategis dapat secara signifikan memperkuat kehadiran merek dalam percakapan digital.

Pada intinya, social listening bukan hanya alat pemantauan pasif; Ini adalah penggerak strategi aktif. Ini menjembatani kesenjangan antara merek dan audiens mereka, memungkinkan pendekatan yang lebih otentik, responsif, dan berbasis data untuk pemasaran modern.



Gambar 7.8 Peningkatan Penyebutan Merek dari Waktu ke Waktu Karena Kampanye Tertentu.

Pembahasan Penyebutan di Media Sosial dan Pangsa Suara

Pembahasan Penyebutan di Media Sosial

Setiap hari, konsumen menggunakan platform media sosial untuk membahas merek, berbagi pengalaman, dan memberikan umpan balik. Diskusi ini, yang dikenal sebagai penyebutan di media sosial, merupakan sumber data yang sangat berharga bagi merek untuk memahami persepsi publik dan menyesuaikan strategi.

Saat pembahasan penyebutan di media sosial, sangat penting untuk menggali lebih dalam daripada sekadar volume penyebutan. Meskipun menggembirakan melihat jumlah penyebutan merek yang tinggi, memahami konteksnya sangat penting. Penyebutan dapat bersifat positif, negatif, atau netral. Merek harus mengkategorikan penyebutan ini berdasarkan sentimen untuk mengukur kesehatan merek secara keseluruhan.

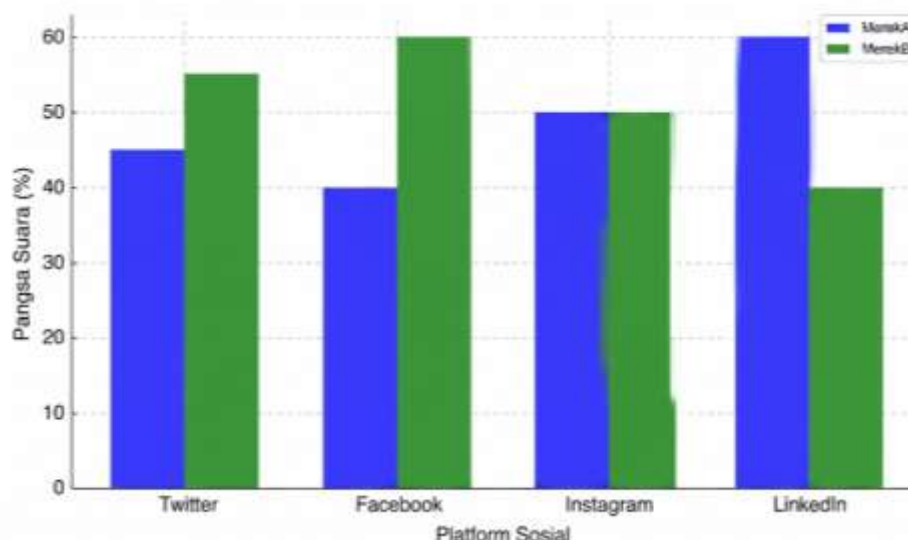
Misalnya, jika produk baru diluncurkan dan terjadi lonjakan penyebutan negatif, merek dapat dengan cepat mengidentifikasi potensi masalah dengan produk tersebut dan memperbaikinya. Sebaliknya, lonjakan penyebutan positif dapat menunjukkan kampanye yang sukses atau produk yang diterima dengan baik, yang memandu upaya di masa mendatang.

Pangsa Suara

Pangsa suara (Share of Voice/SOV) adalah metrik yang menunjukkan kehadiran suatu merek dalam percakapan industrinya. Sederhananya, ini menjawab pertanyaan, “Dari semua obrolan di industri kita, berapa banyak yang membahas tentang kita?”

SOV ditentukan dengan membandingkan jumlah penyebutan yang diterima suatu merek relatif terhadap total penyebutan industrinya atau pesaing utamanya. Pembahasan SOV membantu merek memahami posisi mereka di pasar (lihat Gambar 7.9).

Jika SOV suatu merek meningkat, ini menunjukkan peningkatan kesadaran merek dan posisi yang semakin kuat dalam percakapan pasar. Sebaliknya, penurunan SOV dapat menandakan bahwa pesaing menjadi lebih dominan dalam dialog industri, dan penilaian ulang strategi pemasaran mungkin diperlukan.



Gambar 7.9 Membandingkan SOV untuk Dua Merek yang Bersaing di Berbagai Platform Media Sosial.

Manajemen Krisis Melalui Social Listening

Di era digital saat ini, ketika berita menyebar dengan kecepatan kilat, mengelola reputasi merek selama krisis menjadi tantangan utama. Di sinilah social listening memainkan peran penting.

Deteksi Dini

Sebelum masalah kecil berubah menjadi krisis besar, biasanya ada tanda-tanda peringatan di platform media sosial. Lonjakan mendadak dalam penyebutan atau meningkatnya jumlah sentimen negatif dapat menjadi indikator awal. Dengan memantau metrik ini secara real-time, merek berpotensi mengidentifikasi dan mengatasi masalah sebelum masalah tersebut menjadi tidak terkendali.

Memahami Sifat Krisis

Setelah masalah terdeteksi, social listening dapat membantu merek memahami akar penyebabnya. Apakah itu cacat produk? Mungkin iklan yang tidak pantas? Atau mungkin pernyataan yang di buat oleh perwakilan perusahaan? Menentukan sifat pasti dari masalah tersebut adalah langkah pertama dalam merancang respons yang tepat.

Keterlibatan dan Komunikasi

Selama krisis, komunikasi adalah kunci. Platform media sosial menjadi saluran utama tempat konsumen mencari respons dan pembaruan dari merek. Dengan mendengarkan secara aktif, merek dapat memilih kapan, di mana, dan bagaimana merespons. Mengatasi kekhawatiran, menawarkan solusi, dan berkomunikasi secara transparan dapat membantu mengurangi kerusakan dan membangun kembali kepercayaan.

Pembahasan Pasca-Krisis

Setelah badai berlalu, mendengarkan media sosial tetap sangat berharga. Merek dapat pembahasan diskusi untuk memahami efektivitas strategi respons mereka. Apakah sentimen konsumen membaik setelah intervensi merek? Apakah ada kekhawatiran yang masih ada yang perlu ditangani?

Pada intinya, meskipun tidak ada merek yang ingin menghadapi krisis, dilengkapi dengan alat yang tepat, seperti mendengarkan media sosial, dapat menjadi perbedaan antara pemulihan yang cepat dan kerusakan yang berkepanjangan.

Konsep Kunci dalam Mendengarkan dan Melacak Media Sosial

Mendengarkan dan melacak media sosial mencakup beberapa konsep kunci, termasuk pembahasan sentimen, penyebutan merek, pelacakan influencer, dan KPI media sosial, di antara lainnya. Masing-masing konsep ini memainkan peran penting dalam memantau kehadiran merek dan memahami persepsi pelanggan secara online.

- Pembahasan sentimen. Juga disebut sebagai penambangan opini, pembahasan sentimen melibatkan penentuan nada emosional di balik serangkaian kata untuk mendapatkan pemahaman tentang sikap, opini, dan emosi yang diungkapkan dalam

penyebutan online. Ini bukan hanya tentang memantau komentar positif, netral, dan negatif, tetapi juga tentang memahami intensitas sentimen.

- **Penyebutan merek.** Penyebutan merek mengacu pada setiap kali perusahaan atau produknya disebutkan di media sosial. Memantau penyebutan ini dapat membantu merek memahami kehadiran pasarnya dan mengukur reputasinya di antara pelanggan dan pesaing.
- **Pelacakan influencer.** Pelacakan influencer adalah tentang mengidentifikasi dan memantau individu-individu kunci yang memiliki pengaruh signifikan terhadap calon pelanggan atau industri secara keseluruhan. Dengan memantau para influencer ini, perusahaan dapat mengidentifikasi peluang untuk kolaborasi, dukungan, atau kemitraan.
- **KPI media sosial.** KPI adalah metrik yang digunakan untuk mengevaluasi keberhasilan kampanye media sosial. KPI umum meliputi jangkauan, tingkat keterlibatan, tingkat konversi, dan ROI. KPI ini membantu bisnis memahami seberapa baik kinerja kampanye media sosial mereka dan memberikan informasi untuk strategi di masa mendatang.

Singkatnya, konsep-konsep kunci dalam mendengarkan dan melacak media sosial ini menyediakan alat yang diperlukan untuk mengukur efektivitas strategi dan tindakan media sosial, membuka jalan bagi proses pengambilan keputusan berbasis data dalam pemasaran.

Contoh Praktis: Pemantauan Media Sosial untuk Manajemen Reputasi Merek

Salah satu aplikasi praktis pemantauan media sosial adalah dalam manajemen reputasi merek secara online. Mari bergabung dengan kami dalam perjalanan melalui aplikasi dunia nyata, ketika pemantauan media sosial membuktikan kemampuannya dalam mengelola reputasi merek:

Pertimbangkan skenario hipotetis yang melibatkan perusahaan minuman global, SoftDrink Inc. Perusahaan ini memiliki kehadiran di beberapa platform media sosial, termasuk X (sebelumnya Twitter), Facebook, Instagram, dan LinkedIn, dan mereka baru-baru ini meluncurkan lini produk baru berupa minuman sehat dan bebas gula.

Untuk mengukur reaksi konsumen terhadap lini produk baru tersebut, SoftDrink Inc. menggunakan alat pemantauan media sosial untuk memantau penyebutan merek mereka dan lini produk baru tersebut di berbagai platform media sosial. Mereka sangat tertarik untuk memahami sentimen mengenai penyebutan ini, yaitu, apakah reaksinya positif, negatif, atau netral.

Pembahasan sistematis dari data yang dikumpulkan mengungkapkan bahwa meskipun ada sentimen yang umumnya positif terhadap lini produk baru tersebut, beberapa konsumen telah menyuarakan kekhawatiran tentang rasanya. SoftDrink Inc. mempertimbangkan umpan balik ini dan memulai dialog dengan konsumen yang peduli untuk lebih memahami preferensi mereka.

Selain itu, SoftDrink Inc. mengidentifikasi sejumlah influencer yang telah memberikan ulasan positif terhadap produk mereka. Dengan menghubungi para influencer ini untuk potensi kolaborasi, mereka dapat meningkatkan visibilitas dan jangkauan produk mereka di demografi target mereka.

Terakhir, perusahaan melacak KPI seperti jumlah penyebutan, skor sentimen, jangkauan, dan tingkat keterlibatan. Mereka menggunakan data ini untuk menilai keberhasilan peluncuran produk mereka dan menggunakan wawasan ini untuk menginformasikan pengembangan produk dan strategi pemasaran mereka di masa mendatang.

Pemantauan media sosial dengan cara ini membantu SoftDrink Inc. untuk tetap unggul dalam menghadapi potensi masalah reputasi, berinteraksi dengan audiens mereka secara bermakna, dan memperoleh wawasan yang dapat ditindaklanjuti untuk meningkatkan produk dan strategi pemasaran mereka.

7.5 OPTIMASI TINGKAT KONVERSI

Dengan wawasan yang telah diperoleh, langkah terakhir adalah optimasi. Optimasi tingkat konversi menjadi kunci untuk membuka potensi penuh sebuah situs web.

Gambaran Umum Optimasi Tingkat Konversi

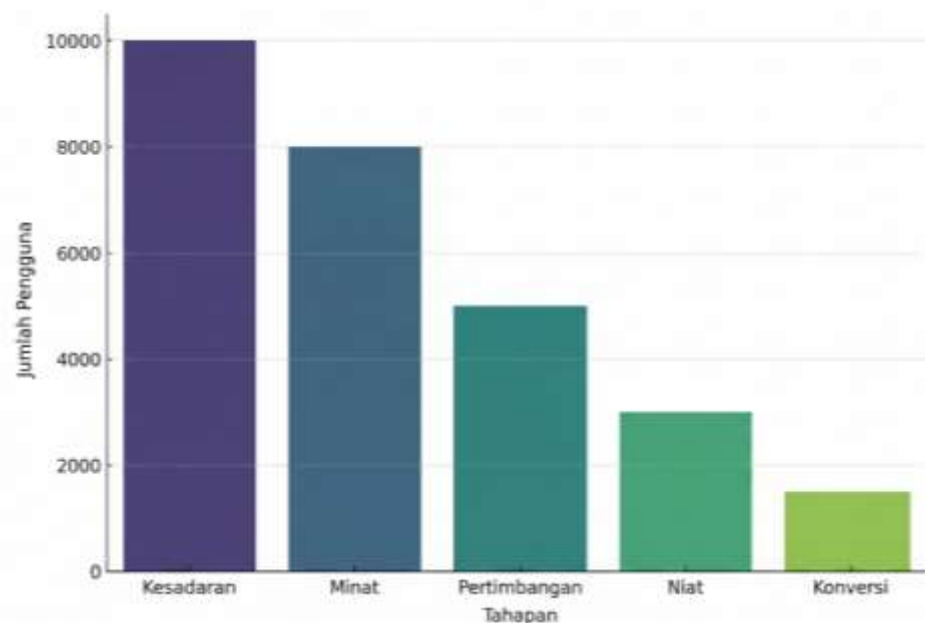
Optimasi tingkat konversi (CRO) adalah proses sistematis yang melibatkan peningkatan persentase pengunjung situs web yang menyelesaikan tindakan yang diinginkan situs web, yang dikenal sebagai konversi. Ini bisa berupa apa saja, mulai dari melakukan pembelian, mengisi formulir, mengunduh sumber daya, mendaftar untuk buletin, atau tindakan lain yang selaras dengan tujuan organisasi.

Tujuan utama CRO adalah untuk memaksimalkan lalu lintas web yang ada dengan mengoptimalkan UX untuk memandu pengunjung menuju penyelesaian tindakan yang diinginkan. Fondasi CRO yang sukses terletak pada pemahaman bagaimana pengguna bergerak melalui situs Anda, tindakan apa yang mereka lakukan, dan apa yang menghalangi mereka untuk menyelesaikan tujuan Anda.

Salah satu langkah pertama dalam CRO melibatkan identifikasi KPI yang berkaitan dengan tujuan situs. Ini bisa berupa tingkat konversi untuk halaman tertentu, jumlah penyelesaian formulir, atau jumlah total pendaftaran baru. Kemudian, melalui kombinasi metode pengumpulan data kuantitatif dan kualitatif, seperti analitik web, peta panas, rekaman pengunjung, survei, dan pengujian pengguna, hipotesis dikembangkan tentang perubahan apa yang dapat meningkatkan kinerja situs.

Hipotesis ini kemudian diuji, biasanya menggunakan pengujian A/B atau multivariat. Hasil pengujian ini memberikan informasi untuk langkah-langkah optimasi lebih lanjut, menjadikan CRO sebagai proses berkelanjutan dan berulang. Untuk memberikan peta jalan visual dari proses CRO, Gambar 7.10 menyajikan bagan batang yang menggambarkan berbagai tahapan dalam CRO, dari kesadaran hingga konversi.

CRO yang efektif dapat menghasilkan peningkatan ROI, peningkatan UX, dan wawasan tentang pelanggan yang dapat memberikan informasi untuk area lain dalam pemasaran digital dan pengembangan produk.



Gambar 7.10 Tahapan dalam CRO.

Pengujian A/B untuk Landing Page

Pada intinya, pengujian A/B adalah eksperimen di mana dua atau lebih versi halaman web ditampilkan kepada pengguna secara acak, dan pembahasan statistik digunakan untuk menentukan versi mana yang berkinerja lebih baik untuk tujuan konversi tertentu. Ketika diterapkan pada landing page, pengujian A/B berfungsi sebagai alat penting bagi pemasar untuk mengoptimalkan konversi.

Landing page adalah pintu masuk digital ke merek, produk, atau layanan. Biasanya, halaman inilah yang dikunjungi pengunjung setelah mengklik CTA pemasaran, seperti iklan atau tautan email. Mengingat pentingnya landing page dalam saluran konversi, memastikan optimasinya sangat penting.

Dalam pengujian A/B untuk landing page, pemasar dapat membuat variasi pada judul, tombol CTA, gambar, testimoni, atau elemen lainnya. Tujuannya adalah untuk menentukan versi mana yang lebih efektif mendorong pengunjung untuk melakukan tindakan yang diinginkan, baik itu mendaftar untuk buletin, membeli produk, atau tujuan konversi lainnya.

Misalnya, jika sebuah perusahaan merasa tingkat pendaftaran mereka kurang memuaskan, mereka mungkin berhipotesis bahwa warna atau teks tombol CTA tidak cukup menarik. Mereka kemudian dapat membuat dua versi halaman arahan: satu dengan tombol asli dan yang lain dengan warna atau teks yang berbeda. Dengan mengarahkan setengah dari lalu lintas ke setiap versi, mereka dapat mengumpulkan data tentang versi mana yang mencapai tingkat konversi yang lebih baik.

Namun, sangat penting untuk hanya mengubah satu elemen pada satu waktu selama pengujian A/B. Ini memastikan bahwa setiap perbedaan kinerja dapat dikaitkan dengan satu perubahan tersebut.

Praktik Terbaik Pengalaman Pengguna untuk Konversi

UX adalah pendekatan holistik yang mencakup semua aspek interaksi pengguna dengan perusahaan, layanan, dan produknya. Ketika dioptimalkan, UX memastikan pengguna menemukan nilai yang nyata, yang mengarah pada peningkatan konversi.

Untuk halaman arahan (landing page), UX yang sangat baik berarti pengunjung dapat dengan cepat memahami proposisi nilai dan dengan mudah diarahkan menuju tujuan konversi. Untuk mencapai hal ini, pastikan untuk mempertimbangkan hal-hal berikut selama pembuatan halaman arahan Anda:

- **Kesederhanaan adalah kuncinya.** Halaman arahan harus bersih, rapi, dan bebas dari informasi yang tidak perlu. Setiap elemen harus berfungsi untuk mengarahkan pengunjung menuju tujuan konversi.
- **Proposisi nilai yang jelas.** Dalam hitungan detik setelah mendarat di halaman, pengunjung harus dapat memahami apa yang ditawarkan dan manfaatnya. Ini dapat dicapai melalui judul yang ringkas, deskripsi yang jelas, dan visual yang relevan.
- **Gambar yang relevan.** Visual memainkan peran penting dalam menarik perhatian. Menggunakan gambar atau video berkualitas tinggi dan relevan dapat membantu mengilustrasikan manfaat penawaran dan sesuai dengan kebutuhan atau keinginan pengunjung.
- **CTA yang menarik.** Tombol CTA harus menonjol dan menggunakan teks yang berorientasi pada tindakan. Alih-alih teks umum seperti 'Klik di sini', menggunakan ajakan yang lebih spesifik seperti 'Dapatkan e-book gratis saya' dapat lebih efektif.
- **Optimasi seluler.** Dengan sebagian besar lalu lintas web berasal dari perangkat seluler, memastikan halaman arahan dioptimalkan untuk seluler dapat sangat memengaruhi konversi. Ini berarti waktu pemuatan yang cepat, ukuran teks yang mudah dibaca, dan tombol yang ramah sentuhan.

Untuk lebih memahami dampak nyata dari praktik terbaik UX pada tingkat konversi, Tabel 7.5 menyajikan daftar praktik-praktik ini, masing-masing disertai dengan penjelasan singkat tentang pengaruh spesifiknya dalam meningkatkan keterlibatan pengguna dan mendorong konversi.

Tabel 7.5 Praktik Terbaik UX dengan Penjelasan Singkat.

Praktik Terbaik	Dampak pada Tingkat Konversi
Ajakan Bertindak (CTA) yang Jelas	Mengarahkan pengguna ke tindakan yang diinginkan, meningkatkan kemungkinan konversi.
Desain Responsif	Memastikan UX yang konsisten di berbagai perangkat, menjangkau audiens yang lebih luas.
Waktu Pemuatan Cepat	Mengurangi rasio pentalan dengan cepat menyajikan konten kepada pengguna.
Navigasi Intuitif	Membantu pengguna menemukan apa yang mereka cari dengan mudah, sehingga meningkatkan keterlibatan.
Gambar Berkualitas Tinggi	Menawarkan tampilan dan nuansa profesional, meningkatkan nilai produk yang dirasakan.

Sinyal Kepercayaan	Meningkatkan kredibilitas, membuat pengguna lebih nyaman untuk mengambil tindakan.
Ulasan/Testimoni Pengguna	Memberikan bukti sosial, meningkatkan kepercayaan pengguna pada produk atau layanan.
Gangguan Minimal	Membuat pengguna tetap fokus pada tujuan utama, meningkatkan peluang konversi.

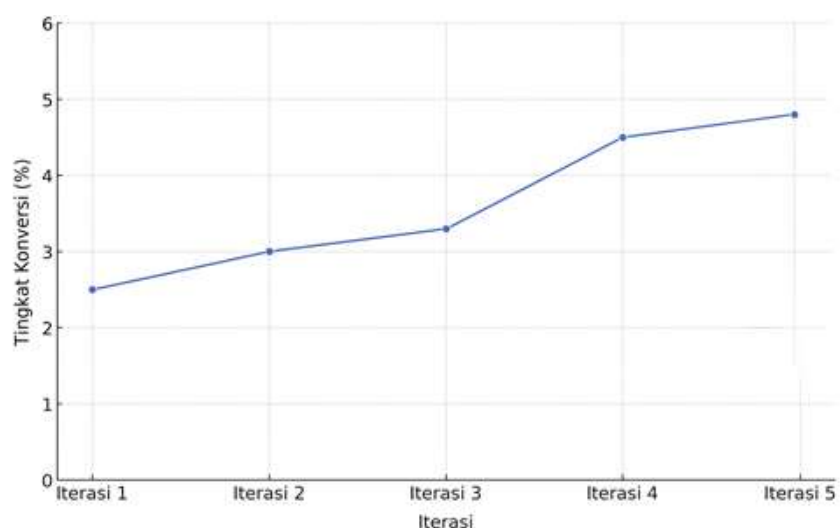
Pembahasan dan Mengulang Strategi Optimasi Tingkat Konversi

CRO bukanlah proses sekali jalan, melainkan siklus pembahasan dan pengulangan yang berkelanjutan. Setelah menerapkan strategi, langkah penting selanjutnya adalah mengukur efektivitasnya.

Pembahasan melibatkan pengumpulan data tentang tingkat konversi baru, metrik keterlibatan pengguna, dan KPI relevan lainnya. Alat seperti Google Analytics dapat menawarkan wawasan mendalam, mulai dari rasio pentalan hingga waktu yang dihabiskan di halaman. Data ini memberikan gambaran yang jelas tentang seberapa baik perubahan yang diterapkan mendorong konversi.

Namun, CRO tidak berhenti pada pembahasan. Wawasan yang diperoleh kemudian harus menjadi dasar strategi selanjutnya. Misalnya, jika suatu perubahan menghasilkan konversi yang lebih tinggi tetapi meningkatkan rasio pentalan, mungkin ada elemen yang membuat pengguna menjauh setelah konversi awal. Hal ini dapat mendorong putaran pengujian A/B baru untuk mengoptimalkan elemen tersebut.

Selain itu, umpan balik pengguna dapat sangat berharga. Alat umpan balik langsung atau survei dapat mengungkap frustrasi atau keinginan pengguna yang mungkin terlewatkan oleh data kuantitatif. Untuk mengilustrasikan sifat dinamis dan terus berkembang dari CRO, Gambar 7.11 menyajikan grafik garis yang menunjukkan peningkatan rasio konversi setelah beberapa iterasi CRO, menyoroti efektivitas pembahasan berkelanjutan dan penyempurnaan strategi.



Gambar 7.11 Peningkatan Tingkat Konversi Setelah Beberapa Iterasi CRO.

Pada intinya, siklus implementasi, pembahasan, dan iterasi adalah jantung dari CRO. Ini adalah perjalanan berkelanjutan menuju penyempurnaan UX dan memaksimalkan konversi.

Konsep Kunci dalam Optimasi Tingkat Konversi

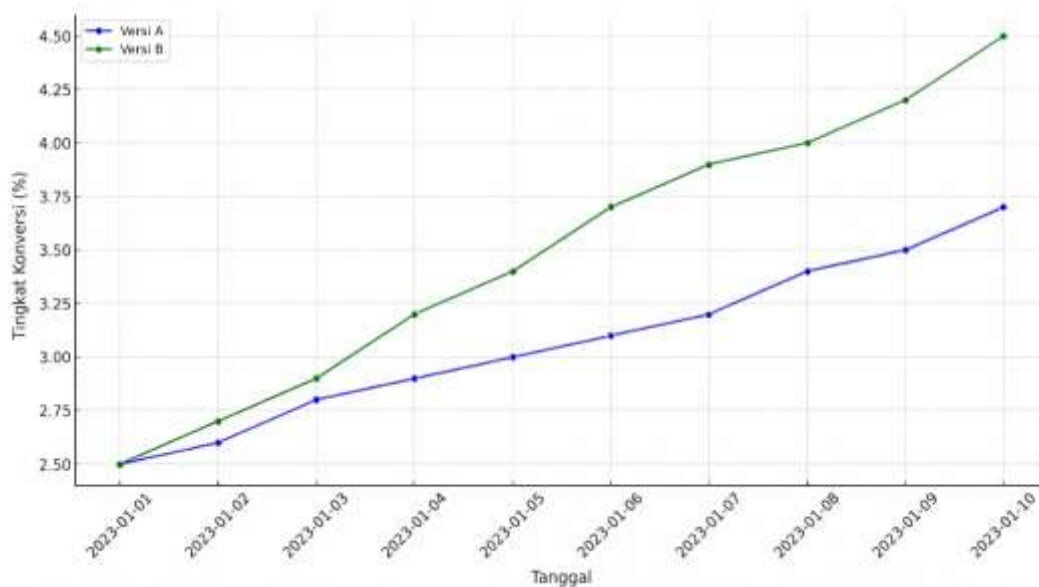
Ada beberapa konsep kunci dalam CRO yang penting untuk memahami dan menjalankan strategi CRO yang efektif:

- **Corong konversi.** Ini adalah representasi perjalanan pelanggan dari kunjungan awal ke situs web hingga konversi akhir. Memahami corong konversi sangat penting untuk mengidentifikasi di mana pelanggan potensial berhenti dan apa yang dapat dilakukan untuk mencegah hal ini.
- **Pengujian A/B.** Ini adalah metode yang digunakan untuk menguji perubahan pada halaman web terhadap desain saat ini dan menentukan mana yang menghasilkan hasil yang lebih unggul. Ini melibatkan pengujian dua versi (A dan B) dengan menunjukkannya kepada pengunjung yang serupa pada waktu yang sama. Versi yang memberikan tingkat konversi yang lebih baik akan menang.
- **Pengujian multivariat.** Mirip dengan pengujian A/B, tetapi alih-alih menguji satu perubahan, beberapa variabel diubah dan diuji untuk memahami interaksi antara berbagai elemen halaman.
- **UX.** Ini mengacu pada pengalaman keseluruhan yang dimiliki pengguna saat berinteraksi dengan situs web. UX yang positif sangat penting untuk CRO karena secara langsung memengaruhi kemungkinan pengguna untuk melakukan konversi.
- **CTA.** CTA adalah ajakan di situs web yang memberi tahu pengguna untuk melakukan tindakan tertentu. CTA biasanya ditulis sebagai perintah atau frasa tindakan, seperti 'Daftar' atau 'Beli Sekarang' dan umumnya berbentuk tombol atau hyperlink.
- **Optimasi halaman arahan.** Proses peningkatan elemen di situs web untuk meningkatkan konversi. Optimasi halaman arahan adalah bagian dari CRO dan melibatkan penggunaan metode seperti pengujian A/B untuk meningkatkan tujuan konversi dari halaman arahan tertentu.

Contoh Praktis: Pengujian A/B untuk Optimasi Tingkat Konversi

Seperti yang telah dibahas, pengujian A/B adalah metode populer yang digunakan oleh bisnis untuk mengoptimalkan situs web mereka dan meningkatkan konversi. Untuk contoh praktis ini, kita akan meneliti toko e-commerce hipotetis yang menjual sepatu.

Toko e-commerce tersebut menyadari bahwa meskipun mereka memiliki arus lalu lintas yang stabil ke situs mereka, tingkat konversi mereka (persentase pengunjung yang melakukan pembelian) lebih rendah daripada rata-rata industri. Untuk mengatasi hal ini, mereka memutuskan untuk menjalankan pengujian A/B pada halaman produk mereka, yang merupakan langkah terakhir sebelum pengunjung menambahkan barang ke keranjang belanja mereka.



Gambar 7.12 Hasil Pengujian A/B yang Menunjukkan Perbedaan Kinerja Antara Versi A dan Versi B.

Mereka berhipotesis bahwa tombol 'Tambahkan ke Keranjang' tidak cukup menonjol, dan pengunjung mungkin melewatkannya. Dalam desain saat ini (Versi A), tombol 'Tambahkan ke Keranjang' berada di bawah deskripsi produk dan memiliki warna yang sama dengan bagian situs web lainnya. Mereka membuat desain baru (Versi B) di mana tombol 'Tambahkan ke Keranjang' dipindahkan ke atas deskripsi produk dan diubah menjadi warna yang kontras.

Setengah dari pengunjung situs web diperlihatkan Versi A, dan setengah lainnya diperlihatkan Versi B. Kemudian mereka mengumpulkan data tentang jumlah pengunjung yang diterima setiap versi dan berapa banyak dari pengunjung tersebut yang menambahkan produk ke keranjang belanja mereka.

Setelah satu bulan, mereka pembahasan hasilnya. Mereka menemukan bahwa Versi B, dengan tombol 'Tambahkan ke Keranjang' yang lebih menonjol, memiliki tingkat konversi yang jauh lebih tinggi daripada Versi A. Berdasarkan hal ini, mereka memutuskan untuk menerapkan Versi B untuk semua pengunjung (lihat Gambar 7.12).

Contoh ini menggambarkan proses dan potensi manfaat pengujian A/B. Dengan membuat keputusan berdasarkan data, bisnis dapat secara signifikan meningkatkan tingkat konversi dan kinerja keseluruhan mereka.

7.6 KESIMPULAN

Di era digital, analitik telah menjadi sangat penting untuk kesuksesan pemasaran. Dengan platform media sosial yang menampung miliaran pengguna di seluruh dunia, pemasar memiliki peluang yang belum pernah terjadi sebelumnya untuk menjangkau audiens target mereka. SNA menyediakan cara untuk memahami dinamika kompleks platform ini dan dapat memandu strategi untuk pemasaran influencer, pemasaran viral, dan banyak lagi.

Pembahasan web, bagaimanapun, menawarkan wawasan tentang bagaimana pengguna berinteraksi dengan situs web suatu merek. Pembahasan web dapat memberikan metrik penting seperti rasio pentalan, durasi sesi, rasio konversi, dan banyak lagi. Dengan menggunakan wawasan ini, pemasar dapat mengoptimalkan situs web mereka untuk meningkatkan UX dan meningkatkan konversi. Pengujian A/B, seperti yang telah kita lihat, dapat memainkan peran penting dalam proses optimasi ini.

Terakhir, mendengarkan dan melacak media sosial membantu pemasar memahami percakapan tentang merek mereka di platform media sosial. Dengan melacak penyebutan, sentimen, dan tren, merek dapat mengelola reputasi mereka, menanggapi kekhawatiran pelanggan, dan mengidentifikasi peluang untuk keterlibatan.

Kesimpulannya, pembahasan media sosial dan pembahasan web adalah alat penting bagi pemasar modern. Pembahasan web memberikan wawasan yang dibutuhkan untuk memahami audiens, mengoptimalkan properti digital, dan berinteraksi dengan pelanggan secara efektif. Teknik dan contoh yang dibahas dalam bab ini hanya sebagian kecil dari apa yang mungkin dilakukan di bidang yang menarik ini.

LATIHAN 7.1

PEMBAHASAN JARINGAN SOSIAL (SNA) DALAM PEMASARAN

Tujuan:

Memahami penerapan pembahasan jaringan sosial dalam mengidentifikasi pengguna berpengaruh dalam konteks pemasaran.

Tugas:

1. Visualisasikan Jaringan Sosial:

- Gunakan networkx untuk membuat representasi visual jaringan.
- Tandai node kunci yang mungkin mewakili pengguna berpengaruh.

2. Hitung Ukuran Sentralitas:

- Hitung derajat, antara, dan sentralitas vektor eigen untuk setiap node.
- Identifikasi lima pengguna berpengaruh teratas berdasarkan ukuran sentralitas ini.

3. Diskusi:

- Diskusikan bagaimana ukuran-ukuran ini dapat membantu mengidentifikasi calon influencer untuk kampanye pemasaran.
- Apa saja keterbatasan pendekatan ini?

Langkah-langkah:

1. Muat Data dan Impor Pustaka:

```
1. import matplotlib.pyplot as plt
2. import networkx as nx
3. sna_data=pd.read_csv("/data/Social_Network_Analysis_Data.csv")
```

Baris ini mengimpor pustaka matplotlib dan networkx serta memuat data jejaring sosial dari file CSV ke dalam DataFrame pandas.

2. Buat Grafik:

```
4. G = nx.Graph()
5. for index, row in sna_data.iterrows():
```

```
6. G.add_node(row['User'],followers=row['Followers'],engagement_rate=row['Engagement Rate'])
```

Di sini, kita membuat grafik kosong menggunakan networkx dan kemudian menambahkan node (pengguna) dari DataFrame. Setiap node ditambahkan dengan atribut 'pengikut' dan 'tingkat keterlibatan'. Grafik saat ini memiliki 25 node. Berikut adalah atribut untuk lima node pertama:

- Pengguna 0: 2.832 pengikut, tingkat keterlibatan 5,93%
- Pengguna 1: 3.364 pengikut, tingkat keterlibatan 6,03%
- Pengguna 2: 9.325 pengikut, tingkat keterlibatan 6,24%
- Pengguna 3: 5.974 pengikut, tingkat keterlibatan 4,38%
- Pengguna 4: 6.844 pengikut, tingkat keterlibatan 2,73%

7. Tambahkan Tepi Acak:

```
7. for _ in range(100):
```

```
8. G.add_edge(np.random.choice(sna_data['User']),np.random.choice(sna_data['User']))
```

- Kami mensimulasikan koneksi sosial dengan membuat garis penghubung secara acak antar simpul.

8. Visualisasikan Jaringan Sosial:

```
9. plt.figure(figsize=(12, 8))
```

```
10. pos = nx.spring_layout(G)
```

```
11. nx.draw_networkx_nodes(G, pos, node_size=700)
```

```
12. nx.draw_networkx_edges(G, pos, width=1.0, alpha=0.5)
```

```
13. nx.draw_networkx_labels(G, pos, font_size=10)
```

```
14. plt.title("Social Network Graph")
```

```
15. plt.axis('off')
```

```
16. plt.show()
```

Blok kode ini memvisualisasikan jaringan sosial. Kita menggunakan **spring_layout** untuk memposisikan node dan kemudian menggambar node, edge, dan label.

9. Hitung Ukuran Sentralitas:

```
17. degree centrality = nx.degree_centrality(G)
```

```
18. betweenness centrality = nx.betweenness_centrality(G)
```

```
19. eigenvector centrality = nx.eigenvector_centrality(G)
```

- Di sini, kita menghitung tiga ukuran sentralitas untuk setiap node: sentralitas derajat, sentralitas antara, dan sentralitas vektor eigen.

10. Mengidentifikasi Pengguna Paling Berpengaruh:

```
20. degree centrality = nx.degree_centrality(G)
```

```
21. betweenness centrality = nx.betweenness_centrality(G)
```

```
22. eigenvector centrality = nx.eigenvector_centrality(G)
```

Kami mengidentifikasi lima pengguna teratas berdasarkan setiap ukuran sentralitas dengan mengurutkan node dan memilih lima teratas.

LATIHAN 7.2

ANALITIK WEB UNTUK WAWASAN PEMASARAN

Tujuan

Memahami bagaimana analitik web dapat digunakan untuk mendapatkan wawasan tentang perilaku pelanggan dan meningkatkan kinerja situs web.

Tugas:

1. Pembahasan Data:

- Pembahasan perilaku pengguna: halaman yang paling sering dikunjungi, waktu rata-rata yang dihabiskan per halaman, rasio pentalan, dan sebagainya.
- Mengidentifikasi pola yang mengarah pada konversi.

2. Optimasi Tingkat Konversi (CRO):

- Menyarankan perubahan pada situs web berdasarkan pembahasan untuk meningkatkan tingkat konversi.
- Mendiskusikan bagaimana pengujian A/B dapat digunakan untuk menguji perubahan ini.

3. Diskusi:

- Mendiskusikan peran analitik web dalam memahami perilaku pelanggan.
- Bagaimana wawasan ini dapat diintegrasikan dengan strategi pemasaran yang lebih luas?

Langkah-langkah:

1. Impor Pustaka yang Diperlukan dan Muat Data Analitik Web:

```
1. import seaborn as sns
2. import pandas as pd
3. web_analytics_data = pd.read_csv("/data/Web_Analytics_Data.csv")
```

Baris ini mengimpor pustaka yang diperlukan untuk pembahasan data dan memuat data analitik web dari file CSV ke dalam DataFrame pandas.

2. Konversi 'Session Timestamp':

Di sini, kita mengkonversi kolom 'Session_Timestamp' ke format tanggal dan waktu untuk pembahasan yang lebih mudah.

```
4. web_analytics_data['Session_Timestamp']=pd.to_datetime(web_analytics_data['Session_Timestamp'])
```

3. Pembahasan Perilaku Pengguna:

▪ Halaman yang Paling Sering Dikunjungi:

```
5. most_visited_pages=web_analytics_data['Page_Visited'].value_counts()
```

Baris ini menghitung frekuensi kunjungan ke setiap halaman.

▪ Rata-rata Waktu yang Dihabiskan di Halaman:

```
6. web_analytics_data['Time_Spent']=np.random.randint(1,300,size=len(web_analytics_data))
7. avg_time_spent=web_analytics_data.groupby('Page_Visited')['Time_Spent'].mean()
```

Di sini, kami mensimulasikan waktu rata-rata yang dihabiskan di setiap halaman. Kemudian kami menghitung waktu rata-rata yang dihabiskan per halaman.

▪ Perhitungan Rasio Pentalan:

```
8. bounce_rate=web_analytics_data[web_analytics_data['Action']=='View']  
   .groupby('User_ID').size()  
9. bounce_rate = (bounce_rate == 1).sum() / len(bounce_rate)
```

Kami menghitung rasio pentalan (bounce rate) dengan mencari persentase sesi di mana hanya satu halaman yang dilihat.

- **Perhitungan Rasio Konversi:**

```
10. conversion_rate = web_analytics_data['Conversion'].mean()
```

Rasio konversi dihitung sebagai rata-rata kolom 'Konversi'.

4. Visualisasi Data:

- **Distribusi Kunjungan Halaman:**

```
11. sns.countplot(x='Page_Visited', data=web_analytics_data)
```

Garis ini membuat plot hitungan yang menunjukkan distribusi kunjungan halaman.

- **Rata-rata Waktu yang Dhabiskan di Setiap Halaman:**

```
12. sns.barplot(x=avg_time_spent.index, y=avg_time_spent.values)
```

Baris ini membuat plot batang yang menunjukkan waktu rata-rata yang dihabiskan di setiap halaman.

Output terdiri dari hal-hal berikut:

- Jumlah kunjungan ke setiap halaman (**most_visited_pages**)
- Waktu rata-rata yang dihabiskan di setiap halaman (**avg_time_spent**)
- Rasio pentalan yang dihitung (**bounce:rate**)
- Rasio konversi keseluruhan (**conversion_rate**)

Pembahasan dan visualisasi ini membantu dalam memahami perilaku pengguna di situs web, mengidentifikasi konten populer, dan mengukur efektivitas situs web dalam menjaga pengunjung tetap terlibat dan mendorong konversi.

BAB 8

PEMODELAN BAURAN PEMASARAN DAN ATRIBUSI

8.1 PENDAHULUAN

Dalam dunia pemasaran modern yang kompleks dan multi-saluran, memahami dampak berbagai aktivitas pemasaran dan secara akurat mengaitkan hasilnya dengan aktivitas tersebut merupakan tugas yang menantang namun penting. Bab ini akan membahas konsep pemodelan bauran pemasaran (MMM) dan atribusi, dua metodologi ampuh yang membantu pemasar mengukur efektivitas upaya mereka dan mengalokasikan sumber daya secara lebih efisien.

MMM adalah teknik statistik yang menggunakan data historis untuk mengukur dampak berbagai komponen pemasaran terhadap penjualan atau indikator kinerja utama (KPI) lainnya. Namun, model atribusi membantu dalam memberikan kredit yang tepat kepada berbagai titik kontak pemasaran dalam perjalanan pelanggan menuju konversi. Memahami dan menerapkan metodologi ini dapat memberikan keunggulan kompetitif dengan menginformasikan keputusan strategis, mengoptimalkan pengeluaran pemasaran, dan, pada akhirnya, memaksimalkan pengembalian investasi pemasaran (ROMI).

Seiring dengan terus berkembangnya pemasaran digital dan semakin berbasis data, penerapan teknik-teknik ini menjadi semakin canggih. Bab ini bertujuan untuk memberikan pemahaman mendalam tentang metodologi ini, penerapannya, dan signifikansinya dalam lanskap pemasaran saat ini.

8.2 KONSEP PEMODELAN BAURAN PEMASARAN

Setelah meletakkan dasar pemahaman tentang MMM dan atribusi, mari kita telusuri lebih dalam konsep-konsep kunci MMM, yang telah lama menjadi landasan dalam dunia analitik pemasaran.

Gambaran Umum Pemodelan Bauran Pemasaran

Seperti yang dijelaskan dalam pendahuluan, MMM adalah pendekatan statistik yang digunakan untuk mengukur dampak berbagai upaya pemasaran terhadap penjualan perusahaan atau KPI lainnya. Teknik ini muncul sebagai respons terhadap meningkatnya kebutuhan akan akuntabilitas dalam aktivitas pemasaran dan dorongan menuju pengambilan keputusan berbasis data.

Pendekatan ini menggabungkan pembahasan regresi berganda untuk memprediksi hasilnya (biasanya penjualan atau pangsa pasar) berdasarkan berbagai masukan pemasaran. Input-input ini, yang sering disebut sebagai bauran pemasaran, biasanya mencakup empat P (produk, harga, promosi, dan tempat), meskipun dapat diperluas ke faktor-faktor lain seperti indikator makroekonomi, musiman, dan aktivitas pesaing.

Salah satu kekuatan utama MMM adalah kemampuannya untuk memisahkan pengaruh berbagai aktivitas pemasaran terhadap penjualan. Dengan demikian, hal ini memungkinkan pemasar untuk memahami efektivitas individu dan gabungan dari berbagai

saluran dan kampanye pemasaran. Hal ini juga membantu dalam menentukan alokasi anggaran pemasaran yang optimal di berbagai saluran dan aktivitas.

Namun, MMM juga memiliki keterbatasan. Misalnya, MMM terutama merupakan pembahasan historis, yang berarti mungkin tidak secara akurat memprediksi kinerja di masa depan, terutama di pasar yang berubah dengan cepat. Selain itu, MMM mungkin tidak sepenuhnya menangkap seluk-beluk perilaku pelanggan atau efek tidak langsung dari aktivitas pemasaran.

Terlepas dari keterbatasan ini, MMM tetap menjadi alat penting dalam perangkat pemasar, memberikan wawasan berharga yang dapat memandu pengambilan keputusan strategis dan alokasi sumber daya.

Komponen Utama: Periklanan, Promosi, Distribusi, dan Penetapan Harga

Setelah memahami gambaran yang lebih luas tentang MMM, mari kita uraikan menjadi komponen-komponen utamanya: periklanan, promosi, distribusi, dan penetapan harga, yang membentuk pilar pendekatan ini.

Memahami komponen-komponen yang membentuk bauran pemasaran sangat penting untuk penerapan MMM. Kerangka kerja empat P tradisional mencakup hal-hal berikut (lihat Gambar 8.1):



Gambar 8.1 Distribusi Anggaran di Berbagai Komponen: Periklanan, Promosi, Distribusi, dan Penetapan Harga.

- **Periklanan (produk).** Ini adalah bentuk komunikasi non-pribadi berbayar yang digunakan untuk mempromosikan atau menjual produk, layanan, atau ide. Iklan yang efektif dapat membangun kesadaran merek, mendorong permintaan produk, dan memfasilitasi kesuksesan pasar jangka panjang. Alat-alatnya meliputi iklan cetak, iklan TV, iklan online, dan papan reklame luar ruangan.
- **Promosi.** Aktivitas promosi bertujuan untuk merangsang permintaan terhadap suatu produk atau layanan. Mereka sering menawarkan keuntungan sementara seperti diskon, kupon, atau sampel gratis untuk menarik pembelian. Promosi dapat

mendorong lonjakan penjualan jangka pendek dan dapat sangat efektif ketika meluncurkan produk baru atau memasuki pasar baru.

- **Distribusi (tempat).** Ini mengacu pada bagaimana suatu produk sampai ke konsumen. Distribusi yang efektif memastikan bahwa produk tersedia di lokasi yang tepat dan pada waktu yang tepat. Saluran dapat mencakup pengecer fisik, platform e-commerce, atau metode langsung ke konsumen. Strategi distribusi dapat secara signifikan memengaruhi kinerja penjualan dan persepsi merek.
- **Penetapan Harga.** Keputusan penetapan harga menentukan berapa banyak konsumen membayar untuk suatu produk. Ini adalah komponen penting yang memengaruhi profitabilitas, permintaan, dan posisi merek. Faktor-faktor yang perlu dipertimbangkan meliputi biaya produksi, harga pesaing, nilai yang dirasakan, dan elastisitas permintaan. Penetapan harga dinamis dan penetapan harga psikologis adalah strategi populer di industri tertentu.

Masing-masing komponen ini dapat secara signifikan memengaruhi penjualan dan KPI lainnya. Dalam MMM, memahami efek individual dan gabungannya sangat penting untuk merancang strategi pemasaran yang efektif (lihat Tabel 8.1).

Tabel 8.1 Komponen Utama Pemodelan Bauran Pemasaran beserta Deskripsi dan Contohnya.

Komponen	Deskripsi
Data penjualan	Data penjualan historis, biasanya pada tingkat mingguan atau bulanan
Pengeluaran iklan	Detail pengeluaran untuk berbagai saluran periklanan
Data promosi	Informasi tentang acara promosi atau diskon
Data pesaing	Data penjualan dan pemasaran dari pesaing
Indikator ekonomi	Indikator makroekonomi yang dapat memengaruhi penjualan, seperti PDB, tingkat pengangguran, dan sebagainya
Data digital	Data dari saluran digital seperti kunjungan situs web, klik, dan sebagainya

Teknik Ekonometrik dan Pembahasan Regresi dalam Pemodelan Bauran Pemasaran

Meskipun komponen-komponen tersebut memberikan struktur, inti dari pemodelan terletak pada teknik yang digunakan. Penerapan teknik ekonometrik, khususnya pembahasan regresi, merupakan dasar dari MMM. Model ekonometrik menggunakan metode statistik untuk menguji hipotesis dan memperkirakan hubungan antar variabel.

Pembahasan regresi, khususnya, digunakan untuk memprediksi variabel hasil (misalnya, penjualan) berdasarkan satu atau lebih variabel prediktor (misalnya, pengeluaran iklan, promosi). Pembahasan regresi berganda, yang menggabungkan beberapa variabel prediktor, paling umum digunakan dalam MMM.

Koefisien yang dihasilkan dari regresi menunjukkan hubungan antara variabel prediktor dan hasil. Dalam konteks MMM, koefisien akan menunjukkan seberapa besar penjualan diperkirakan akan berubah untuk perubahan satu unit dalam variabel prediktor, dengan asumsi faktor-faktor lain tetap konstan. Lebih lanjut, statistik kecocokan model (seperti R-

squared) memberikan wawasan tentang seberapa baik model menjelaskan variabilitas dalam variabel hasil.

Untuk memberikan gambaran komprehensif tentang berbagai teknik ekonometrik yang digunakan dalam MMM dan aplikasinya masing-masing, Tabel 8.2 merinci berbagai metodologi seperti regresi linier dan ARIMA, menguraikan bagaimana masing-masing digunakan untuk pembahasan dan memprediksi dampak aktivitas pemasaran terhadap penjualan dan hasil bisnis utama lainnya.

Tabel 8.2 Teknik Ekonometrik yang Digunakan dalam Pemodelan Bauran Pemasaran dan Aplikasinya.

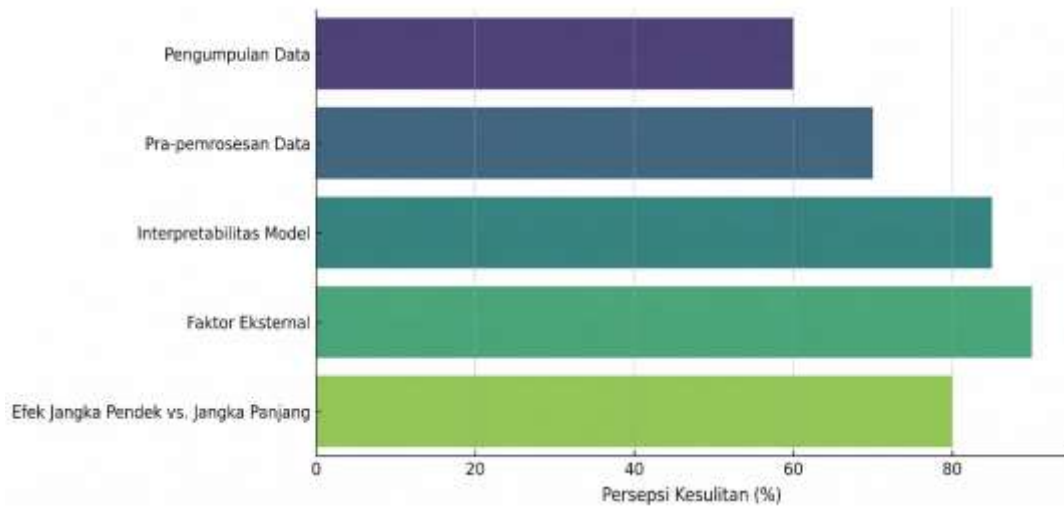
Teknik	Aplikasi
Regresi linier	Menilai dampak berbagai aktivitas pemasaran terhadap penjualan
Pembahasan deret waktu	Pembahasan dan memprediksi data penjualan yang berurutan dari waktu ke waktu
Model linier umum	Memodelkan hubungan dengan distribusi kesalahan non-normal
ARIMA	Memprediksi penjualan menggunakan komponen autoregresif dan rata-rata bergerak
Regresi multivariat	Pembahasan dampak beberapa prediktor pada variabel respons

Tantangan dan Keterbatasan Pemodelan Bauran Pemasaran Tradisional

Meskipun regresi menawarkan pendekatan yang kuat, mirip dengan semua model, MMM memiliki tantangan dan keterbatasannya sendiri (lihat Gambar 8.2). Mari kita bedah apa saja tantangan dan keterbatasan tersebut dan bagaimana pengaruhnya terhadap pembahasan kita:

- **Ketergantungan data historis.** MMM bergantung pada data masa lalu, yang mungkin tidak selalu menunjukkan hasil di masa depan, terutama di pasar yang berkembang pesat.
- **Pengecualian faktor yang tidak dapat dikuantifikasi.** Faktor-faktor berpengaruh tertentu, seperti sentimen merek atau pemasaran dari mulut ke mulut, mungkin tidak mudah dikuantifikasi dan dapat dikecualikan dari model.
- **Kausalitas versus korelasi.** MMM mungkin menunjukkan korelasi, tetapi tidak selalu membuktikan kausalitas.
- **Granularitas.** MMM tradisional mungkin tidak menangkap detail granular dari kampanye atau segmen pelanggan. Kurangnya detail ini dapat berarti bahwa wawasan spesifik tentang kampanye individual atau segmen pelanggan khusus terlewatkan. Untuk MMM modern, hal ini memerlukan integrasi alat pembahasan data yang lebih canggih, seperti yang dijelaskan kemudian dalam bab ini, yang dapat menangkap dan pembahasan data pada tingkat yang lebih detail. Dengan memanfaatkan model dan alat yang lebih canggih, bisnis dapat mengatasi masalah granularitas ini, memastikan bahwa strategi pemasaran mereka disesuaikan lebih tepat dengan audiens dan kampanye tertentu, sehingga mengoptimalkan upaya pemasaran mereka.

- **Tantangan multiplatform.** Di dunia digital saat ini, konsumen berinteraksi di berbagai platform, yang dapat membuat pengumpulan dan interpretasi data menjadi lebih kompleks.
- **Overfitting model.** Jika terlalu banyak variabel yang disertakan, model mungkin terlalu sesuai dengan data historis, sehingga mengurangi akurasi prediksinya untuk data di masa mendatang.



Gambar 8.2 Tantangan yang Dihadapi dalam Pemodelan Bauran Pemasaran Tradisional.

Seperti halnya semua model analitik, sangat penting untuk memahami keterbatasan ini ketika menafsirkan hasil dan membuat keputusan strategis.

Di Luar Regresi: Teknik Analitik Tambahan dalam Pemodelan Bauran Pemasaran

Meskipun regresi adalah teknik yang banyak digunakan dalam MMM, pendekatan analitik yang menyeluruh juga harus mempertimbangkan pengintegrasian metode tambahan untuk memperhitungkan nuansa dan kompleksitas yang tidak ditangkap oleh regresi saja:

- **Pembelajaran mesin dan kecerdasan buatan.** Dengan meningkatnya daya komputasi dan ketersediaan data, model pembelajaran mesin seperti pohon keputusan, hutan acak, dan jaringan saraf menawarkan cara alternatif untuk pembahasan data. Model-model ini dapat mengungkap hubungan nonlinier, interaksi, dan dapat menangani kumpulan data besar, memberikan prediksi yang lebih akurat dalam skenario tertentu.
- **Metode Bayesian.** Teknik Bayesian, yang menggabungkan keyakinan awal dan memperbaruinya dengan data baru, dapat bermanfaat, terutama ketika data jarang atau ketika mengintegrasikan pendapat ahli sangat penting. Untuk MMM, model Bayesian dapat memberikan estimasi yang lebih kuat, terutama dalam situasi dengan data terbatas.
- **Pembahasan segmentasi.** MMM terkadang dapat melewati nuansa segmen konsumen yang berbeda. Pembahasan kluster dan teknik segmentasi lainnya dapat mengidentifikasi dan pembahasan segmen-segmen ini, menawarkan pemahaman yang lebih rinci tentang perilaku konsumen dan respons terhadap rangsangan pemasaran.

- **Koreksi endogenitas.** Masalah seperti simultanitas, di mana pengeluaran pemasaran mungkin dipengaruhi oleh penjualan (dan sebaliknya), dapat menyebabkan koefisien regresi yang bias. Instrumen, fungsi kontrol, dan teknik ekonometrik lainnya dapat membantu mengoreksi endogenitas, memberikan estimasi yang lebih akurat.
- **Pembahasan dekomposisi.** Teknik seperti dekomposisi nilai Shapley dapat membantu dalam mengaitkan efek dari suatu tindakan pemasaran di berbagai saluran, terutama di ruang digital. Hal ini membantu dalam memahami interaksi dan dampak gabungan dari berbagai tindakan pemasaran.

Meskipun regresi tetap menjadi alat dasar dalam MMM, kompleksitas modern membutuhkan pendekatan terintegrasi yang memanfaatkan berbagai metode analitis. Dengan demikian, pemasar dapat memperoleh pandangan yang lebih komprehensif dan bernuansa tentang upaya mereka, yang mengarah pada pengambilan keputusan yang lebih tepat dan strategi pemasaran yang optimal.

Konsep Kunci dalam Pemodelan Bauran Pemasaran

Meskipun penuh tantangan, ada beberapa konsep dasar dalam MMM yang sangat penting. Memahami konsep-konsep ini akan memungkinkan pemahaman dan penerapan yang lebih mendalam. Konsep-konsep kunci dalam MMM adalah sebagai berikut:

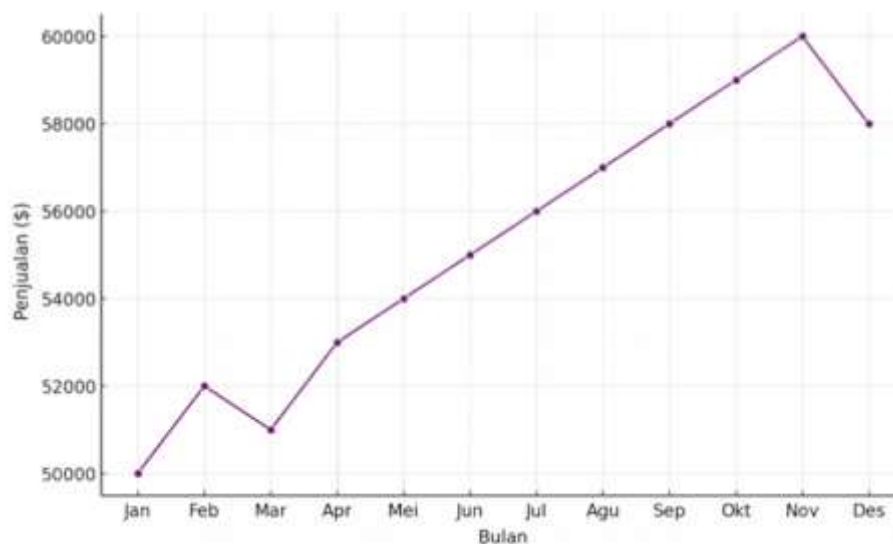
- **Variabel dependen dan independen.** Variabel dependen adalah hasil yang ingin diprediksi atau dipengaruhi oleh bisnis, seringkali penjualan atau pangsa pasar. Variabel independen adalah input pemasaran atau faktor lain yang mungkin memengaruhi variabel dependen, seperti pengeluaran iklan di berbagai saluran, penetapan harga, distribusi, promosi, aktivitas pesaing, dan faktor ekonomi.
- **Pembahasan regresi.** Pembahasan regresi adalah metode statistik yang digunakan untuk memahami hubungan antara variabel dependen dan independen. Dalam konteks MMM, pembahasan regresi berganda biasanya digunakan untuk memahami efek simultan dari beberapa input pemasaran terhadap penjualan atau KPI lainnya.
- **Koefisien.** Koefisien yang dihasilkan dari pembahasan regresi mewakili perkiraan dampak setiap variabel independen terhadap variabel dependen. Dalam konteks MMM, koefisien ini memberi tahu pemasar seberapa besar penjualan atau pangsa pasar yang diharapkan berubah untuk perubahan satu unit dalam input pemasaran tertentu, dengan mempertahankan semua faktor lain tetap konstan.
- **Elastisitas.** Elastisitas mengukur perubahan persentase dalam variabel dependen (misalnya, penjualan) untuk perubahan 1% dalam variabel independen (misalnya, pengeluaran iklan). Konsep ini membantu pemasar memahami respons penjualan terhadap perubahan dalam berbagai input pemasaran dan sangat membantu dalam pengambilan keputusan alokasi anggaran.
- **Efek sinergi.** Dalam banyak kasus, dampak dari aktivitas pemasaran gabungan lebih besar daripada jumlah efek individunya. Efek sinergi ini dapat ditangkap dalam MMM melalui penyertaan istilah interaksi dalam model regresi.

Contoh Praktis: Pemodelan Bauran Pemasaran untuk Perusahaan Barang Konsumen

Teori dan konsep paling baik dipahami ketika dipraktikkan. Mari kita bahas contoh praktis bagaimana MMM diimplementasikan untuk perusahaan barang konsumsi. MMM dapat menjadi alat yang berharga bagi perusahaan barang konsumsi dalam memahami efektivitas upaya pemasaran mereka.

Bayangkan sebuah perusahaan, Sunrise Soaps, yang menjual berbagai produk perawatan rumah dan pribadi. Mereka berinvestasi dalam berbagai saluran pemasaran, termasuk iklan TV, iklan online, media cetak, dan promosi di toko. Mereka juga menghadapi berbagai tingkat persaingan dan dipengaruhi oleh faktor musiman.

Mereka memutuskan untuk menggunakan MMM untuk memahami dampak upaya pemasaran mereka terhadap penjualan triwulanan mereka. Variabel dependen dalam model mereka adalah penjualan triwulanan, dan variabel independen meliputi pengeluaran pemasaran di berbagai saluran, pengeluaran pemasaran pesaing, dan variabel dummy untuk menangkap efek musiman.



Gambar 8.3 Hasil dari Contoh Praktis untuk Perusahaan Barang Konsumen.

Menggunakan data historis, mereka menyesuaikan model regresi berganda dan menemukan hasil berikut (lihat Gambar 8.3):

- Iklan TV memiliki koefisien tertinggi, menunjukkan bahwa iklan TV memiliki dampak terbesar pada penjualan. Hal ini konsisten dengan jangkauan dan paparan luas yang diberikan oleh iklan TV.
- Iklan online memiliki koefisien yang lebih kecil tetapi elastisitas yang tinggi. Hal ini menunjukkan bahwa meskipun dampak iklan online terhadap penjualan saat ini lebih kecil daripada iklan TV, penjualan sangat responsif terhadap perubahan pengeluaran iklan online. Mengingat tren peningkatan konsumsi digital, ini menghadirkan peluang untuk mengoptimalkan anggaran pemasaran.

- Variabel dummy menunjukkan efek musiman yang signifikan, dengan penjualan meningkat pada kuartal liburan. Wawasan ini dapat digunakan untuk mengatur waktu aktivitas pemasaran agar memberikan dampak maksimal.
- Istilah interaksi antara iklan TV dan online bersifat positif dan signifikan, menunjukkan efek sinergi. Ini berarti efek gabungan iklan TV dan online lebih besar daripada jumlah efek masing-masing.

Wawasan ini dapat memandu Sunrise Soaps dalam membuat keputusan berbasis data tentang alokasi anggaran pemasaran, waktu kegiatan pemasaran, dan koordinasi upaya pemasaran di berbagai saluran untuk dampak maksimal.

8.3 MODEL ATRIBUSI BERBASIS DATA

Beralih dari model tradisional, munculnya saluran digital memerlukan pendekatan yang lebih bernuansa: model atribusi berbasis data. Model-model ini memberikan wawasan tentang perjalanan yang ditempuh pelanggan dengan merek Anda.

Gambaran Umum Model Atribusi Berbasis Data

Model atribusi berbasis data adalah jenis model atribusi yang menggunakan pembelajaran mesin dan algoritma statistik untuk memberikan kredit pada titik kontak pemasaran. Tidak seperti model atribusi berbasis aturan, seperti klik terakhir atau klik pertama, yang memberikan kredit berdasarkan aturan yang telah ditentukan sebelumnya, model atribusi berbasis data belajar dari data untuk memahami kontribusi setiap titik kontak dalam jalur konversi.

Dengan meningkatnya pemasaran digital, pelanggan sering berinteraksi dengan merek beberapa kali di berbagai saluran sebelum melakukan pembelian. Interaksi ini dapat mencakup melihat iklan online, mengunjungi situs web perusahaan, mengklik postingan media sosial, atau menerima email. Setiap titik kontak ini dapat memengaruhi keputusan pelanggan untuk membeli, dan model atribusi berbasis data bertujuan untuk mengukur pengaruh ini.

Model atribusi berbasis data menggunakan data konversi historis dan mempertimbangkan semua titik kontak dalam jalur konversi. Dengan menggunakan algoritma pembelajaran mesin, mereka mengidentifikasi pola dalam data, seperti urutan titik kontak yang sering mengarah pada konversi dan memberikan lebih banyak kredit pada titik kontak ini. Dengan cara ini, mereka dapat memperhitungkan kompleksitas dan keragaman perjalanan pelanggan di lingkungan multichannel saat ini.

Model atribusi berbasis data menawarkan beberapa manfaat. Mereka memberikan ukuran ROI yang lebih akurat dari berbagai saluran pemasaran, memungkinkan pemasar untuk mengoptimalkan pengeluaran pemasaran mereka. Mereka juga menawarkan wawasan tentang perjalanan pelanggan, membantu pemasar memahami peran berbagai titik kontak dalam mendorong konversi.

Namun, model atribusi berbasis data juga memiliki tantangannya sendiri. Model atribusi berbasis data membutuhkan sejumlah besar data untuk memberikan hasil yang andal, dan implementasi serta pemahamannya bisa rumit. Selain itu, model ini bersifat korelasional dan tidak dapat membuktikan sebab-akibat. Model atribusi berbasis data, meskipun

memberikan wawasan, terutama menyoroti korelasi, yang hanya menunjukkan hubungan antar variabel tanpa membuktikan bahwa satu variabel menyebabkan variabel lainnya.

Perbedaan ini sangat penting karena menganggap korelasi sebagai sebab-akibat dapat menyebabkan investasi pemasaran yang salah arah, mengabaikan titik kontak utama, dan salah menafsirkan pendorong sebenarnya di balik konversi. Sifat perjalanan pelanggan saat ini yang beragam, keberadaan variabel pengganggu, dan lingkaran umpan balik menambah kompleksitas, sehingga sulit untuk mengisolasi dampak sebab-akibat sebenarnya dari masing-masing titik kontak terhadap konversi. Akibatnya, pemasar harus menafsirkan model ini dengan hati-hati, memastikan mereka tidak mengacaukan pola yang diamati dengan tindakan sebab-akibat yang pasti (lihat Tabel 8.3).

Tabel 8.3 Keuntungan dan Kerugian Model Atribusi Tradisional dan Berbasis Data.

Atribut	Atribusi Tradisional	Atribusi Berbasis Data
Dasar	Memberikan kredit berdasarkan aturan sederhana (misalnya, sentuhan terakhir, sentuhan pertama)	Menggunakan algoritma dan model statistik untuk menetapkan kredit
Keuntungan	Mudah dipahami dan diimplementasikan	Lebih akurat karena mempertimbangkan semua titik kontak
Kerugian	Tidak memperhitungkan semua interaksi dalam perjalanan pelanggan	Membutuhkan data yang substansial dan dapat rumit untuk disiapkan

Model Atribusi Sentuhan Terakhir, Sentuhan Pertama, dan Linier

Terdapat beberapa model atribusi yang digunakan di industri ini. Beberapa model yang paling populer meliputi model sentuhan terakhir, sentuhan pertama, dan linier. Mari kita telusuri seluk-beluk model-model ini. Memahami nilai yang diberikan oleh berbagai titik sentuh terhadap konversi sangat penting.

Berbagai model atribusi membantu menentukan hal ini, dengan masing-masing memberikan wawasan unik:

- **Atribusi sentuhan terakhir.** Ini adalah model yang paling umum digunakan dan mengaitkan 100% nilai konversi ke titik sentuh terakhir sebelum konversi. Ini mudah dipahami tetapi dapat terlalu menekankan titik sentuh terakhir, seringkali dengan mengorbankan upaya pemasaran sebelumnya.
- **Atribusi sentuhan pertama.** Model ini memberikan 100% kredit kepada titik kontak pertama yang mengarahkan pelanggan ke jalur konversi. Ini berguna untuk memahami kampanye membangun kesadaran tetapi dapat mengabaikan titik kontak selanjutnya yang mungkin mendorong konversi akhir.
- **Atribusi linier.** Ini mendistribusikan kredit konversi secara merata di semua titik kontak. Ini mengenali setiap langkah dalam perjalanan pelanggan tetapi dapat terlalu menyederhanakan dengan mengasumsikan setiap titik kontak memiliki dampak yang sama.

Model Atribusi Algoritma atau Probabilistik

Model atribusi algoritma atau probabilistik menggunakan statistik canggih, pembelajaran mesin, dan big data untuk mengaitkan nilai setiap titik kontak. Tidak seperti model berbasis aturan (seperti sentuhan terakhir atau linier), model algoritma pembahasan sejumlah besar data untuk menentukan kontribusi yang paling mungkin dari setiap titik kontak (lihat Tabel 8.4):

- **Pembahasan data.** Model ini mempertimbangkan semua interaksi di sepanjang jalur konversi pengguna, pembahasan pola dan urutan yang mengarah ke konversi.
- **Fleksibilitas.** Model algoritmik dapat beradaptasi dengan perubahan perilaku pelanggan dan dinamika pasar, terus diperbarui berdasarkan data terbaru.

Mengevaluasi Kinerja Model

Setelah Anda memanfaatkan model algoritmik, langkah penting selanjutnya adalah mengevaluasi kinerjanya untuk memastikan model tersebut selaras dengan tujuan bisnis:

- **Akurasi.** Nilai apakah model tersebut secara tepat menetapkan nilai pada titik kontak. Membandingkan prediksi model dengan set data uji atau menjalankan uji A/B (lihat Bagian 4.4) dapat memvalidasi akurasi.
- **Granularitas.** Model yang baik harus memberikan wawasan granular, menangkap nuansa dari berbagai saluran, kampanye, atau segmen pelanggan.
- **Kemampuan Adaptasi.** Lanskap pemasaran berubah. Kinerja suatu model mungkin menurun seiring waktu, sehingga model tersebut harus mampu beradaptasi dengan perubahan perilaku pelanggan atau kondisi pasar.

Tabel 8.4 Gambaran Umum dan Perbandingan Model Atribusi Sentuhan Terakhir, Sentuhan Pertama, Linier, dan Algoritmik.

Model	Deskripsi
Sentuhan terakhir	Memberikan kredit 100% pada titik kontak terakhir sebelum konversi
Sentuhan pertama	Memberikan kredit 100% pada titik kontak pertama dalam perjalanan pelanggan
Linear	Mendistribusikan kredit secara merata di semua titik kontak
Algoritma	Menggunakan teknik berbasis data untuk mendistribusikan kredit berdasarkan pengaruh setiap titik kontak

Tantangan dan Jebakan Implementasi

Evaluasi yang berhasil tidak menjamin implementasi yang lancar. Ada berbagai tantangan dan jebakan yang harus kita waspadai saat mengimplementasikan model atribusi berbasis data:

- **Silo data.** Data mungkin ada dalam silo di berbagai platform, sehingga konsolidasi dan pembahasan menjadi sulit.
- **Keterlalu kompleksitas.** Model yang terlalu kompleks mungkin akurat tetapi dapat menjadi sulit untuk diinterpretasikan dan ditindaklanjuti.

- **Mengabaikan faktor eksternal.** Hanya berfokus pada data internal dapat mengabaikan pengaruh eksternal, seperti kondisi ekonomi atau tindakan pesaing, yang dapat memengaruhi kinerja.
- **Menganggap perilaku statis.** Perilaku dan preferensi pelanggan berkembang. Mengasumsikan perilaku statis dapat menyebabkan atribusi yang tidak akurat.

Meskipun model atribusi menawarkan wawasan yang berharga, sangat penting untuk mendekatinya dengan bijaksana, memahami kekuatan, keterbatasan, dan potensi tantangan dalam implementasinya.

Konsep Kunci dalam Model Atribusi Berbasis Data

Untuk mengatasi tantangan ini, pemahaman tentang konsep kunci model atribusi berbasis data sangat penting. Model atribusi berbasis data memanfaatkan konsep kunci yang memungkinkan pemasar untuk memberikan representasi yang lebih akurat tentang nilai dari berbagai titik kontak pemasaran. Berikut adalah beberapa konsep tersebut:

- **Jalur konversi.** Jalur konversi mengacu pada urutan titik kontak yang diinteraksikan pelanggan sebelum menyelesaikan konversi, seperti melakukan pembelian atau mengisi formulir. Model atribusi berbasis data membahas jalur konversi untuk memahami pola yang mengarah pada konversi.
- **Titik kontak.** Titik kontak adalah setiap interaksi yang dilakukan pelanggan dengan suatu merek. Dalam konteks pemasaran digital, titik sentuh dapat mencakup tayangan iklan, klik, kunjungan situs web, keterlibatan media sosial, dan banyak lagi. Model atribusi berbasis data memberikan kredit kepada berbagai titik sentuh berdasarkan pengaruhnya terhadap konversi.
- **Penugasan kredit.** Penugasan kredit mengacu pada proses mendistribusikan kredit untuk suatu konversi di antara berbagai titik sentuh dalam jalur konversi. Model atribusi berbasis data menggunakan algoritma untuk menentukan berapa banyak kredit yang harus diterima setiap titik sentuh.
- **Pembelajaran mesin.** Model atribusi berbasis data menggunakan algoritma pembelajaran mesin untuk belajar dari data konversi dan mengidentifikasi pola. Algoritma ini dapat menangani sejumlah besar data dan memperhitungkan hubungan kompleks dan nonlinier antar variabel.
- **Signifikansi statistik.** Signifikansi statistik adalah ukuran kemungkinan bahwa hasil yang diamati dalam data terjadi secara kebetulan. Dalam konteks atribusi berbasis data, signifikansi statistik dapat digunakan untuk menentukan apakah perbedaan dalam pemberian kredit antara berbagai titik kontak bermakna.

Contoh Praktis: Atribusi Berbasis Data untuk Perusahaan E-Commerce

Dengan konsep yang telah dipahami, mari kita perkuat pemahaman kita dengan contoh praktis yang berpusat pada perusahaan e-commerce. Menerapkan model atribusi berbasis data untuk perusahaan e-commerce melibatkan berbagai langkah dan praktik. Mari kita pertimbangkan skenario di mana perusahaan e-commerce ingin lebih memahami nilai berbagai saluran pemasarannya dalam mendorong konversi, dan dengan demikian memutuskan untuk menerapkan model atribusi berbasis data.

1. **Pengumpulan data.** Langkah pertama adalah mengumpulkan data tentang interaksi pelanggan di semua saluran pemasaran, termasuk iklan pencarian, iklan display, pemasaran email, media sosial, dan pencarian organik. Data ini biasanya mencakup jenis interaksi (misalnya, klik iklan, pembukaan email), waktu interaksi, dan apakah interaksi tersebut akhirnya menghasilkan konversi.
2. **Pembahasan jalur konversi.** Perusahaan kemudian pembahasan jalur konversi—urutan titik kontak yang mengarah ke konversi—untuk mengidentifikasi pola. Misalnya, perusahaan mungkin menemukan bahwa pelanggan yang berinteraksi dengan iklan display lebih cenderung melakukan pembelian jika mereka kemudian menerima email.
3. **Pengembangan model.** Perusahaan menggunakan algoritma pembelajaran mesin untuk mengembangkan model yang memprediksi konversi berdasarkan urutan titik kontak. Model ini dilatih pada data konversi historis, sehingga mampu mempelajari pola dan hubungan antar variabel.
4. **Penugasan kredit.** Perusahaan menerapkan model untuk menetapkan kredit untuk konversi ke berbagai titik kontak. Misalnya, jika model menemukan bahwa iklan display memainkan peran penting dalam mendorong konversi, model tersebut mungkin akan memberikan persentase kredit yang lebih tinggi kepada iklan display.
5. **Implementasi dan iterasi.** Perusahaan mengimplementasikan model atribusi dan menggunakan hasilnya untuk menginformasikan strategi pemasarannya. Misalnya, jika model memberikan nilai tinggi pada pemasaran email, perusahaan mungkin akan berinvestasi lebih banyak pada saluran ini. Model tersebut diperbarui secara berkala seiring tersedianya data baru.

8.4 ATRIBUSI MULTI-SENTUH (MTA)

Melangkah lebih jauh dari titik sentuh tunggal, atribusi multi-sentuh (MTA) mempertimbangkan beberapa titik sentuh dalam perjalanan pelanggan, memberikan pandangan yang lebih holistik.

Gambaran Umum Atribusi Multi-Sentuh

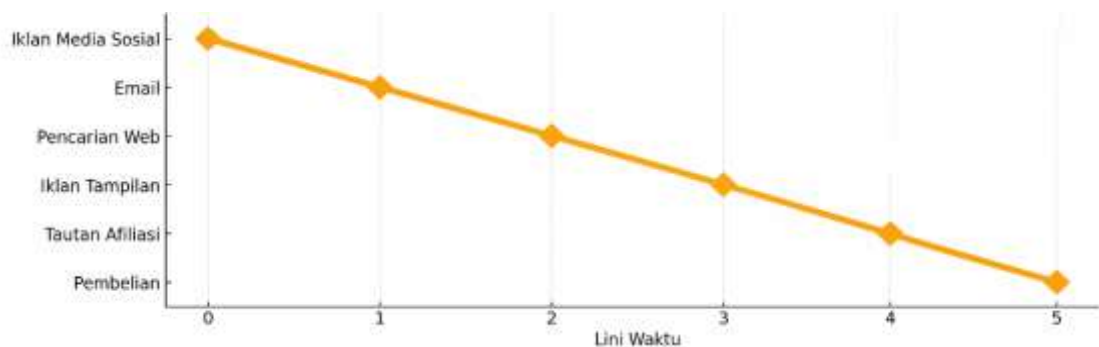
MTA adalah metode canggih untuk memahami dan mengaitkan nilai yang diberikan oleh berbagai titik sentuh pemasaran terhadap konversi akhir. Model ini mengakui kompleksitas perjalanan pelanggan modern yang seringkali mencakup banyak titik kontak di berbagai saluran sebelum pembelian dilakukan.

Dalam model sentuhan tunggal tradisional, semua kredit untuk konversi diberikan kepada satu titik kontak, biasanya yang terakhir sebelum konversi (atribusi klik terakhir) atau yang pertama (atribusi klik pertama). Meskipun model-model ini sederhana dan mudah diimplementasikan, model-model ini tidak secara akurat mencerminkan realitas perjalanan pelanggan, yang biasanya jauh lebih kompleks.

Namun, model MTA mendistribusikan kredit untuk konversi di beberapa titik kontak. Hal ini memungkinkan pemasar untuk memahami dampak setiap titik kontak dan mengoptimalkan upaya pemasaran mereka sesuai dengan itu. Ada beberapa jenis model MTA, termasuk model linier, peluruhan waktu, dan berbentuk U. Pilihan model bergantung pada

kekhususan bisnis dan strategi pemasarannya. Namun, implementasi MTA bukannya tanpa tantangan. Hal ini membutuhkan sejumlah besar data, kemampuan analitik tingkat lanjut, dan seringkali integrasi data di berbagai platform dan saluran. Salah satu keterbatasan penting dari banyak model MTA adalah asumsi dasar independensi antar titik kontak. Ini berarti model-model tersebut sering mengasumsikan bahwa pengaruh setiap titik kontak terhadap konversi bersifat independen dari pengaruh titik kontak lainnya.

Pada kenyataannya, interaksi antar titik kontak dapat bersifat sinergis atau antagonis. Misalnya, melihat unggahan media sosial dapat memperkuat efek kampanye email berikutnya, atau sebaliknya. Ketika titik kontak tidak benar-benar independen, pemberian nilai berdasarkan asumsi ini dapat menyebabkan kesalahan estimasi pengaruh sebenarnya dari setiap titik kontak terhadap konversi. Hal ini kemudian dapat menyesatkan pemasar ketika mengoptimalkan strategi mereka. Sangat penting bagi pemasar untuk menyadari keterbatasan ini ketika menafsirkan hasil dan membuat keputusan berdasarkan MTA (lihat Gambar 8.4).



Gambar 8.4 Titik Sentuh Atribusi Multi-Sentuhan yang Menekankan Pentingnya Interaksi Ganda.

Terlepas dari tantangan ini, wawasan yang diberikan oleh MTA dapat secara signifikan meningkatkan efektivitas dan efisiensi kampanye pemasaran.

Rasionalisasi Pendekatan Multi-Sentuhan

Untuk sepenuhnya memahami MTA, penting untuk memahami rasionalisasi di balik pengembangannya dan semakin pentingnya. Dalam lanskap digital yang rumit saat ini, pelanggan berinteraksi dengan merek melalui berbagai saluran dan titik sentuh sebelum membuat keputusan pembelian. Dari penemuan merek awal hingga konversi akhir, konsumen mungkin terlibat dengan unggahan media sosial perusahaan, mengklik iklan mesin pencari, membuka email, dan banyak lagi. Mengingat kompleksitas ini, menjadi jelas bahwa mengandalkan satu titik kontak—seperti iklan terakhir yang diklik atau situs web pertama yang dikunjungi—dapat memberikan pandangan yang menyimpang tentang apa yang benar-benar memengaruhi keputusan konsumen.

Inti dari pendekatan MTA terletak pada pengakuannya terhadap sifat multifaset dari perjalanan pelanggan modern. Alih-alih menyederhanakan perjalanan ini secara berlebihan, model multi-sentuh bertujuan untuk menangkap pandangan yang lebih holistik tentang peran dan nilai setiap titik sentuh. Dengan memberikan pemahaman yang lebih mendalam tentang bagaimana berbagai titik sentuh berkontribusi pada tujuan akhir, pemasar dapat

mengalokasikan sumber daya secara lebih efektif, menyempurnakan strategi, dan meningkatkan keterlibatan pelanggan. Keterkaitan ekosistem digital membutuhkan metode yang mencerminkan kompleksitasnya, dan model MTA mampu memenuhi kebutuhan tersebut.

Pendekatan Peluruhan Waktu, Berbasis Posisi, dan Algoritma

Pendekatan Peluruhan Waktu

Model ini menilai titik sentuh berdasarkan kedekatannya dengan konversi akhir, secara progresif memberikan nilai yang lebih besar pada titik sentuh saat semakin dekat dengan momen konversi. Hal ini mengakui bahwa meskipun interaksi awal berperan dalam meningkatkan kesadaran, titik sentuh yang lebih dekat dengan konversi biasanya lebih berpengaruh dalam proses pengambilan keputusan akhir.

Pendekatan Berbasis Posisi

Sering digambarkan sebagai model berbentuk U, pendekatan berbasis posisi menetapkan persentase nilai spesifik untuk titik sentuh pertama dan terakhir, mengakui peran penting mereka dalam memperkenalkan dan memantapkan keputusan konsumen. Nilai yang tersisa kemudian didistribusikan secara merata di seluruh titik sentuh perantara. Model ini menangkap signifikansi penemuan merek awal dan dorongan terakhir menuju konversi sambil tetap menghargai titik sentuh yang membina di antaranya.

Pendekatan Algoritma

Dengan menggunakan pembelajaran mesin dan teknik statistik yang canggih, model algoritma menyelami data historis untuk memastikan pengaruh yang paling mungkin dari setiap titik sentuh. Alih-alih aturan yang telah ditentukan sebelumnya, pendekatan ini secara dinamis menyesuaikan berdasarkan pola dan hubungan yang ditemukan dalam data, menawarkan model atribusi yang sangat adaptif dan disesuaikan.

Untuk menjelaskan perbedaan di antara berbagai model atribusi multi-sentuhan, Tabel 8.5 menyajikan perbandingan fitur-fitur utama dan kasus penggunaan terbaik untuk model peluruhan waktu, berbasis posisi, dan algoritmik, yang menawarkan wawasan tentang bagaimana setiap model memberikan kredit kepada berbagai titik sentuh dalam perjalanan pelanggan.

Tabel 8.5 Fitur Utama dan Perbedaan Antara Model Atribusi Multi-Sentuh Berbasis Peluruhan Waktu, Berbasis Posisi, dan Algoritma.

Model	Deskripsi	Terbaik untuk
Peluruhan waktu	Memberikan lebih banyak poin pada titik kontak yang lebih dekat dengan konversi	Siklus penjualan pendek di mana interaksi terkini lebih penting
Berbasis posisi	Memberikan lebih banyak poin pada titik kontak pertama dan terakhir, dan mendistribusikan sisanya secara merata di antara titik kontak tengah	Perjalanan pelanggan di mana pendahuluan dan kesimpulan sangat penting

Algoritma	Menggunakan algoritma dan teknik statistik untuk memberikan poin berdasarkan pengaruh aktual dari titik kontak	Perjalanan pelanggan yang kompleks dengan banyak data yang tersedia

Alat dan Platform untuk Atribusi Multi-Sentuhan

Dengan pemahaman tentang berbagai pendekatan, alat dan platform yang tersedia membantu dalam memanfaatkan MTA secara efektif (lihat Tabel 8.6).

Tabel 8.6 Alat dan Platform untuk Atribusi Multi-Sentuh dengan Fitur.

Alat/Platform	Fitur
Google Analytics	Pembahasan perjalanan pengguna, pelacakan peristiwa, segmentasi
Adobe Analytics	Atribusi lintas saluran, analitik waktu nyata, segmentasi
Facebook Attribution	Pelacakan lintas perangkat, kinerja iklan, jalur konversi
Visual IQ	Pelacakan multi-saluran, visualisasi perjalanan pelanggan, pembahasan ROI

Munculnya era digital telah disertai dengan pengembangan berbagai alat dan platform yang dirancang untuk membantu pemasar menavigasi MTA (Marketing Theory Analysis). Solusi seperti Google Analytics, Adobe Analytics, dan Visual IQ memberikan wawasan komprehensif tentang perjalanan pelanggan. Platform ini memanfaatkan sejumlah besar data, memprosesnya melalui algoritma canggih untuk memetakan dampak dari berbagai titik kontak.

Google Analytics, misalnya, menawarkan model yang memungkinkan pandangan komparatif tentang bagaimana berbagai pendekatan atribusi dapat menggambarkan perjalanan pelanggan, memfasilitasi pengambilan keputusan yang tepat. Namun, platform seperti Visual IQ menekankan pentingnya interaksi lintas saluran, menyoroti sinergi antara berbagai aktivitas pemasaran.

Mengintegrasikan Data Offline dan Online untuk Atribusi Holistik

Di dunia yang serba terintegrasi saat ini, mengintegrasikan data offline dan online menjadi sangat penting untuk mencapai pandangan atribusi 360 derajat. Meskipun dunia digital menawarkan banyak sekali titik data, penting untuk diingat bahwa konsumen masih berinteraksi dengan merek secara offline. Mengintegrasikan data offline—seperti pembelian di toko, interaksi pusat panggilan, atau kehadiran acara fisik—ke dalam model atribusi sangat penting untuk pemahaman yang holistik.

Alat modern telah mulai menjembatani kesenjangan ini. Misalnya, Kunjungan Toko Google di Google Ads mencoba menghubungkan titik-titik antara klik iklan online dan kunjungan toko offline. Demikian pula, sistem CRM dapat diintegrasikan ke dalam platform analitik digital, membawa data dari penjualan dan interaksi offline ke dalam lingkup atribusi digital.

Menggabungkan data offline tidak hanya menawarkan gambaran lengkap tentang perjalanan pelanggan tetapi juga membantu dalam memahami interaksi antara titik sentuh

online dan offline. Mengenali pengaruh iklan digital pada pembelian di toko, atau bagaimana pengalaman di toko mendorong pencarian online, dapat memberikan wawasan yang tak ternilai bagi pemasar, memandu strategi yang benar-benar sesuai dengan perjalanan konsumen.

Konsep Kunci dalam Atribusi Multi-Sentuhan

Menguasai MTA membutuhkan fondasi yang kuat dalam konsep-konsep kuncinya. Model MTA mendistribusikan kredit untuk konversi di berbagai titik kontak, yang mencerminkan kontribusi setiap titik kontak dalam perjalanan pelanggan menuju konversi. Ada beberapa konsep kunci yang terlibat dalam MTA yang sangat penting untuk memahami dan menerapkan model ini secara efektif:

- **Titik kontak.** Titik kontak mengacu pada setiap interaksi yang dilakukan pelanggan dengan merek, produk, atau layanan. Ini dapat mencakup melihat iklan, mengunjungi situs web, membaca buletin email, atau berinteraksi dengan perwakilan penjualan. Setiap titik kontak memberikan kesempatan bagi merek untuk memengaruhi proses pengambilan keputusan pelanggan.
- **Perjalanan pelanggan.** Perjalanan pelanggan adalah rangkaian lengkap pengalaman yang dialami pelanggan saat berinteraksi dengan merek atau produk. Dalam MTA, perjalanan pelanggan dipertimbangkan untuk mengidentifikasi urutan dan dampak titik kontak yang mengarah pada konversi.
- **Model atribusi.** Model atribusi adalah aturan atau seperangkat aturan yang menentukan bagaimana kredit untuk konversi diberikan kepada titik kontak dalam perjalanan pelanggan. Model yang berbeda memberikan kredit secara berbeda, tergantung pada spesifikasi bisnis dan strategi pemasarannya.
- **Model linier, peluruhan waktu, dan berbentuk U.** Ini adalah contoh model MTA. Model linier memberikan kredit yang sama untuk semua titik kontak, model peluruhan waktu memberikan lebih banyak kredit untuk titik kontak yang lebih dekat dengan konversi, dan model berbentuk U memberikan lebih banyak kredit untuk titik kontak pertama dan terakhir.

Memahami dan menerapkan konsep-konsep ini dengan benar dapat membantu pemasar mendapatkan gambaran yang lebih akurat tentang efektivitas upaya pemasaran mereka dan membuat keputusan yang lebih tepat tentang di mana harus menginvestasikan sumber daya pemasaran mereka.

Contoh Praktis: Atribusi Multi-Sentuhan untuk Pengecer Online

Sekarang, mari kita terapkan pengetahuan kita melalui contoh praktis yang berfokus pada pengecer online. Pertimbangkan pengecer online yang bertujuan untuk mengoptimalkan strategi pemasaran digitalnya. Pengecer tersebut menggunakan beberapa saluran pemasaran: optimasi mesin pencari, iklan bayar per klik (PPC), pemasaran email, media sosial, dan iklan tampilan. Masing-masing saluran ini mewakili titik kontak potensial dalam perjalanan pelanggan, dan pengecer ingin memahami kontribusi setiap titik kontak terhadap konversi akhir: pembelian di situs web mereka.

Untuk mengimplementasikan model MTA, pengecer pertama-tama mengumpulkan data tentang titik kontak pelanggan. Data ini mencakup urutan dan waktu titik kontak dan konversi, serta detail tentang pelanggan dan konteks setiap titik kontak. Peritel kemudian menerapkan model MTA untuk memberikan kredit untuk setiap konversi ke titik kontak yang berkontribusi. Misalnya, mereka mungkin menggunakan model berbentuk U, yang memberikan lebih banyak kredit ke titik kontak pertama dan terakhir dalam urutan tersebut.

Pembahasan hasil model MTA, peritel mungkin menemukan bahwa meskipun iklan PPC sering memulai perjalanan pelanggan, pemasaran email paling efektif dalam menutup penjualan. Dengan wawasan ini, peritel dapat membuat keputusan yang lebih tepat tentang di mana harus menginvestasikan anggaran pemasarannya dan bagaimana menyusun pesan pemasarannya untuk efek maksimal.

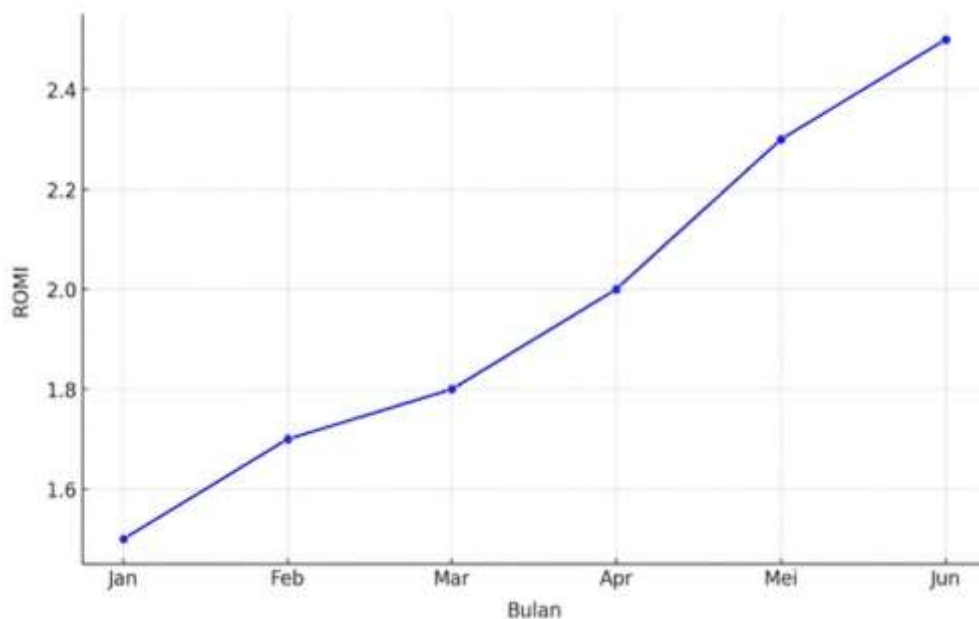
Namun, penting untuk dicatat bahwa model MTA memiliki keterbatasan. Model ini bergantung pada asumsi bahwa semua titik kontak bersifat independen dan efeknya bersifat aditif, yang mungkin tidak selalu demikian. Selain itu, model ini tidak memperhitungkan pengaruh titik kontak offline, seperti pengalaman di toko atau rekomendasi dari mulut ke mulut.

8.5 EFEKTIVITAS STRATEGI PEMASARAN PERUSAHAAN

Beralih dari atribusi, mari kita jelajahi ukuran utama keberhasilan pemasaran: ROMI.

Gambaran Umum Pengembalian Investasi Pemasaran

ROMI adalah metrik yang digunakan untuk mengukur efektivitas strategi pemasaran perusahaan. Tidak seperti metrik pengembalian investasi (ROI) tradisional, ROMI secara khusus berfokus pada pengembalian yang dapat diatribusikan kepada investasi pemasaran.



Gambar 8.5 Pengembalian Investasi Pemasaran dari Waktu ke Waktu.

ROMI memberikan cara untuk mengukur dampak aktivitas pemasaran terhadap profitabilitas perusahaan. Ini dihitung dengan membandingkan nilai finansial tambahan yang diperoleh sebagai hasil dari pengeluaran pemasaran tertentu dengan biaya pengeluaran pemasaran itu sendiri. Perhitungan ini membantu pemasar memahami efektivitas strategi mereka dalam hal menghasilkan pendapatan dan berkontribusi pada laba bersih.

ROMI adalah alat penting untuk pengambilan keputusan pemasaran (lihat Gambar 8.5). Seiring meningkatnya tekanan yang dihadapi perusahaan untuk membenarkan pengeluaran pemasaran mereka, ROMI menawarkan cara untuk menunjukkan nilai aktivitas pemasaran dalam bentuk keuangan. Ini membantu dalam alokasi anggaran dengan mengidentifikasi aktivitas pemasaran yang menghasilkan pengembalian tertinggi, sehingga memungkinkan penggunaan sumber daya pemasaran yang lebih efisien.

Namun, mengukur ROMI bukannya tanpa tantangan. Hal ini membutuhkan atribusi pendapatan yang akurat terhadap aktivitas pemasaran tertentu, yang dapat sulit dilakukan dalam praktiknya, terutama ketika banyak saluran dan titik kontak terlibat dalam perjalanan pelanggan. Hal ini juga membutuhkan pemahaman yang jelas tentang jeda waktu antara aktivitas pemasaran dan dampaknya terhadap perilaku pelanggan, yang dapat sangat bervariasi di berbagai aktivitas pemasaran dan segmen pelanggan.

Menghitung Pengembalian Investasi Pemasaran: Prinsip dan Rumus

Sebelum membahas lebih dalam, mari kita pahami prinsip dan rumus di balik perhitungan ROMI. ROMI merupakan metrik dasar bagi pemasar untuk menentukan efektivitas aktivitas pemasaran mereka. Pada dasarnya, hal ini bertujuan untuk menghubungkan pengeluaran pemasaran dengan manfaat finansial yang dihasilkan oleh aktivitas tersebut.

Prinsip

Prinsip inti di balik ROMI adalah bahwa pemasaran tidak boleh hanya dilihat sebagai pengeluaran, tetapi sebagai investasi yang menghasilkan pengembalian finansial. ROMI membantu mengukur hal ini dengan menyatakan laba bersih dari upaya pemasaran relatif terhadap biaya upaya tersebut.

Rumus

ROMI dapat dihitung menggunakan rumus berikut:

$$ROMI = \frac{\text{Pendapatan Tambahan yang Diatribusikan pada Pemasaran} - \text{Biaya Pemasaran},}{\text{Biaya Pemasaran}}$$

di mana

- Pendapatan tambahan yang diatribusikan ke pemasaran menandakan pendapatan tambahan yang dapat diatribusikan ke aktivitas pemasaran.
- Biaya pemasaran mencakup total pengeluaran untuk inisiatif pemasaran.

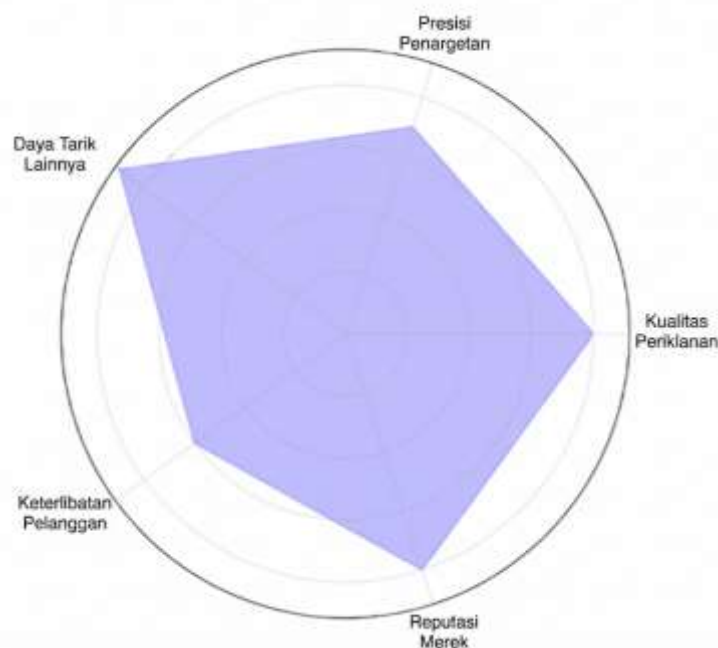
Perlu dicatat bahwa meskipun rumusnya tampak sederhana, mendapatkan nilai yang akurat untuk 'Pendapatan Tambahan yang Diatribusikan ke Pemasaran' bisa rumit, karena

membutuhkan mekanisme atribusi yang tepat untuk menetapkan penjualan atau pendapatan dengan benar ke aktivitas pemasaran tertentu.

Faktor-faktor yang Mempengaruhi Pengembalian Investasi Pemasaran

ROMI, meskipun merupakan metrik tunggal, dipengaruhi oleh banyak faktor (lihat Gambar 8.6). Mari kita jelajahi faktor-faktor apa yang mendorong angka penting ini:

- **Kematangan pasar.** Di pasar yang matang, di mana kesadaran merek sudah tinggi, pendapatan tambahan dari pemasaran tambahan mungkin lebih rendah daripada di pasar berkembang.
- **Siklus hidup produk.** Peluncuran produk baru seringkali membutuhkan investasi pemasaran yang signifikan, tetapi juga dapat menghasilkan pendapatan tambahan yang lebih tinggi karena pasar diperkenalkan dengan penawaran tersebut.
- **Saluran pemasaran yang digunakan.** Saluran yang berbeda memiliki struktur biaya dan jangkauan potensial yang berbeda. Iklan digital mungkin memiliki biaya awal yang lebih rendah daripada kampanye TV nasional tetapi juga dapat memiliki tingkat konversi yang berbeda.
- **Tindakan kompetitif.** Aktivitas pemasaran pesaing dapat mengurangi atau meningkatkan efektivitas pemasaran perusahaan, yang memengaruhi ROMI-nya.
- **Kualitas pelaksanaan pemasaran.** Bahkan dalam saluran atau format yang sama, kualitas, kreativitas, dan relevansi pesan pemasaran memainkan peran penting dalam efektivitasnya.



Gambar 8.6 Faktor-faktor yang Mempengaruhi ROMI.

Mengoptimalkan Biaya Pemasaran demi ROI

Berbekal pengetahuan tentang faktor-faktor yang memengaruhi, fokus bergeser ke optimasi—bagaimana kita bisa mendapatkan hasil terbaik dari anggaran pemasaran kita?

Mencapai ROMI yang optimal seringkali menjadi tujuan utama bagi pemasar. Meningkatkan metrik ini berarti memperoleh pengembalian yang lebih tinggi (pendapatan atau laba) untuk setiap dolar yang dihabiskan untuk pemasaran. Beberapa strategi dapat dilakukan:

- **MMM.** Dengan memahami kontribusi setiap saluran pemasaran terhadap pendapatan keseluruhan, bisnis dapat mengalokasikan anggaran mereka dengan lebih efisien.
- **Segmentasi pelanggan.** Menargetkan segmen pelanggan tertentu yang berpotensi lebih menguntungkan dapat meningkatkan tingkat konversi dan efektivitas kampanye pemasaran.
- **Pengujian berkelanjutan.** Pengujian A/B, pengujian multivariat, atau pendekatan eksperimental lainnya dapat membantu pemasar menyempurnakan strategi mereka, menilai efektivitas taktik baru, dan meninggalkan taktik yang kurang efektif.
- **Komunikasi pemasaran terintegrasi.** Memastikan pesan yang konsisten dan saling melengkapi di semua saluran dapat meningkatkan dampak keseluruhan dari upaya pemasaran, yang mengarah pada ROMI yang lebih baik.

Optimalisasi Pengembalian Investasi Pemasaran yang Sukses

Teori dan prinsip menjadi nyata ketika dilihat dalam praktik. Mari kita tinjau beberapa studi kasus dunia nyata yang menunjukkan optimalisasi ROMI yang sukses.

Perusahaan A. Sebuah platform e-commerce terkemuka, meskipun memiliki kehadiran online yang signifikan, mengalami kesulitan dengan penjualan yang stagnan. Dengan menerapkan MMM, mereka mengidentifikasi kurangnya investasi dalam pemasaran email. Dengan mengalihkan dana dari saluran lain yang kurang efektif ke kampanye email yang ditargetkan, perusahaan tersebut melihat peningkatan penjualan sebesar 20%, yang secara signifikan meningkatkan ROMI mereka.

Perusahaan B. Sebuah merek kecantikan multinasional memutuskan untuk mengoptimalkan ROMI-nya dengan mengintegrasikan data offline dan online. Melalui penggunaan kode QR unik di toko fisik dan profil pelanggan yang terpadu, mereka menghubungkan pembelian offline dengan iklan online. Dengan memahami perjalanan pelanggan secara holistik ini, mereka mengoptimalkan iklan digital mereka, menghasilkan peningkatan penjualan keseluruhan sebesar 15% dan peningkatan yang signifikan dalam ROM.

Studi kasus seperti ini menggarisbawahi pentingnya pendekatan berbasis data dan berpusat pada pelanggan dalam mengoptimalkan ROMI. Dengan memahami perjalanan pelanggan, memanfaatkan saluran yang tepat, dan memastikan pesan yang konsisten, perusahaan dapat secara signifikan meningkatkan ROMI mereka.

Konsep Dasar Pengukuran ROMI

Dari aplikasi praktis ini, ada beberapa konsep dasar dalam ROMI yang menonjol dan patut disorot:

- **Penjualan inkremental.** Ini mengacu pada penjualan tambahan yang dihasilkan karena aktivitas pemasaran tertentu. Ini adalah selisih antara penjualan selama periode promosi dan periode yang sebanding tanpa promosi. Penjualan inkremental seringkali sulit diukur karena faktor eksternal dan fluktuasi pasar.

- **Labamarginal.** Ini adalah laba yang dihasilkan oleh penjualan inkremental. Ini dihitung dengan mengalikan penjualan inkremental dengan margin laba per unit. Tidak semua penjualan inkremental menghasilkan laba inkremental, terutama jika berasal dari promosi harga yang mengurangi margin laba.
- **Investasi pemasaran.** Ini adalah total biaya aktivitas pemasaran, termasuk biaya produk yang dijual, biaya kampanye pemasaran (misalnya, iklan, hubungan masyarakat), dan biaya terkait lainnya.
- **Perhitungan ROMI.** ROMI biasanya dinyatakan sebagai rasio atau persentase, dihitung sebagai (Pendapatan Penjualan Inkremental – Biaya Pemasaran)/Biaya Pemasaran. ROMI sebesar 0,2, misalnya, berarti bahwa untuk setiap dolar yang dihabiskan untuk pemasaran, perusahaan memperoleh keuntungan sebesar 20 sen.

Penting untuk dicatat bahwa meskipun ROMI memberikan ukuran yang berguna untuk efektivitas pemasaran, ia memiliki keterbatasan. ROMI tidak memperhitungkan efek jangka panjang dari aktivitas pemasaran, seperti kesadaran merek dan loyalitas pelanggan.

Selain itu, ROMI mungkin tidak sepenuhnya menangkap dampak aktivitas pemasaran digital, yang sering memengaruhi perjalanan pelanggan dengan cara yang tidak linier dan kompleks.

Menghitung Pengembalian Investasi Pemasaran untuk Kampanye Pemasaran Digital

Berdasarkan eksplorasi teoritis dan berbasis kasus kita, mari kita selami contoh praktis, menghitung ROMI untuk kampanye pemasaran digital. Pertimbangkan contoh hipotetis di mana sebuah perusahaan bernama E-Fashions meluncurkan kampanye pemasaran digital untuk lini pakaian barunya. Total biaya kampanye, yang mencakup pengeluaran untuk pembuatan konten, penempatan iklan, dan biaya agensi, berjumlah Rp500.000.000.

Setelah kampanye, E-Fashions melihat peningkatan penjualan online mereka. Mereka mampu melacak pendapatan sebesar Rp1.200.000.000 langsung kembali ke pelanggan yang mengklik iklan digital mereka, melalui platform analitik web mereka. Margin keuntungan dari penjualan ini adalah 60%.

Untuk menghitung ROMI, pertama-tama kita perlu mencari keuntungan tambahan. Ini akan menjadi pendapatan dari kampanye dikalikan dengan margin keuntungan, yaitu, $Rp1.200.000.000 \times 0,60 = Rp720.000.000$.

ROMI kemudian dihitung sebagai berikut:

$$ROMI = (Keuntungan Tambahan - Investasi Pemasaran) / Investasi Pemasaran$$

$$ROMI = (Rp720.000.000 - Rp500.000.000) / Rp500.000.000$$

$$ROMI = 0,44 \text{ atau } 44\%$$

Ini berarti bahwa untuk setiap dolar yang dihabiskan untuk kampanye, E-Fashions memperoleh keuntungan sebesar 44 sen. Hasil ini menunjukkan bahwa kampanye pemasaran digital cukup efektif.

Ingatlah bahwa meskipun ini adalah contoh yang disederhanakan, pada kenyataannya, perhitungannya bisa lebih kompleks. Misalnya, mungkin sulit untuk melacak secara akurat semua penjualan yang dihasilkan dari kampanye tersebut. Selain itu, kampanye tersebut

mungkin memiliki manfaat yang tidak tercakup dalam ROMI, seperti peningkatan kesadaran merek atau loyalitas pelanggan.

8.6 KESIMPULAN

Bab 8 memberikan pemantauan komprehensif tentang MMM dan atribusi, dua komponen penting dalam lanskap pemasaran modern. MMM menawarkan perspektif holistik tentang efektivitas berbagai saluran dan inisiatif pemasaran, memungkinkan organisasi untuk mengoptimalkan strategi dan anggaran pemasaran mereka. Namun, model atribusi, khususnya model berbasis data dan multi-sentuhan, menawarkan wawasan mendalam tentang perjalanan pelanggan, menyoroti peran titik sentuh individual dalam menghasilkan konversi.

Kedua metode ini, jika digunakan secara efektif, dapat secara signifikan meningkatkan ROMI, metrik kunci yang mengukur profitabilitas aktivitas pemasaran. Dengan mengukur ROMI, organisasi dapat memastikan bahwa pengeluaran pemasaran mereka menghasilkan pengembalian positif dan berkontribusi pada laba bersih.

Seiring lanskap pemasaran menjadi semakin digital dan berbasis data, pentingnya konsep-konsep ini diperkirakan akan meningkat. Pemasar yang dapat memanfaatkan teknik-teknik ini secara efektif akan memiliki keunggulan kompetitif dalam mendorong keterlibatan pelanggan dan mencapai tujuan bisnis.

LATIHAN 8.1

PEMODELAN BAURAN PEMASARAN (MMM)

Tujuan

Mengembangkan model regresi berganda untuk memahami dampak berbagai upaya pemasaran terhadap penjualan perusahaan hipotetis.

Tugas:

1. Muat data yang dihasilkan ke dalam DataFrame.
2. Lakukan pembahasan data eksplorasi (EDA) untuk memahami distribusi dan korelasi data.
3. Bangun model regresi berganda untuk pembahasan pengaruh setiap saluran pemasaran terhadap penjualan.
4. Interpretasikan koefisien dan evaluasi kinerja model.

Langkah-langkah:

1. Memuat Pustaka:

- Pertama, kita perlu mengimpor pustaka yang diperlukan untuk manipulasi data dan pembahasan statistik.

```
1. import pandas as pd
2. import numpy as np
3. import statsmodels.api as sm
```

- **pandas** digunakan untuk manipulasi dan pembahasan data.
- **numpy** digunakan untuk operasi numerik.

- **statsmodels** digunakan untuk memperkirakan dan menginterpretasikan model untuk pembahasan statistik.

2. Memuat Data:

Selanjutnya, kita memuat file CSV yang dihasilkan ke dalam DataFrame. Di sinilah data MMM kita berada.

```
4. df = pd.read_csv('/mnt/data/marketing_mix_modeling_data.csv')
```

- Kita menggunakan **"pd.read_csv"** untuk membaca file CSV dan memuatnya ke dalam DataFrame bernama **df**.

3. Pembahasan Data Eksplorasi (EDA):

Sebelum melakukan pemodelan, sangat penting untuk memahami data. Mari kita lihat sekilas dan periksa anomali atau pola apa pun.

```
5. print(df.describe())
```

```
6. print(df.corr())
```

- **df.describe()** memberikan ringkasan statistik dari DataFrame, termasuk rata-rata, deviasi standar, dan kuartil.
- **df.corr()** menghitung matriks korelasi, membantu kita memahami hubungan antara variabel yang berbeda.

4. Mempersiapkan Data untuk Regresi:

Kita memisahkan variabel dependen ('Penjualan') dan variabel independen (pengeluaran pemasaran dan diskon).

```
7. X=df[['TV_Ad_Spend', 'Online_Ad_Spend', 'Radio_Ad_Spend', 'Promotional_Discount']] # Independent variables
```

```
8. y = df['Sales'] # Dependent variable
```

5. Menambahkan Konstanta ke Model:

Untuk model regresi kita, kita menambahkan konstanta ke variabel independen. Ini adalah langkah umum dalam regresi linier untuk menyertakan intersep dalam model.

```
9. X = sm.add_constant(X)
```

6. Membangun Model Regresi:

Sekarang, kita menggunakan regresi kuadrat terkecil biasa (OLS) untuk memodelkan hubungan antara variabel independen dan dependen.

```
10. model = sm.OLS(y, X).fit()
```

- **sm.OLS** digunakan untuk membuat model regresi OLS. Fungsi **fit()** kemudian dipanggil pada model ini untuk menyesuaikannya dengan data.

7. Melihat Hasil Regresi:

Terakhir, kita mencetak ringkasan model regresi kita untuk melihat koefisien dan ukuran statistik lainnya.

```
11. print(model.summary())
```

- **model.summary()** memberikan ringkasan rinci hasil regresi, termasuk koefisien, nilai R-squared, nilai p, dan sebagainya.

LATIHAN 8.2

ATRIBUSI BERBASIS DATA

Tujuan

Pembahasan data perjalanan pelanggan untuk mengaitkan konversi dengan berbagai titik kontak pemasaran menggunakan model probabilitas.

Tugas:

1. Muat data yang dihasilkan ke dalam DataFrame.
2. Lakukan pra-pemrosesan data untuk menyusun data titik kontak.
3. Terapkan model probabilitas untuk menetapkan kredit konversi ke setiap titik kontak.
4. Pembahasan hasilnya untuk mengidentifikasi titik kontak mana yang memiliki pengaruh paling signifikan terhadap konversi.

Langkah-langkah:

1. Memuat Pustaka:

Seperti pada latihan sebelumnya, kita mulai dengan mengimpor pustaka yang diperlukan.

```
1. import pandas as pd
2. from sklearn.preprocessing import MultiLabelBinarizer
3. from sklearn.linear_model import LogisticRegression
```

- **Pandas** digunakan untuk manipulasi data.
- **MultiLabelBinarizer** dari **sklearn.preprocessing** digunakan untuk mengubah data titik kontak menjadi format biner yang sesuai untuk pemodelan.
- **LogisticRegression** dari **sklearn.linear_model** digunakan untuk menjalankan model 'LogisticRegression'.

2. Memuat Data:

Selanjutnya, muat data sintetis untuk atribusi berbasis data.

```
4.df_attribution=pd.read_csv('/mnt/data/data_driven_attribution_data.csv')
```

3. Praproses Data:

Kolom 'Titik Sentuh' berisi daftar titik sentuh, yang perlu diubah ke dalam format yang dapat digunakan untuk pemodelan.

```
5. # Splitting the touchpoint strings into lists
6. df_attribution['Touchpoints']=df_attribution['Touchpoints'].apply(lambda x: x.split(', '))
7. # Using MultiLabelBinarizer to transform the touchpoint lists into binary format
8. mlb = MultiLabelBinarizer()
9. touchpoints_binary=mlb.fit_transform(df_attribution['Touchpoints'])
10.# Creating a DataFrame for the binary touchpoints
11. df_touchpoints=pd.DataFrame(touchpoints_binary,columns=mlb.classes_)
```

4. Mempersiapkan Model:

Sekarang, siapkan data untuk pembahasan regresi logistik, yang akan kita gunakan untuk atribusi.

```
12. X = df_touchpoints # Independent variables (binary touchpoints)
13. y = df_attribution['Conversion'] # Dependent variable (conversion)
```

5. Membangun dan Menyesuaikan Model 'Regresi Logistik':

Dengan data yang telah disiapkan, kita dapat membangun dan menyesuaikan model 'Regresi Logistik'.

```
14. model = LogisticRegression()
15. model.fit(X, y)
```

6. Menginterpretasikan Koefisien Model:

Koefisien dari regresi logistik akan membantu kita memahami dampak setiap titik kontak terhadap kemungkinan konversi.

```
16. coefficients=pd.DataFrame({"Touchpoint":mlb.classes_, "Coefficient":
    model.coef_[0]})
```

7. Mengurutkan Koefisien:

Untuk menafsirkan hasil dengan lebih baik, kami mengurutkan titik kontak berdasarkan koefisiennya. Koefisien yang lebih tinggi menunjukkan dampak positif yang lebih besar pada konversi.

```
17. sorted_coefficients=coefficients.sort_values(by="Coefficient", ascending=False)
```

Mari kita jalankan kode ini untuk pembahasan titik sentuh dan pengaruhnya terhadap konversi dalam dataset sintetis kita.

BAB 9

PEMBAHASAN PERJALANAN PELANGGAN

9.1 PENDAHULUAN

Perjalanan pelanggan modern itu kompleks dan multifaset, melintasi berbagai saluran dan titik kontak dari waktu ke waktu. Perjalanan ini dimulai sejak pelanggan menyadari keberadaan suatu merek atau produk, dan berlanjut melalui pertimbangan, evaluasi, pembelian, dan pengalaman pasca-pembelian. Perilaku konsumen yang terus berkembang ini telah menjadikan pemahaman dan pembahasan perjalanan pelanggan sebagai hal yang penting bagi pemasar untuk mengoptimalkan strategi pemasaran dan memberikan pengalaman pelanggan yang unggul.

Pembahasan perjalanan pelanggan adalah proses pelacakan dan pembahasan bagaimana pelanggan menggunakan kombinasi saluran untuk berinteraksi dengan perusahaan, dan kemudian menggunakan wawasan tersebut untuk memungkinkan keterlibatan pelanggan dengan cara yang paling optimal.

Pembahasan ini mencakup berbagai teknik, termasuk pemetaan perjalanan pelanggan, pembahasan titik kontak, optimasi pemasaran lintas saluran, dan pembahasan jalur pembelian dan atribusi. Dengan memanfaatkan teknik-teknik ini, bisnis dapat memperoleh pandangan holistik tentang perjalanan pelanggan, mengungkap wawasan pelanggan yang tersembunyi, dan mengidentifikasi peluang untuk menyederhanakan perjalanan dan meningkatkan pengalaman pelanggan secara keseluruhan.

Bab ini akan membahas konsep-konsep ini dan memberikan contoh praktis tentang bagaimana bisnis dapat memanfaatkan analitik perjalanan pelanggan untuk mendorong keberhasilan pemasaran.

9.2 PEMETAAN PERJALANAN PELANGGAN

Gambaran Umum Pemetaan Perjalanan Pelanggan

Pemetaan perjalanan pelanggan adalah teknik yang digunakan oleh pemasar untuk menggambarkan secara visual interaksi pelanggan dengan suatu merek di berbagai titik kontak dan saluran. Metode ini membantu menggambarkan jalur pelanggan dari kontak awal hingga proses keterlibatan, pembelian, dan seterusnya.

Peta perjalanan biasanya dibuat dari perspektif pelanggan dan dirancang untuk menggambarkan pengalaman, emosi, dan harapan pelanggan di setiap titik kontak. Pemetaan perjalanan pelanggan dapat digunakan untuk mengungkap momen-momen gesekan atau titik-titik kesulitan dalam perjalanan pelanggan, serta peluang untuk meningkatkan pengalaman pelanggan.

Pemetaan perjalanan pelanggan adalah alat yang ampuh untuk mendorong sentralitas pelanggan dalam suatu organisasi. Ini membantu memecah silo dengan mendorong kolaborasi lintas fungsi dan menumbuhkan pemahaman bersama tentang perjalanan pelanggan di seluruh organisasi. Lebih jauh lagi, ini memberikan wawasan berharga yang dapat memandu

pengambilan keputusan strategis dan menginformasikan desain intervensi pemasaran yang lebih efektif dan personal.

Tahapan Perjalanan Pelanggan: Kesadaran hingga Advokasi

Setiap interaksi yang dilakukan pelanggan dengan suatu merek menambah perjalanan mereka. Perjalanan ini secara umum dapat dibagi menjadi beberapa tahapan yang berbeda:

1. **Kesadaran.** Ini adalah tahap di mana calon pelanggan pertama kali mempelajari tentang suatu merek atau produk. Mereka mungkin menemukan iklan, unggahan media sosial, atau mendengarnya melalui dari mulut ke mulut. Fokus merek pada tahap ini adalah untuk menarik perhatian dan membangkitkan minat.
2. **Pertimbangan.** Setelah memperoleh beberapa pengetahuan tentang merek, calon pelanggan sekarang secara aktif meneliti dan membandingkan pilihan. Mereka mungkin mengunjungi situs web merek, membaca ulasan, atau mencari rekomendasi. Merek harus menawarkan informasi yang berharga dan mudah diakses pada tahap ini untuk memengaruhi keputusan pelanggan.
3. **Pembelian/Keputusan.** Pelanggan telah memutuskan untuk melakukan pembelian. Pengalaman pada tahap ini, termasuk kemudahan proses pembelian dan kualitas layanan pelanggan, dapat sangat memengaruhi persepsi keseluruhan mereka terhadap merek.
4. **Retensi/pasca-pembelian.** Setelah pembelian, perjalanan tidak berakhir. Bagaimana merek mendukung pelanggan, baik melalui layanan purna jual, dukungan garansi, atau sekadar melalui pesan terima kasih, memainkan peran penting dalam menentukan apakah pelanggan akan kembali.
5. **Advokasi.** Tujuan utama bagi banyak merek adalah mengubah pelanggan menjadi pendukung. Pelanggan yang puas dapat berbagi pengalaman positif mereka, merekomendasikan merek kepada orang lain, atau bahkan menulis ulasan yang menguntungkan. Tahap advokasi ini dapat memberikan pertumbuhan organik yang signifikan dan kepercayaan merek.

Alat dan Teknik untuk Pemetaan Perjalanan Pelanggan

Untuk membuat peta perjalanan pelanggan yang efektif, bisnis sering menggunakan campuran alat dan teknik, termasuk yang berikut:

- **Wawancara dan survei.** Umpan balik langsung dari pelanggan dapat memberikan banyak informasi. Merek dapat melakukan wawancara atau mendistribusikan survei untuk mengumpulkan wawasan tentang pengalaman pelanggan di berbagai titik kontak.
- **Pembahasan web dan aplikasi.** Alat seperti Google Analytics dapat memberikan data tentang bagaimana pelanggan berinteraksi dengan platform digital suatu merek, mengungkapkan potensi titik masalah atau tahapan di mana pelanggan mungkin berhenti.
- **Peta panas.** Alat ini secara visual merepresentasikan di mana pengguna mengklik, bergerak, dan menggulir di halaman web, memberikan wawasan tentang perilaku dan preferensi mereka.

- **Platform umpan balik pelanggan.** Platform seperti Userve atau GetSatisfaction memungkinkan pelanggan untuk memberikan umpan balik langsung, melaporkan masalah, atau menyarankan fitur.
- **Lokakarya.** Mengadakan lokakarya internal dengan tim penjualan, pemasaran, dan dukungan dapat membantu mengumpulkan wawasan dari berbagai perspektif departemen, memastikan peta perjalanan yang komprehensif.

Memanfaatkan Peta Perjalanan untuk Pengembangan Strategi

Peta perjalanan, setelah dibuat, dapat berperan penting dalam merumuskan strategi merek:

1. **Mengidentifikasi kesenjangan.** Dengan merepresentasikan perjalanan pelanggan secara visual, merek dapat dengan cepat mengidentifikasi area yang mungkin kurang dalam hal keterlibatan atau dukungan.
2. **Optimalkan titik kontak.** Dengan pandangan yang jelas tentang semua titik kontak, merek dapat mengalokasikan sumber daya secara lebih efektif, memastikan bahwa setiap interaksi menambah nilai pada pengalaman pelanggan.
3. **Tingkatkan personalisasi.** Dengan memahami perilaku, preferensi, dan masalah pelanggan, merek dapat membuat kampanye pemasaran yang lebih personal dan efektif.
4. **Dorong kolaborasi.** Pemahaman bersama tentang perjalanan pelanggan dapat mendorong kolaborasi lintas departemen, memastikan bahwa tim selaras dalam upaya mereka untuk meningkatkan pengalaman pelanggan.
5. **Ukur dampak.** Dengan membandingkan peta perjalanan dari periode yang berbeda, merek dapat mengukur dampak dari setiap perubahan atau peningkatan yang dilakukan.

Kesimpulannya, memahami tahapan perjalanan pelanggan, menggunakan alat untuk pemetaan perjalanan, dan memanfaatkan peta ini untuk pengembangan strategi sangat penting untuk mengoptimalkan pengalaman pelanggan. Wawasan yang diperoleh dapat memandu merek dalam memberikan interaksi yang lebih personal, efisien, dan berdampak di setiap titik kontak.

Konsep Kunci dalam Pemetaan Perjalanan Pelanggan

Pemetaan perjalanan pelanggan adalah proses strategis untuk menangkap total pengalaman pelanggan di semua titik kontak dengan suatu merek. Proses ini melibatkan identifikasi berbagai tahapan yang dilalui pelanggan dalam interaksi mereka dengan merek, dari kontak awal hingga pembelian atau interaksi akhir. Konsep kunci yang terlibat dalam proses ini meliputi:

- **Persona pelanggan.** Persona pelanggan adalah karakter semi-fiktif yang mewakili segmen audiens target perusahaan. Persona ini didasarkan pada data nyata tentang demografi dan perilaku pelanggan, bersama dengan spekulasi yang terinformasi tentang motivasi, tantangan, dan kekhawatiran mereka.

- Titik kontak. Ini adalah titik interaksi antara pelanggan dan merek. Titik kontak dapat terjadi di berbagai saluran (seperti situs web, media sosial, di toko) dan pada berbagai tahapan perjalanan pelanggan.
- Momen kebenaran. Diperkenalkan oleh Procter & Gamble, ini adalah interaksi penting di mana pelanggan menginvestasikan sejumlah besar energi emosional pada hasilnya. Hal ini secara signifikan memengaruhi persepsi pelanggan terhadap merek.
- Titik-titik kesulitan. Ini adalah hambatan atau frustrasi yang dialami pelanggan pada berbagai tahap perjalanan mereka. Mengidentifikasi titik-titik ini dapat membantu merek meningkatkan pengalaman pelanggan mereka.
- Pemetaan emosi. Ini melibatkan penangkapan perjalanan emosional pelanggan bersamaan dengan perjalanan fisik mereka. Ini membantu dalam memahami bagaimana perasaan pelanggan pada berbagai tahap, yang dapat sangat memengaruhi pengalaman keseluruhan mereka.

Pemetaan Perjalanan Pelanggan untuk Perusahaan Ritel

Pemetaan perjalanan pelanggan dapat menjadi alat penting bagi perusahaan ritel dalam memahami pengalaman pelanggan mereka dan mengidentifikasi area untuk perbaikan. Berikut adalah contoh praktis bagaimana perusahaan ritel dapat menggunakan pemetaan perjalanan pelanggan:

1. Pembuatan persona pelanggan. Perusahaan pertama-tama membuat persona pelanggan berdasarkan target audiensnya. Ini dapat mencakup orang tua yang sibuk bekerja, mahasiswa yang memperhatikan harga, atau pensiunan yang lebih tua. Persona ini dibuat menggunakan kombinasi data demografis, analitik perilaku pelanggan, riset pasar, dan umpan balik pelanggan.
2. Mengidentifikasi titik kontak. Perusahaan kemudian mengidentifikasi semua titik kontak potensial yang mungkin dimiliki pelanggan dengan merek tersebut. Ini dapat mencakup melihat iklan online, mengunjungi situs web, menerima buletin email, mengunjungi toko fisik, menghubungi layanan pelanggan, atau melakukan pembelian.
3. Memetakan perjalanan pelanggan. Perusahaan memetakan perjalanan untuk setiap persona, dari kesadaran awal hingga pembelian dan interaksi pasca-pembelian. Mereka mengidentifikasi momen-momen penting—kejadian-kejadian yang memiliki dampak signifikan pada pengalaman pelanggan secara keseluruhan.
4. Mengidentifikasi titik-titik masalah. Perusahaan menggunakan umpan balik pelanggan, pemantauan media sosial, dan pembahasan data untuk mengidentifikasi titik-titik masalah dalam perjalanan pelanggan. Misalnya, mereka mungkin menemukan bahwa pelanggan frustrasi dengan lamanya waktu yang dibutuhkan untuk menerima pesanan online atau kesulitan dalam menemukan produk di toko.
5. Mengembangkan solusi. Perusahaan menggunakan wawasan yang diperoleh dari peta perjalanan pelanggan untuk mengembangkan solusi terhadap titik-titik masalah yang telah diidentifikasi. Ini dapat mencakup pengenalan opsi pengiriman ekspres, peningkatan papan petunjuk di toko, atau pengenalan desain situs web yang lebih ramah pengguna.

Dengan menggunakan pemetaan perjalanan pelanggan, perusahaan ritel dapat memperoleh pemahaman yang lebih baik tentang pengalaman pelanggan mereka, mengidentifikasi area untuk perbaikan, dan mengambil tindakan untuk meningkatkan perjalanan pelanggan secara keseluruhan.

9.3 NAVIGASI TOUCHPOINT PELANGGAN

Gambaran Umum Pembahasan Titik Kontak

Pembahasan titik kontak merupakan komponen penting dari analitik perjalanan pelanggan. Pembahasan ini berfokus pada identifikasi dan evaluasi semua titik interaksi antara pelanggan dan merek, mulai dari fase penemuan awal hingga proses pembelian, dan bahkan hingga interaksi pasca-pembelian.

Titik kontak ini dapat terjadi di berbagai saluran dan mencakup berbagai macam interaksi, seperti menjelajahi situs web perusahaan, berbicara dengan perwakilan penjualan, menerima email pemasaran, berinteraksi dengan merek di media sosial, atau mengunjungi toko fisik. Setiap titik kontak ini dapat secara signifikan memengaruhi persepsi pelanggan terhadap merek, proses pengambilan keputusan mereka, dan pengalaman pelanggan secara keseluruhan.

Pembahasan titik kontak yang efektif melibatkan pemetaan semua titik kontak potensial, memahami bagaimana pelanggan berinteraksi dengan titik kontak ini, dan menilai kualitas interaksi tersebut. Hal ini memerlukan kombinasi metode pengumpulan data, termasuk survei pelanggan, pemantauan media sosial, analitik situs web, dan data penjualan.

Pembahasan titik sentuh sangat berharga untuk mengidentifikasi titik-titik masalah, yang kemudian dapat diatasi untuk meningkatkan kepuasan dan loyalitas pelanggan. Hal ini juga membantu merek memahami titik sentuh mana yang memiliki dampak paling signifikan pada keputusan pelanggan, sehingga memungkinkan mereka untuk memfokuskan sumber daya mereka secara efektif.

Mengidentifikasi Titik Sentuh Utama dalam Perjalanan Pelanggan

Mengenali titik sentuh utama dalam perjalanan pelanggan sangat penting untuk memberikan pengalaman pelanggan yang optimal (lihat Tabel 9.1). Munculnya evolusi digital telah memperkuat titik sentuh ini, sehingga merek perlu memetakan setiap interaksi secara cermat untuk memastikan pemahaman yang holistik. Menggunakan metode seperti survei pelanggan, formulir umpan balik, dan alat analitik digital dapat memberikan wawasan kepada merek tentang kapan dan di mana interaksi ini terjadi, sehingga memungkinkan mereka untuk mengalokasikan sumber daya secara efektif.

Meskipun lanskap digital telah memperluas berbagai titik sentuh secara signifikan, penting untuk menyadari pentingnya interaksi offline dalam membentuk pengalaman pelanggan. Titik kontak offline, yang sering disebut sebagai titik kontak tradisional, memiliki nilai yang sangat besar karena sering menawarkan interaksi langsung yang berpusat pada manusia yang dapat sangat memengaruhi persepsi dan emosi. Misalnya, suasana fisik toko, sikap staf penjualan, atau kemudahan pertanyaan melalui telepon dapat meninggalkan kesan

yang mendalam pada pelanggan. Tabel 9.2 merupakan kelanjutan dari Tabel 9.1, yang menyoroti beberapa potensi titik kontak offline.

Tabel 9.1 Potensi Titik Kontak Utama dalam Perjalanan Pelanggan.

Titik Kontak	Deskripsi
Kunjungan situs web	Pengalaman pertama mengunjungi situs web e-commerce
Pencarian produk	Menggunakan fitur pencarian untuk menemukan produk tertentu
Detail produk	Melihat spesifikasi detail dan ulasan produk
Menambahkan ke keranjang	Memilih barang dan menambahkannya ke keranjang belanja
Pembayaran	Proses penyelesaian pembelian, termasuk memilih opsi pengiriman
Proses pembayaran	Menyelesaikan transaksi keuangan
Email pasca pembelian	Email yang diterima setelah pembelian, termasuk tanda terima dan pemberitahuan pengiriman
Pengiriman produk	Saat pelanggan menerima dan membuka kemasan produk yang dipesan
Pengembalian atau dukungan	Berinteraksi dengan layanan pelanggan untuk dukungan pasca-pembelian atau memproses pengembalian

Tabel 9.2 Potensi Titik Kontak Utama Offline dalam Perjalanan Pelanggan.

Titik Kontak	Deskripsi
Pengalaman di Toko	Suasana, tata letak, dan kemudahan navigasi di dalam toko fisik; termasuk interaksi dengan produk dan perwakilan penjualan
Pertanyaan melalui Telepon	Panggilan telepon untuk menanyakan tentang produk, layanan, atau informasi lainnya
Katalog Fisik	Katalog atau brosur cetak yang diberikan kepada pelanggan
Promosi di Toko	Acara khusus atau promosi yang diadakan di dalam toko fisik
Surat Langsung	Surat fisik atau penawaran promosi yang dikirim ke rumah pelanggan
Demonstrasi Produk	Demonstrasi langsung atau sesi uji coba produk
Dukungan Pelanggan Tatap Muka	Interaksi langsung dengan perwakilan layanan pelanggan atau help desk
Promosi dari Mulut ke Mulut	Rekomendasi atau umpan balik pribadi dari teman, keluarga, atau kenalan
Acara dan Lokakarya	Acara atau lokakarya yang diadakan oleh merek yang dapat dihadiri pelanggan

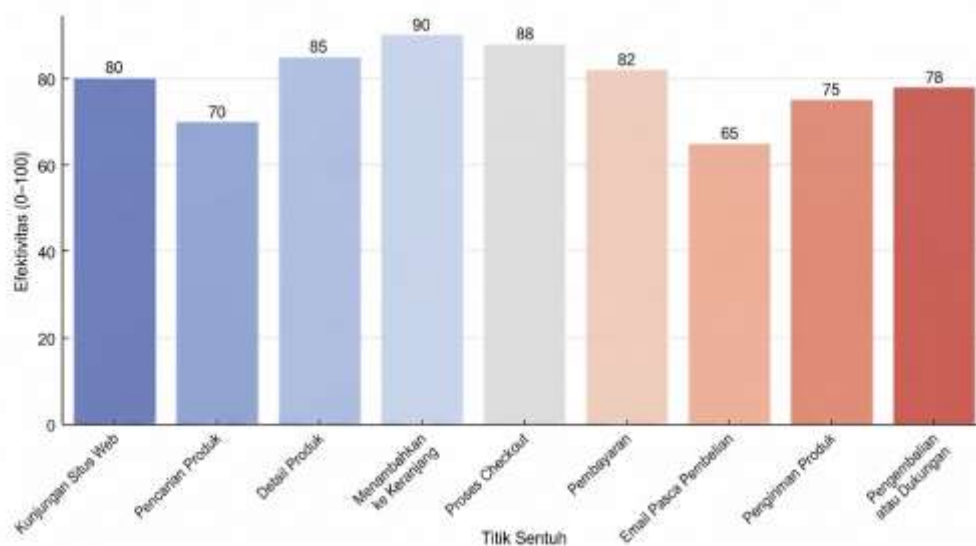
Menggabungkan titik kontak online dan offline menawarkan pandangan komprehensif tentang perjalanan pelanggan yang beragam. Untuk benar-benar menguasai perjalanan pelanggan, merek perlu memastikan kohesi antara titik kontak ini, mendorong transisi yang mulus dan pengalaman yang konsisten.

Mengukur Dampak dan Efektivitas Titik Kontak

Mengingat banyaknya titik kontak dalam perjalanan pelanggan, sangat penting bagi bisnis untuk memastikan dampak dan efektivitas masing-masing titik kontak. Efektivitas titik

kontak dapat diukur dengan pembahasan metrik seperti skor kepuasan pelanggan, tingkat konversi, dan frekuensi keterlibatan. Misalnya, sebuah merek mungkin mengamati tingkat konversi yang lebih tinggi di situs web mereka setelah interaksi melalui kampanye media sosial tertentu, yang menunjukkan efektivitasnya.

Alat analitik canggih, dikombinasikan dengan model atribusi, memungkinkan bisnis untuk mengukur dampak langsung dan tidak langsung dari titik kontak, memberi mereka kemampuan untuk membedakan mana yang paling kuat memengaruhi keputusan pembelian atau meningkatkan loyalitas merek. Wawasan seperti itu sangat penting, tidak hanya untuk mengukur efektivitas titik kontak tetapi juga untuk mengidentifikasi area potensial untuk perbaikan (lihat Gambar 9.1).



Gambar 9.1 Dampak dan Efektivitas Setiap Titik Sentuh.

Kuncinya terletak pada identifikasi metrik yang tepat yang dapat secara akurat menangkap dampak ini:

- **Nilai seumur hidup pelanggan (CLV).** Metrik ini memperkirakan total nilai yang dibawa pelanggan ke bisnis selama durasi hubungan mereka. CLV memberikan pandangan ke dalam nilai jangka panjang pelanggan, dengan memperhitungkan semua titik kontak yang telah mereka interaksi. Dengan membandingkan CLV pelanggan yang telah mengalami titik kontak yang berbeda, pemasar dapat memastikan interaksi mana yang paling signifikan berkontribusi pada profitabilitas jangka panjang.
- **Tingkat churn.** Metrik ini mengungkapkan persentase pelanggan yang mengakhiri hubungan mereka dengan perusahaan selama periode tertentu. Tingkat churn yang tinggi dapat mengindikasikan masalah dengan titik kontak tertentu atau perjalanan pelanggan secara keseluruhan. Dengan mempelajari tingkat churn bersamaan dengan interaksi titik kontak, bisnis dapat mengidentifikasi saluran mana yang mungkin berkontribusi pada pengurangan pelanggan dan memperbaikinya.

- **Skor promotor bersih (NPS).** NPS mengukur kemungkinan pelanggan untuk merekomendasikan merek kepada orang lain. Dengan mengkorelasikan skor NPS dengan titik kontak tertentu, merek dapat mengetahui interaksi mana yang menghasilkan advokasi merek yang lebih tinggi. Penurunan NPS setelah interaksi pada titik kontak tertentu dapat menjadi tanda peringatan, yang menandakan perlunya penyelidikan lebih lanjut dan kemungkinan perbaikan.
- **Skor upaya pelanggan (Customer effort score/CES).** CES mengukur kemudahan pelanggan mencapai tujuan mereka dengan suatu merek, baik itu membeli produk atau menyelesaikan masalah. CES tinggi setelah interaksi dapat menunjukkan adanya hambatan pada titik sentuh tersebut, sehingga perlu dilakukan optimasi.
- **Nilai pesanan rata-rata (Average order value/AOV).** Melacak AOV bersamaan dengan data titik sentuh dapat memberikan wawasan tentang saluran mana yang mendorong transaksi bernilai lebih tinggi. Misalnya, jika kampanye email tertentu menghasilkan lonjakan AOV, hal itu menunjukkan pengaruh kuat kampanye tersebut terhadap perilaku pembelian.
- **Tingkat keterlibatan (engagement rate).** Biasanya digunakan untuk saluran digital seperti media sosial atau kampanye email, tingkat keterlibatan mengukur persentase audiens yang berinteraksi dengan konten. Tingkat keterlibatan yang tinggi sering kali menandakan efektivitas suatu titik sentuh dalam menarik minat pelanggan dan mendorong tindakan.

Singkatnya, dengan menggunakan metrik ini, bisnis dapat memperoleh gambaran komprehensif tentang efektivitas titik sentuh mereka. Dengan secara teratur menilai angka-angka ini, mereka dapat terus mengoptimalkan perjalanan pelanggan, memastikan pertumbuhan berkelanjutan dan loyalitas merek. Ingat, dalam dunia pemasaran yang serba cepat, apa yang diukur akan ditingkatkan.

Meningkatkan dan Memperbaiki Titik Kontak

Dengan pemahaman yang jelas tentang dampak dan efektivitas titik kontak, bisnis dapat memulai proses perbaikan. Peningkatan berkelanjutan sangat penting dalam lanskap interaksi pelanggan yang dinamis. Titik kontak yang terbukti efektif hari ini mungkin akan berkembang atau menjadi usang besok. Mekanisme umpan balik, seperti ulasan pelanggan atau NPS, memberikan wawasan langsung ke area peningkatan.

Selain itu, pengujian A/B, terutama di saluran digital, memungkinkan bisnis untuk membandingkan kinerja strategi titik kontak yang berbeda, sehingga membantu dalam memperbaiki strategi yang paling berdampak. Misalnya, jika pelanggan sering meninggalkan keranjang belanja online, meningkatkan kesederhanaan dan kecepatan proses pembayaran dapat mengurangi tingkat putus sekolah.

Mengintegrasikan Titik Sentuh untuk Pengalaman yang Lancar

Karena pelanggan sering berinteraksi dengan merek di berbagai titik sentuh, pengalaman yang terputus-putus dapat mengakibatkan kebingungan atau ketidakpuasan. Oleh karena itu, penting untuk mengintegrasikan titik sentuh untuk memberikan perjalanan pelanggan yang konsisten dan lancar. Repositori data terpusat atau sistem manajemen

hubungan pelanggan (CRM) terintegrasi dapat menawarkan bisnis pandangan holistik tentang interaksi pelanggan di berbagai titik sentuh. Integrasi ini memastikan bahwa jika pelanggan menanyakan tentang suatu produk di media sosial dan kemudian mengunjungi situs web merek, pengalaman mereka tetap konsisten. Lebih lanjut, teknologi canggih seperti AI dapat memanfaatkan data terintegrasi ini untuk memprediksi interaksi pelanggan di masa mendatang, memungkinkan merek untuk secara proaktif menyesuaikan titik sentuh, sehingga meningkatkan pengalaman pelanggan secara keseluruhan.

Konsep Kunci dalam Pembahasan Titik Sentuh

Konsep kunci dalam pembahasan titik sentuh didasarkan pada pemahaman bagaimana pelanggan berinteraksi dengan merek di berbagai tahap perjalanan mereka:

- **Titik Sentuh.** Titik sentuh adalah setiap titik interaksi antara pelanggan dan merek. Ini termasuk iklan, interaksi media sosial, pertemuan layanan pelanggan, pengalaman di toko, dan media lain apa pun yang digunakan pelanggan untuk berinteraksi dengan merek.
- **Perjalanan pelanggan.** Perjalanan pelanggan adalah rangkaian lengkap pengalaman yang dilalui pelanggan saat berinteraksi dengan perusahaan atau merek. Ini mencakup semua tahapan mulai dari kesadaran dan pertimbangan awal, melalui pembelian, hingga interaksi pasca-pembelian.
- **Titik masalah.** Titik masalah adalah setiap aspek interaksi pelanggan dengan merek yang menyebabkan frustrasi atau ketidakpuasan. Mengidentifikasi dan mengatasi hal ini merupakan bagian penting dari pembahasan titik sentuh.
- **Momen kebenaran.** Ini adalah titik sentuh penting yang secara signifikan memengaruhi persepsi pelanggan terhadap merek. Ini bisa menjadi momen penentu bagi kepuasan dan loyalitas pelanggan.
- **Pengalaman omnichannel.** Di era digital saat ini, pelanggan sering berinteraksi dengan merek melalui berbagai saluran (misalnya, online, di toko, media sosial). Pembahasan titik sentuh yang efektif harus mempertimbangkan semua saluran ini dan bagaimana kontribusinya terhadap pengalaman pelanggan secara keseluruhan.

Contoh Praktis: Pembahasan Titik Sentuh untuk Perusahaan E-Commerce

Pada bagian ini, kita akan mengeksplorasi bagaimana sebuah perusahaan e-commerce, E-shop, menggunakan pembahasan titik sentuh untuk lebih memahami interaksi dan pengalaman pelanggan mereka. Sebagai permulaan, E-shop memetakan semua titik sentuh yang berbeda di mana pelanggan berinteraksi dengan merek mereka. Ini termasuk situs web, komunikasi email, saluran media sosial, interaksi layanan pelanggan, dan platform digital lainnya seperti aplikasi seluler. Setiap titik sentuh dipembahasan secara detail, dengan mempertimbangkan faktor-faktor seperti antarmuka pengguna, kemudahan navigasi, waktu respons, dan keterlibatan yang dipersonalisasi.

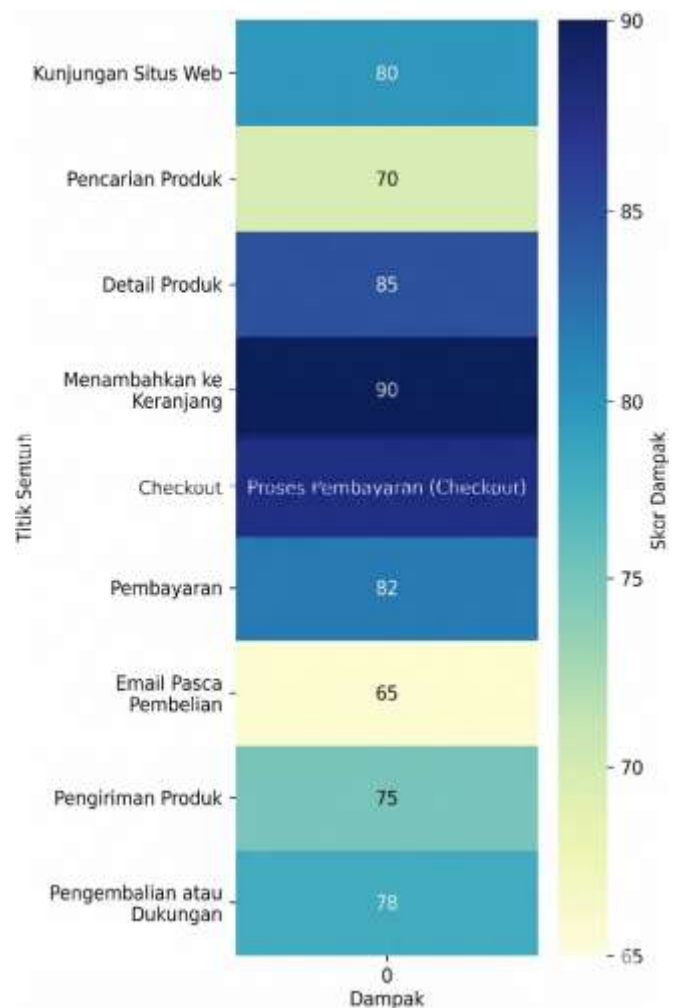
Kemudian, toko online mengumpulkan data dari berbagai sumber, seperti analitik situs web, survei pelanggan, alat pemantauan media sosial, dan catatan layanan pelanggan, untuk memahami perilaku dan persepsi pelanggan di setiap titik kontak. Mereka memberikan perhatian khusus pada titik-titik permasalahan—area di mana pelanggan mengungkapkan

ketidakpuasan atau frustrasi. Misalnya, jika pelanggan mengeluh tentang proses pembayaran yang terlalu rumit, ini menjadi area fokus untuk perbaikan.

Selanjutnya, toko online mengidentifikasi momen-momen penting—tahap-tahap krusial dalam perjalanan pelanggan yang secara signifikan memengaruhi keputusan pelanggan untuk membeli atau membeli kembali. Ini bisa berupa proses pemilihan produk, kemudahan menemukan informasi, atau kualitas interaksi layanan pelanggan.

Berdasarkan pembahasan ini, toko online menerapkan perubahan untuk meningkatkan pengalaman pelanggan di setiap titik kontak. Mereka dapat mendesain ulang bagian-bagian situs web agar navigasi lebih mudah, menyempurnakan strategi komunikasi email mereka untuk memberikan konten yang lebih personal, atau berinvestasi dalam pelatihan layanan pelanggan yang lebih baik. Mereka terus memantau umpan balik dan metrik pelanggan untuk menilai efektivitas perubahan ini dan melakukan perbaikan lebih lanjut jika diperlukan.

Pembahasan titik kontak toko online (lihat Gambar 9.2) menunjukkan bagaimana pemantauan sistematis interaksi pelanggan di semua titik kontak dapat memberikan wawasan berharga untuk meningkatkan pengalaman pelanggan dan meningkatkan kinerja bisnis.

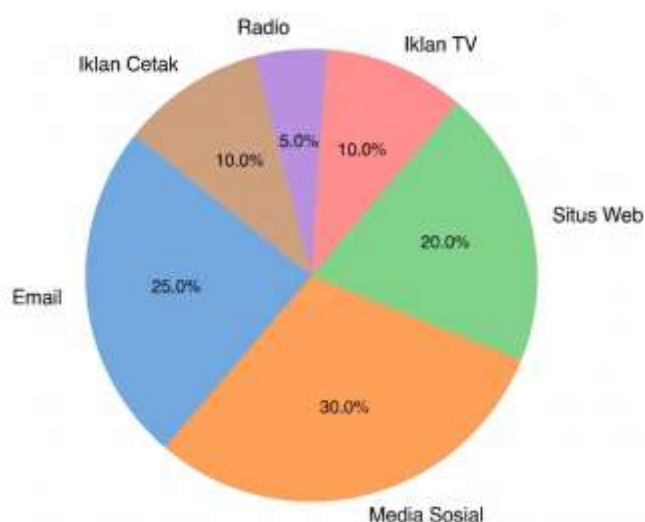


Gambar 9.2 Pembahasan Titik Sentuh untuk Perusahaan E-Commerce, Menandai Titik Sentuh dengan Dampak Tinggi dan Rendah.

9.4 OPTIMASI PEMASARAN LINTAS SALURAN

Gambaran Umum Optimasi Pemasaran Lintas Saluran

Optimasi pemasaran lintas saluran adalah strategi pemasaran tingkat lanjut yang melibatkan penciptaan pengalaman pelanggan yang konsisten dan lancar di berbagai saluran. Pelanggan saat ini berinteraksi dengan merek melalui berbagai titik sentuh: situs web, aplikasi seluler, media sosial, email, toko fisik, dan banyak lagi. Masing-masing saluran ini menawarkan peluang unik untuk terhubung dengan pelanggan dan memengaruhi keputusan pembelian mereka.



Gambar 9.3 Berbagai Saluran Pemasaran.

Inti dari optimasi pemasaran lintas saluran adalah pemahaman bahwa pelanggan bergerak secara fleksibel di berbagai saluran dan mengharapkan pengalaman yang konsisten di mana pun mereka berinteraksi dengan merek. Mereka mungkin menelusuri produk di aplikasi seluler, mencari saran di media sosial, membandingkan harga di situs web, dan akhirnya melakukan pembelian di toko fisik. Setiap interaksi ini berkontribusi pada persepsi keseluruhan mereka terhadap merek dan keputusan mereka untuk membeli.

Optimasi pemasaran lintas saluran melibatkan pengintegrasian dan koordinasi upaya pemasaran di seluruh saluran ini untuk memastikan pengalaman pelanggan yang konsisten dan personal. Hal ini membutuhkan pemahaman mendalam tentang perilaku dan preferensi pelanggan, kemampuan analitik data yang kuat untuk melacak dan pembahasan interaksi pelanggan di seluruh saluran, dan ketangkasan untuk menanggapi kebutuhan pelanggan secara real-time (lihat Gambar 9.3).

Dengan mengoptimalkan upaya pemasaran di seluruh saluran, perusahaan dapat melibatkan pelanggan secara lebih efektif, memberikan pengalaman yang lebih personal, dan pada akhirnya mendorong hasil bisnis yang lebih baik. Strategi pemasaran lintas saluran yang sukses dapat menghasilkan peningkatan keterlibatan pelanggan, tingkat konversi yang lebih tinggi, dan peningkatan loyalitas pelanggan.

Kebutuhan akan Strategi Lintas Saluran Terpadu

Di era konektivitas saat ini, konsumen berinteraksi dengan merek melalui berbagai saluran seperti situs web, aplikasi seluler, media sosial, dan toko fisik. Setiap interaksi ini dapat membentuk persepsi, loyalitas, dan keputusan pembelian konsumen. Pendekatan yang terfragmentasi, di mana setiap saluran beroperasi secara terpisah, dapat menyebabkan pesan yang tidak konsisten, potensi peluang yang terlewatkan, dan kebingungan pelanggan.

Namun, strategi lintas saluran terpadu memastikan pesan yang konsisten dan sinkron, menciptakan narasi merek yang kohesif terlepas dari medianya. Strategi holistik seperti itu sangat penting bagi merek untuk mempertahankan relevansi, karena pelanggan mengharapkan pengalaman yang mulus yang beresonansi terlepas dari titik interaksi mereka.

Selain itu, strategi yang harmonis dapat mengoptimalkan pengeluaran pemasaran, meningkatkan kepuasan pelanggan, dan menawarkan ROI yang lebih baik dengan memanfaatkan kekuatan setiap saluran secara bersamaan daripada secara terpisah.

Untuk memberikan perspektif perbandingan yang jelas, Tabel 9.3 menguraikan keuntungan dan kerugian dari penggunaan strategi lintas saluran tunggal versus terintegrasi, menyoroti potensi dampak masing-masing terhadap pengalaman pelanggan dan efektivitas pemasaran.

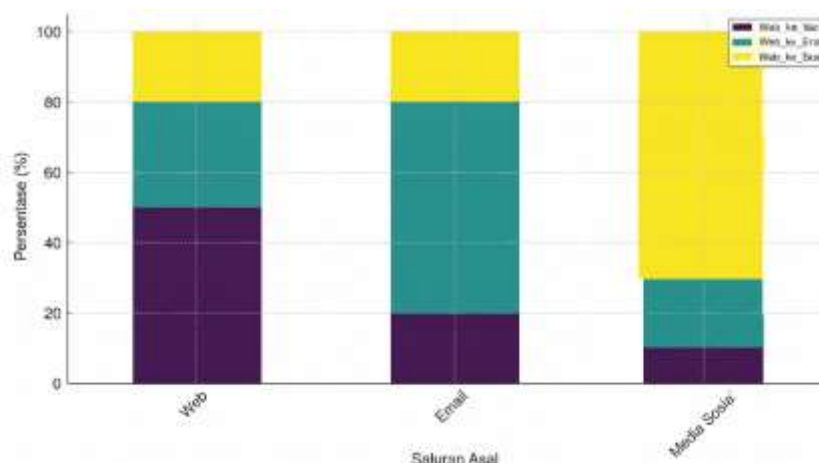
Tabel 9.3 Keuntungan dan Kerugian Strategi Tunggal Dibandingkan Strategi Lintas Saluran.

Strategi	Keunggulan	Kelemahan
Saluran tunggal	Lebih mudah dikelola dan dioptimalkan Pendekatan terfokus dapat menghasilkan wawasan yang lebih mendalam untuk saluran spesifik tersebut	Tidak mendapatkan wawasan dari saluran lain Mungkin tidak memberikan gambaran komprehensif tentang perilaku pelanggan
Lintas saluran terintegrasi	Memberikan pandangan holistik tentang perjalanan pelanggan Memungkinkan keterlibatan yang lebih personal dan tepat waktu	Membutuhkan alat dan keahlian yang lebih canggih Risiko kelebihan informasi jika tidak dikelola dengan benar

Pembahasan Perilaku dan Pola Lintas Saluran

Memahami perilaku pelanggan di berbagai saluran sangat penting untuk strategi terpadu yang efektif. Ini melibatkan pelacakan dan pembahasan bagaimana konsumen berpindah antar saluran, preferensi mereka di setiap saluran, dan urutan interaksi mereka. Misalnya, seorang konsumen mungkin menemukan produk di media sosial, menelitinya di situs web perusahaan, tetapi menyelesaikan pembelian mereka di toko fisik. Mengenali pola-pola tersebut membantu dalam memprediksi perilaku pelanggan, memungkinkan pemasar untuk menyesuaikan strategi mereka untuk memfasilitasi dan mengoptimalkan perjalanan lintas saluran ini. Alat analitik canggih dapat memberikan wawasan mendalam tentang perilaku ini, menunjukkan di mana konsumen berhenti, di mana mereka paling banyak terlibat, dan potensi alasan untuk perilaku tersebut (lihat Gambar 9.4).

Wawasan yang dapat ditindaklanjuti yang diperoleh dari pembahasan perilaku lintas saluran merupakan kunci untuk beralih dari sekadar pengamatan ke optimasi strategis. Wawasan ini, jika diterjemahkan secara efektif, dapat berfungsi sebagai peta jalan, mengarahkan pemasar tentang ke mana harus mengalokasikan sumber daya, titik kontak mana yang perlu diperkuat, dan bagaimana mengatasi potensi jebakan dalam perjalanan pelanggan. Misalnya, menggunakan teknik seperti pembahasan urutan dapat membantu pemasar mengidentifikasi jalur umum yang diambil pelanggan, menyoroti urutan optimal yang mengarah pada konversi dan urutan di mana pelanggan sering berhenti.



Gambar 9.4 Perilaku dan Pola Lintas Saluran.

Metodologi lain adalah pembahasan kluster, yang mengelompokkan pelanggan berdasarkan perilaku lintas saluran yang serupa, memberdayakan pemasar untuk membuat strategi yang dipersonalisasi untuk setiap segmen. Lebih lanjut, model atribusi multi-sentuhan dapat menunjukkan saluran mana yang memiliki pengaruh paling signifikan pada proses pengambilan keputusan pelanggan. Dengan wawasan yang dapat ditindaklanjuti tersebut, bisnis dapat secara proaktif memperbaiki strategi mereka, memastikan mereka tidak hanya memenuhi tetapi juga mengantisipasi dan melampaui harapan pelanggan.

Dengan menyematkan metodologi ini ke dalam proses analitis mereka, merek dapat mengubah data menjadi senjata strategis, selalu selangkah lebih maju dalam permainan keterlibatan pelanggan yang terus berkembang.

Alat untuk Pembahasan Lintas Saluran

Untuk pembahasan perilaku lintas saluran secara efektif, alat dan platform yang andal sangat diperlukan. Banyak platform analitik modern menawarkan kemampuan pelacakan multi-saluran. Google Analytics, misalnya, memberikan wawasan tentang bagaimana pengguna menavigasi berbagai saluran sebelum menyelesaikan tindakan yang diinginkan. Platform yang lebih khusus seperti Adobe Experience Cloud menawarkan serangkaian solusi yang dirancang untuk mengukur dan mengoptimalkan interaksi pelanggan di berbagai saluran.

Selain itu, sistem CRM seperti Salesforce atau HubSpot dapat mengintegrasikan data dari berbagai titik kontak, memberikan merek pandangan komprehensif tentang perjalanan pelanggan. Menggunakan alat-alat ini dapat memberikan bisnis wawasan yang dapat ditindaklanjuti, mulai dari menentukan titik kontak yang sukses hingga mengidentifikasi area yang perlu diperbaiki (lihat Tabel 9.4).

Tabel 9.4 Perbandingan Alat yang Tersedia untuk Pembahasan Lintas Saluran.

Alat	Fitur	Harga	Kemudahan Penggunaan
ChannelMaster	Pembahasan waktu nyata Segmentasi	RpRpRp	Sedang

	Pelacakan ROI		
OptiFlow	Laporan otomatis Pemetaan perjalanan pelanggan Segmentasi	RpRp	Mudah
JourneyScope	Visualisasi tingkat lanjut Pembahasan waktu nyata Pelacakan konversi	RpRpRpRp	Sulit
TrackPulse	Pelacakan ROI Laporan otomatis Visualisasi tingkat lanjut	Rp	Mudah

Studi Kasus: Kampanye Lintas Saluran yang Sukses

Kampanye **“The Ten” Nike**. Nike berkolaborasi dengan desainer Virgil Abloh untuk menciptakan perpaduan unik antara keterlibatan online dan offline. Mereka menggunakan aplikasi SNKRS mereka untuk konten eksklusif dan rilis awal, ditambah dengan lokakarya pop-up di kota-kota besar di seluruh dunia. Pendekatan multi-cabang ini membuat konsumen merasa terlibat dan terhubung di ruang digital dan fisik, yang mengarah pada peningkatan loyalitas merek dan penjualan yang substansial.

Pengalaman toko digital Sephora. Sephora mengintegrasikan pengalaman online dan di toko dengan mulus. Di toko, mereka memperkenalkan Sephora + Pantone Color IQ, sebuah layanan yang memindai permukaan kulit Anda dan mencocokkan warna alas bedak yang tersedia di toko. Pelanggan kemudian dapat menyimpan warna ini ke profil online mereka, memfasilitasi pembelian online di masa mendatang. Sinergi antara pengalaman fisik dan digital ini menghasilkan peningkatan kepuasan pelanggan dan pengalaman berbelanja yang lebih efisien.

Studi kasus ini menggarisbawahi pentingnya dan efektivitas strategi lintas saluran terpadu. Merek yang berhasil menyatukan pengalaman online dan offline akan memperoleh keuntungan dalam hal kepercayaan pelanggan, loyalitas, dan pendapatan.

Konsep Kunci dalam Optimalisasi Pemasaran Lintas Saluran

Optimasi pemasaran lintas saluran melibatkan beberapa konsep kunci:

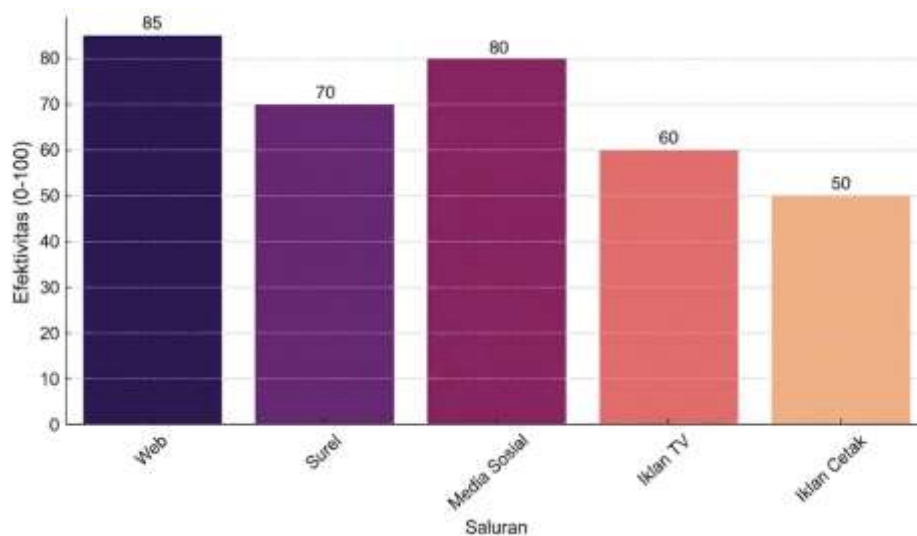
- **Pengalaman omnichannel.** Istilah omnichannel mengacu pada jenis ritel yang mengintegrasikan berbagai metode belanja yang tersedia bagi konsumen (online, di toko fisik, atau melalui telepon). Pengalaman pelanggan berada di pusat pendekatan ini, di mana tujuannya adalah untuk memberikan pengalaman yang mulus dan konsisten di semua saluran.
- **Integrasi data pelanggan.** Untuk optimalisasi lintas saluran yang sukses, merek perlu mengintegrasikan data pelanggan dari semua saluran. Ini termasuk perilaku online,

data transaksi, dan bahkan interaksi offline. Pandangan terpadu tentang pelanggan membantu merek memberikan pengalaman yang dipersonalisasi di seluruh saluran.

- Personalisasi. Personalisasi adalah proses penyampaian konten individual melalui pembahasan data dan teknologi digital. Ini adalah aspek penting dari optimasi pemasaran lintas saluran karena memungkinkan bisnis untuk menyesuaikan pesan dan penawaran mereka berdasarkan perilaku dan preferensi pelanggan di masa lalu.
- Manajemen interaksi waktu nyata. Di era pemasaran digital, manajemen interaksi waktu nyata menjadi sangat penting. Merek perlu menanggapi tindakan pelanggan dengan segera, baik itu komentar di media sosial, pencarian produk di situs web, atau pembelian di toko.
- Pemodelan atribusi. Pemodelan atribusi melibatkan penentuan titik sentuh pemasaran mana yang berkontribusi pada konversi dan sejauh mana. Memahami hal ini membantu pemasar mengoptimalkan upaya pemasaran mereka di berbagai saluran.

Optimasi Pemasaran Lintas Saluran untuk Perusahaan Perjalanan

Industri perjalanan adalah salah satu industri di mana perjalanan pelanggan seringkali kompleks dan beragam, melibatkan berbagai titik sentuh di berbagai saluran, menjadikannya contoh utama untuk memahami optimasi pemasaran lintas saluran.



Gambar 9.5 Optimalisasi Pemasaran Lintas Saluran untuk TravelCo.

Misalkan kita mempertimbangkan perusahaan hipotetis, TravelCo. TravelCo adalah perusahaan mapan yang menawarkan berbagai layanan mulai dari pemesanan penerbangan dan reservasi hotel hingga paket liburan. Mereka memiliki kehadiran online yang signifikan dengan situs web yang ramah pengguna, aplikasi seluler, dan profil media sosial yang aktif. Mereka juga memiliki kehadiran offline melalui agen perjalanan lokal dan surat langsung.

TravelCo memutuskan untuk menerapkan optimasi pemasaran lintas saluran. Langkah pertama mereka adalah mengintegrasikan data pelanggan dari semua saluran untuk menciptakan pandangan terpadu tentang pelanggan mereka. Ini termasuk melacak perilaku

online di situs web dan aplikasi seluler mereka, data interaksi dari media sosial, dan data offline dari agen perjalanan.

Dengan data ini, mereka memanfaatkan algoritma pembelajaran mesin untuk memprediksi preferensi pelanggan dan mempersonalisasi penawaran mereka. Misalnya, jika pelanggan sering mencari destinasi pantai di situs web mereka, TravelCo dapat mempersonalisasi kampanye email mereka untuk menyertakan lebih banyak paket liburan pantai. TravelCo juga menerapkan manajemen interaksi waktu nyata. Jika pelanggan berkomentar pada postingan media sosial mereka menanyakan tentang paket liburan tertentu, mereka segera merespons dengan informasi yang dibutuhkan.

Selain itu, mereka memberikan rekomendasi yang dipersonalisasi berdasarkan riwayat interaksi pelanggan. Terakhir, TravelCo menggunakan pemodelan atribusi untuk memahami saluran mana yang lebih berkontribusi pada konversi mereka. Mereka memperhatikan bahwa pelanggan yang berinteraksi dengan kampanye email mereka lebih cenderung memesan paket liburan.

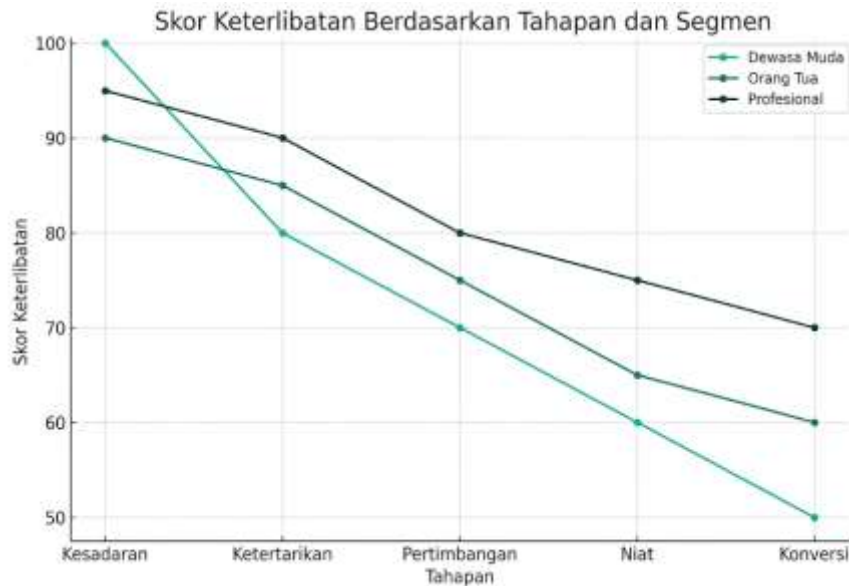
Oleh karena itu, mereka memutuskan untuk berinvestasi lebih banyak dalam mengoptimalkan upaya pemasaran email mereka(. Dengan cara ini, dengan menerapkan optimasi pemasaran lintas saluran, TravelCo dapat meningkatkan pengalaman pelanggan dan meningkatkan konversi mereka.

9.5 OPTIMASI ATRIBUSI PERJALANAN PEMBELIAN

Gambaran Umum Pembahasan Jalur Pembelian dan Atribusi

Jalur pembelian, yang sering disebut sebagai perjalanan pelanggan, mewakili serangkaian langkah yang diambil konsumen mulai dari kesadaran awal akan kebutuhan atau keinginan hingga pembelian akhir. Kompleksitas perjalanan ini telah meningkat seiring dengan proliferasi saluran dan titik kontak digital, sehingga penting bagi bisnis untuk memahami dan memetakan perjalanan ini untuk mengoptimalkan strategi pemasaran mereka.

Pembahasan atribusi memainkan peran penting dalam memahami jalur pembelian. Atribusi adalah proses mengidentifikasi serangkaian tindakan pengguna, atau peristiwa, yang berkontribusi dalam beberapa cara terhadap hasil yang diinginkan, dan kemudian menetapkan nilai untuk setiap peristiwa ini. Dalam konteks pemasaran, ini adalah praktik menentukan peran yang dimainkan oleh berbagai saluran dalam menginformasikan, membujuk, dan mengkonversi konsumen.



Gambar 9.6 Corong Jalur Pembelian Khas.

Misalnya, pelanggan mungkin pertama kali mengetahui suatu produk melalui iklan media sosial, meneliti produk di situs web perusahaan, membaca ulasan di situs pihak ketiga, dan akhirnya melakukan pembelian melalui promosi email yang ditargetkan. Pembahasan atribusi dalam konteks ini bertujuan untuk memahami pengaruh masing-masing titik kontak ini terhadap keputusan pembelian akhir (lihat Gambar 9.6).

Memahami alur pembelian dan melakukan pembahasan atribusi sangat penting untuk mengoptimalkan pengeluaran pemasaran, mempersonalisasi interaksi pelanggan, dan pada akhirnya mendorong konversi dan loyalitas pelanggan.

Memahami Alur Pembelian: Variasi dan Kompleksitas

Di era digital, alur pembelian tidak linier. Perpaduan titik sentuh digital dan fisik membuat perjalanan konsumen lebih rumit dari sebelumnya. Secara tradisional, perjalanan dari kesadaran hingga pembelian dipandang sebagai perkembangan lurus melalui corong. Namun, saat ini, konsumen mungkin bersilasi antar tahapan, dipengaruhi oleh berbagai titik sentuh, mulai dari ulasan online hingga pengalaman di toko. Variasi ini muncul dari perbedaan preferensi konsumen, pengaruh eksternal, dan sifat pembelian itu sendiri. Misalnya, membeli rumah akan melibatkan alur yang lebih berbelit-belit daripada membeli buku. Pemahaman tentang berbagai variasi dan kompleksitas ini sangat penting bagi pemasar untuk menciptakan strategi yang disesuaikan yang sesuai dengan perilaku dan preferensi konsumen individu.

Teknik Kuantitatif untuk Pembahasan Jalur

Teknik kuantitatif menyediakan pendekatan berbasis data untuk memahami dan memetakan jalur pembelian (lihat Tabel 9.5). Algoritma pembelajaran mesin, misalnya, dapat digunakan untuk melacak jejak digital, mengidentifikasi pola dalam perilaku konsumen. Pembahasan urutan dapat digunakan untuk menentukan urutan di mana konsumen berinteraksi dengan titik kontak, menawarkan wawasan tentang preferensi dan kebiasaan mereka. Selain itu, teknik pengelompokan dapat mensegmentasikan konsumen berdasarkan kesamaan dalam jalur pembelian mereka, memungkinkan strategi pemasaran yang lebih

personal. Dengan menggunakan teknik analitik canggih ini, bisnis dapat memperoleh wawasan yang dapat ditindaklanjuti, memprediksi perilaku di masa mendatang, dan menyederhanakan jalur pembelian.

Tabel 9.5 Gambaran Umum Teknik Kuantitatif yang Digunakan untuk Pembahasan Jalur.

Teknik	Deskripsi
Pembahasan urutan	Pembahasan urutan peristiwa atau tahapan dalam perjalanan pelanggan
Rantai Markov	Model statistik untuk merepresentasikan transisi antar tahapan
Model probabilistik	Model untuk memprediksi kemungkinan jalur atau hasil tertentu
Pembahasan kelangsungan hidup	Pembahasan untuk memperkirakan waktu hingga satu atau lebih peristiwa terjadi
Pembahasan kluster	Pengelompokan perjalanan pelanggan ke dalam kluster berdasarkan kesamaan

Implikasi untuk Pemodelan Atribusi

Dengan jalur pembelian yang rumit, menentukan titik sentuh mana yang secara signifikan memengaruhi keputusan konsumen menjadi sangat penting. Di sinilah pemodelan atribusi berperan. Seiring jalur menjadi lebih kompleks, model tradisional, seperti atribusi klik pertama atau klik terakhir, mungkin menawarkan pandangan yang terlalu sederhana tentang perjalanan. Skenario modern membutuhkan model yang lebih rumit yang mempertimbangkan seluruh perjalanan. Model atribusi berbasis data, misalnya, pembahasan kumpulan data besar untuk menentukan dampak tepat dari setiap titik sentuh. Memahami kerumitan dan nuansa jalur memastikan bahwa model atribusi yang dipilih sesuai dengan pengalaman konsumen yang sebenarnya, sehingga menghasilkan wawasan yang lebih akurat dan dapat ditindaklanjuti.

Strategi untuk Mempersingkat dan Mengoptimalkan Jalur Pembelian

Mengoptimalkan jalur pembelian dapat menghasilkan konversi yang lebih cepat dan peningkatan kepuasan pelanggan. Beberapa strategi dapat digunakan untuk tujuan ini:

- **Pengalaman tanpa hambatan.** Penyederhanaan pengalaman digital dengan memastikan waktu pemuatan yang cepat, navigasi yang intuitif, dan proses pembayaran yang mudah dapat secara signifikan mengurangi angka pelanggan yang meninggalkan toko.
- **Personalisasi.** Menyesuaikan konten dengan preferensi konsumen individu, menggunakan wawasan berbasis data, dapat membuat perjalanan pembelian lebih menarik dan efisien.
- **Konsistensi omnichannel.** Memastikan konsistensi di semua titik kontak, baik di toko, di situs web, atau aplikasi seluler, dapat mencegah kebingungan dan menumbuhkan kepercayaan.
- **Konten edukatif.** Memberikan konsumen detail produk, ulasan, dan perbandingan yang komprehensif dapat mempercepat proses pengambilan keputusan.

Dengan mengadopsi strategi ini, bisnis dapat menciptakan jalur pembelian yang lebih efisien dan lancar, yang mengarah pada peningkatan konversi dan peningkatan loyalitas pelanggan.

Konsep Kunci dalam Pembahasan Jalur Pembelian dan Atribusi

Memahami konsep kunci dalam pembahasan jalur pembelian dan atribusi dapat memberdayakan bisnis untuk mengoptimalkan strategi pemasaran mereka dan meningkatkan pengalaman pelanggan. Jalur pembelian sering divisualisasikan sebagai corong, dimulai dengan kesadaran dan minat di bagian atas, diikuti oleh pertimbangan, niat, evaluasi, dan akhirnya pembelian di bagian bawah. Setiap tahap mewakili titik kontak yang berbeda di mana konsumen berinteraksi dengan merek atau produk. Pembahasan atribusi digunakan untuk memberikan kredit kepada berbagai titik kontak ini berdasarkan dampaknya pada keputusan pembelian akhir. Ada beberapa model untuk pembahasan atribusi:

- **Atribusi klik terakhir.** Memberikan kredit pada titik kontak terakhir sebelum pembelian. Model ini sederhana, tetapi dapat terlalu menekankan peran titik kontak terakhir dan mengabaikan pengaruh interaksi sebelumnya.
- **Atribusi klik pertama.** Memberikan kredit pada titik kontak pertama yang mengarahkan pelanggan ke produk atau layanan. Mirip dengan model klik terakhir, ini dapat terlalu menyederhanakan perjalanan pembelian.
- **Atribusi linier.** Memberikan kredit yang sama untuk setiap titik kontak. Model ini mengakui bahwa setiap interaksi memainkan peran tetapi tidak memperhitungkan dampak yang bervariasi dari berbagai titik kontak.
- **Atribusi peluruhan waktu.** Memberikan kredit lebih besar kepada titik kontak yang terjadi lebih dekat dengan waktu pembelian. Model ini mengakui bahwa interaksi selanjutnya cenderung lebih berpengaruh dalam keputusan akhir.
- **Atribusi berbasis data.** Menggunakan algoritma untuk memberikan kredit kepada titik kontak berdasarkan bagaimana kontribusinya terhadap kemungkinan pembelian. Model ini membutuhkan data dan kemampuan analitis yang signifikan tetapi memberikan pandangan yang lebih bernuansa tentang perjalanan pelanggan.

Memahami konsep-konsep kunci ini sangat penting bagi bisnis untuk secara akurat mengukur efektivitas upaya pemasaran mereka dan untuk membuat keputusan yang tepat tentang di mana harus menginvestasikan sumber daya mereka.

Jalur Pembelian dan Pembahasan Atribusi untuk Perusahaan Perangkat Lunak

Di bidang perusahaan perangkat lunak yang kompetitif, memahami jalur pembelian dan menerapkan pembahasan atribusi yang efektif sangat penting untuk keberhasilan. Sebagai contoh, sebuah perusahaan perangkat lunak B2B mungkin memiliki proses penjualan yang kompleks yang melibatkan banyak titik kontak, termasuk iklan digital, kampanye email, webinar, demo produk, dan panggilan penjualan.

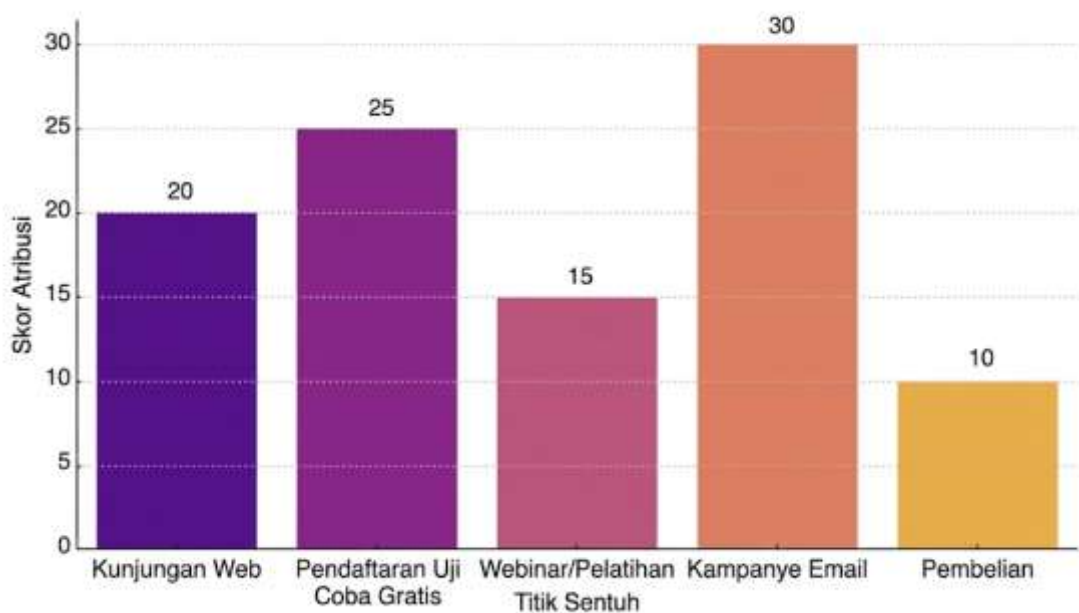
Misalkan perusahaan tersebut menggunakan model atribusi berbasis data untuk pembahasan dampak titik kontak ini terhadap keputusan pembelian akhir. Setelah mengumpulkan data selama beberapa bulan, mereka mungkin menemukan bahwa iklan digital mereka efektif dalam meningkatkan kesadaran dan minat, tetapi webinar dan demo produk mereka lebih berpengaruh dalam mendorong keputusan pembelian akhir.

Akibatnya, mereka mungkin memutuskan untuk mengalokasikan lebih banyak sumber daya untuk mengembangkan webinar dan demo berkualitas tinggi, sambil mempertahankan

pengeluaran iklan digital mereka untuk memastikan kesadaran dan minat yang berkelanjutan. Hal ini dapat melibatkan pembuatan konten webinar yang lebih tertarget atau berinvestasi dalam perangkat lunak demo yang lebih baik untuk memberikan pengalaman yang lebih interaktif dan menarik bagi calon pelanggan.

Selain itu, mereka dapat menggunakan informasi ini untuk mengoptimalkan saluran penjualan mereka, memastikan bahwa calon pelanggan beralih dengan lancar dari tahap kesadaran dan minat ke tahap pertimbangan dan niat, yang pada akhirnya mengarah pada konversi yang lebih sukses (lihat Gambar 9.7).

Dengan cara ini, pemahaman tentang jalur pembelian dan pembahasan atribusi dapat memungkinkan perusahaan perangkat lunak untuk mengoptimalkan strategi pemasarannya, meningkatkan pengalaman pelanggan, dan meningkatkan penjualan.



Gambar 9.7 Rincian Alur Pembelian dan Pembahasan Atribusi untuk Perusahaan Perangkat Lunak, Menunjukkan Titik Sentuh yang Berpengaruh.

9.6 KESIMPULAN

Memahami perjalanan pelanggan merupakan aspek integral dari strategi pemasaran modern. Dengan memanfaatkan analitik canggih, pemasar dapat memperoleh wawasan berharga tentang bagaimana konsumen berinteraksi dengan merek mereka di berbagai titik sentuh dan saluran, dan bagaimana interaksi ini memengaruhi keputusan pembelian mereka.

Analitik perjalanan pelanggan dapat memberikan pandangan holistik tentang pengalaman pelanggan, memungkinkan pemasar untuk mengidentifikasi momen-momen penting keterlibatan dan potensi area gesekan. Hal ini dapat menghasilkan strategi pemasaran yang lebih efektif dan personal, yang pada akhirnya menghasilkan peningkatan kepuasan dan loyalitas pelanggan.

Selain itu, optimasi pemasaran lintas saluran dan pembahasan atribusi dapat memungkinkan pemasar untuk mengalokasikan sumber daya mereka secara lebih efektif, memastikan bahwa setiap titik sentuh dan saluran berkontribusi pada tujuan pemasaran secara keseluruhan. Hal ini dapat menghasilkan peningkatan ROI pemasaran dan kinerja bisnis.

Seiring dengan terus berkembangnya teknologi, analitik perjalanan pelanggan kemungkinan akan menjadi alat yang lebih penting bagi pemasar, memungkinkan mereka untuk menavigasi lanskap digital yang semakin kompleks dan saling terhubung.

Kesimpulannya, konsep dan contoh yang disajikan dalam bab ini menyoroti nilai luar biasa dari analitik perjalanan pelanggan dalam pemasaran. Dengan memahami dan menerapkan konsep-konsep ini, pemasar dapat menciptakan strategi yang lebih efektif dan berpusat pada pelanggan, mendorong hasil jangka pendek dan kesuksesan bisnis jangka panjang.

LATIHAN 9.1

MEMBUAT PETA PERJALANAN PELANGGAN

Tujuan: Memahami proses pemetaan perjalanan pelanggan dengan membuat peta sintetis untuk perusahaan fiktif (ZaraTech).

Tugas:

1. **Pengembangan Persona:** Buat profil terperinci untuk setiap persona pelanggan, termasuk tujuan, tantangan, dan preferensi mereka.
2. **Identifikasi Titik Kontak:** Daftarkan semua kemungkinan interaksi yang mungkin dimiliki persona ini dengan ZaraTech di berbagai saluran.
3. **Pemetaan Perjalanan:** Buat peta perjalanan untuk setiap persona. Sertakan tahapan seperti 'Kesadaran', 'Pertimbangan', 'Pembelian', dan 'Loyalitas'. Plot titik kontak dan potensi emosi atau titik kesulitan di setiap tahapan.
4. **Pembahasan:** Identifikasi momen-momen penting dan titik kesulitan untuk setiap persona.

Langkah-langkah:

1. Impor Pustaka yang Diperlukan:

```
1. import pandas as pd
2. import matplotlib.pyplot as plt
3. import seaborn as sns
```

- **Pandas:** Digunakan untuk manipulasi dan pembahasan data.
- **Matplotlib.pyplot** dan **seaborn:** Digunakan untuk visualisasi data.

2. Memuat Data

```
4. file_path = 'Customer_Journey_Map_Data.csv' # Replace with the
   correct file path
5. journey_map_df = pd.read_csv(file_path)
```

- Kita memuat file CSV ke dalam DataFrame menggunakan **pandas**. Pastikan jalur file sudah benar.

3. Pembahasan Data Eksplorasi:


```
6. # Display the first few rows of the DataFrame
7. print(journey_map_df.head())
8. # Get a summary of the dataset
9. print(journey_map_df.describe(include='all'))
```

- **journey_map_df.head():** Menampilkan beberapa baris pertama data untuk gambaran umum yang cepat.
- **journey_map_df.describe(include='all'):** Memberikan ringkasan statistik data, termasuk jumlah, nilai unik, dan frekuensi untuk data kategorikal.

4. Pembahasan Emosi dan Titik Masalah:

```
10. # Count of emotions per persona
11. emotion_count=journey_map_df.groupby(['Persona', 'Emotion']).size().
    unstack()print(emotion_count)
12. # Count of pain points per persona
13. pain_point_count=journey_map_df.groupby(['Persona', 'PainPoint']).si
    ze().unstack() print(pain_point_count)
```

- Mengelompokkan data berdasarkan 'Persona' dan 'Emosi' untuk melihat distribusi emosi untuk setiap persona.
- Demikian pula, mengelompokkan berdasarkan 'Persona' dan 'Titik Masalah' untuk memahami titik masalah umum.

5. Visualisasikan Data:

```
14. # Plotting emotion count
15. sns.barplot(data=emotion_count)
16. plt.title("Emotion Count per Persona")
17. plt.ylabel("Count")
18. plt.show()
19. # Plotting pain point count sns.barplot(data=pain_point_count)
20. plt.title("Pain Point Count per Persona")
21. plt.ylabel("Count")
22. plt.show()
```

- Menggunakan **seaborn** untuk membuat plot batang untuk jumlah emosi dan titik masalah.
- Plot ini akan membantu memvisualisasikan emosi dan titik masalah mana yang paling umum untuk setiap persona.

Kode ini menyediakan kerangka kerja dasar untuk pembahasan data peta perjalanan pelanggan. Ini membantu dalam memahami distribusi emosi dan titik masalah di berbagai persona, yang dapat sangat berharga untuk menyesuaikan strategi pemasaran dan meningkatkan pengalaman pelanggan.

Ingatlah untuk menyesuaikan jalur file dan mungkin memodifikasi kode agar sesuai dengan format dan persyaratan spesifik data dan tujuan pembahasan Anda.

LATIHAN 9.2

PEMBAHASAN EFEKTIVITAS TITIK SENTUH

Tujuan: Pembahasan efektivitas berbagai titik sentuh dalam perjalanan pelanggan.

Tugas:

1. **Pengumpulan Data:** Gunakan data sintetis untuk mensimulasikan interaksi pelanggan di berbagai titik sentuh.
2. **Pembahasan Metrik:** Hitung efektivitas setiap titik sentuh. Misalnya, tentukan tingkat konversi untuk kunjungan situs web dan kampanye email.
3. **Pembuatan Wawasan:** Identifikasi titik sentuh mana yang paling efektif dalam mendorong kepuasan pelanggan dan konversi.
4. **Perumusan Strategi:** Berdasarkan pembahasan, sarankan perbaikan atau pergeseran strategis untuk ZaraTech guna meningkatkan pengalaman pelanggan.

Langkah-langkah:

1. Impor Pustaka yang Diperlukan:

```
1. import pandas as pd
2. import matplotlib.pyplot as plt
3. import seaborn as sns
```

- **Pandas** digunakan untuk manipulasi dan pembahasan data.
- **Matplotlib.pyplot** dan **seaborn** digunakan untuk visualisasi data.

2. Membuat Data Sintetis:

```
4. file_path = 'Touchpoint_Effectiveness_Analysis_Data.csv'
   # Replace with the correct file path
5. touchpoint_df = pd.DataFrame(touchpoint_data)
```

- Kita memuat file CSV ke dalam DataFrame menggunakan **pandas**. Pastikan jalur file sudah benar.

3. Gambaran Umum Data:

```
6. print(touchpoint_df.head())
```

- **touchpoint_df.head():** Menampilkan beberapa baris pertama dari DataFrame.

4. Pembahasan Data:

```
7. # Descriptive statistics of the dataset
8. print(touchpoint_df.describe())
```

- **touchpoint_df.describe():** Memberikan ringkasan statistik dari dataset.

5. Visualisasikan Data:

Kita akan membuat visualisasi untuk setiap metrik untuk lebih memahami distribusi dan efektivitasnya:

```
9. # Visualizing Customer Satisfaction Scores
10. plt.figure(figsize=(10, 6))
11. sns.barplot(x='Touchpoint', y='CustomerSatisfactionScores', data=touchpoint_df)
12. plt.title('Customer Satisfaction Scores by Touchpoint')
13. plt.show()
14. # Visualizing Conversion Rates
15. plt.figure(figsize=(10, 6))
16. sns.barplot(x='Touchpoint', y='Conversion Rates', data=touchpoint_df)
17. plt.title('Conversion Rates by Touchpoint')
18. plt.show()
19. # Visualizing Repeat Visits/Purchases
```

```
20. plt.figure(figsize=(10, 6))
21. sns.barplot(x='Touchpoint', y='RepeatVisits/Purchases', data=touchpoint_df)
22. plt.title('Repeat Visits/Purchases by Touchpoint')
23. plt.show()
```

- Grafik-grafik ini akan secara visual merepresentasikan efektivitas setiap titik kontak dalam hal kepuasan pelanggan, tingkat konversi, dan kunjungan atau pembelian berulang.

Latihan ini akan membantu dalam memahami bagaimana berbagai titik kontak berkontribusi pada kepuasan pelanggan dan hasil bisnis. Visualisasi akan memberikan cara yang jelas dan langsung untuk mengidentifikasi titik kontak mana yang berkinerja baik dan mana yang mungkin perlu ditingkatkan.

BAB 10

DESAIN EKSPERIMENTAL DALAM PEMASARAN

10.1 PENDAHULUAN

Desain eksperimen dalam pemasaran adalah alat penting bagi bisnis yang bertujuan untuk mengoptimalkan strategi pemasaran mereka. Ini merujuk pada prosedur metodis untuk merancang, menerapkan, pembahasan, dan menafsirkan pengujian yang diatur untuk mengevaluasi strategi atau taktik pemasaran. Tujuannya adalah untuk mengetahui dampak aktual dari berbagai masukan pemasaran terhadap hasil seperti penjualan, keterlibatan pelanggan, atau kesadaran merek.

Metode ilmiah ini memungkinkan pemasar untuk membangun hubungan sebab-akibat daripada sekadar korelasi, menawarkan tingkat kepercayaan yang lebih tinggi pada hasilnya. Misalnya, jika suatu bisnis mengubah strategi periklanannya dan kemudian mengamati peningkatan penjualan, desain eksperimen dapat membantu menentukan apakah peningkatan penjualan tersebut memang disebabkan oleh perubahan strategi periklanan atau dipengaruhi oleh faktor eksternal.

Desain eksperimen dapat diterapkan dalam berbagai konteks pemasaran, termasuk penetapan harga produk, strategi promosi, pemasaran digital, dan banyak lagi. Ketangguhan pendekatan ini berasal dari kemampuannya untuk mengendalikan faktor-faktor pengganggu potensial, sehingga memberikan hasil yang lebih akurat.

Namun, seiring kita melanjutkan pembahasan di bab ini, bersiaplah untuk mengungkap lapisan yang lebih dalam dari desain eksperimen. Desain faktorial fraksional menawarkan cara yang bernuansa namun efisien untuk mempelajari banyak faktor secara bersamaan tanpa perlu menguji semua kemungkinan kombinasi—suatu keuntungan nyata ketika sumber daya terbatas. Sementara itu, multi-armed bandits (MAB) adalah pendekatan yang menarik, menyeimbangkan eksplorasi dan eksploitasi dalam eksperimen pemasaran waktu nyata, terbukti sangat berharga dalam skenario yang menuntut pengambilan keputusan cepat dengan data terbatas. Tetaplah bersama kami saat kita memulai perjalanan melalui alat dan metodologi canggih ini, meningkatkan keahlian Anda dalam desain eksperimen dan implikasinya yang mendalam dalam pemasaran.

10.2 DESAIN EKSPERIMEN

Gambaran Umum Desain Eksperimen

Desain eksperimen (DoE) adalah metodologi statistik yang memungkinkan eksplorasi ruang masalah secara terstruktur dan ilmiah. Ini melibatkan metode sistematis untuk menentukan hubungan antara faktor-faktor yang memengaruhi suatu proses dan hasil dari proses tersebut. Dalam pemasaran, faktor-faktor ini dapat berupa berbagai elemen strategi pemasaran seperti pengeluaran iklan, penetapan harga, fitur produk, dan hasilnya dapat berupa penjualan, keterlibatan pelanggan, atau kesadaran merek.

Pendekatan DoE sangat bermanfaat ketika berurusan dengan banyak variabel masukan, karena memungkinkan penilaian dampak faktor individu dan interaksinya terhadap hasil. Tujuannya adalah untuk menyediakan model statistik yang mengidentifikasi faktor-faktor yang dapat memengaruhi variabel yang diminati, dan juga interaksi di antara faktor-faktor tersebut.

Pada intinya, DoE adalah pendekatan strategis untuk melakukan eksperimen yang memungkinkan banyak faktor masukan dimanipulasi untuk menentukan pengaruhnya terhadap hasil (respons) yang diinginkan sambil menggunakan jumlah percobaan minimum. Efisiensi ini sangat penting dalam pemasaran di mana keterbatasan sumber daya dan waktu adalah hal yang umum. Salah satu manfaat utama DoE adalah kemampuannya untuk meminimalkan risiko mencapai kesimpulan yang salah. Dengan memvariasikan beberapa faktor secara bersamaan, dimungkinkan untuk menghindari efek yang membingungkan di mana dampak dari beberapa faktor digabungkan, sehingga sulit untuk mengaitkan perubahan dalam respons dengan faktor individu.

Faktor, Tingkat, dan Respons dalam Eksperimen

Desain eksperimental adalah pendekatan terstruktur yang digunakan untuk menentukan hubungan sebab-akibat antara variabel. Pemahaman yang tepat tentang terminologi sangat penting untuk implementasi dan interpretasi eksperimen yang efektif. Dalam konteks ini, tiga konsep mendasar adalah faktor, tingkat, dan respons.

Faktor dalam sebuah eksperimen mengacu pada variabel independen terkontrol yang dimanipulasi untuk memastikan pengaruhnya terhadap variabel dependen. Misalnya, dalam eksperimen pemasaran, faktor-faktor tersebut dapat berupa metode periklanan, penetapan harga produk, atau strategi promosi (lihat Tabel 10.1). Penting untuk membedakan antara faktor dan variabel eksternal yang tidak sengaja diubah atau dikendalikan oleh peneliti.

Level mewakili nilai atau pengaturan berbeda yang dapat diasumsikan oleh suatu faktor. Misalnya, jika kita meneliti faktor 'metode periklanan', levelnya mungkin 'media sosial', 'media cetak', dan 'televisi'. Pilihan level dapat sangat memengaruhi hasil dan kesimpulan yang diambil dari eksperimen. Peneliti harus dengan cermat memilih level yang relevan secara praktis dan teoritis.

Respons adalah variabel dependen dalam eksperimen, yang mewakili hasil atau keluaran yang kita ukur. Menggunakan contoh sebelumnya, jika kita melihat efek dari berbagai metode periklanan pada penjualan produk, maka 'jumlah unit yang terjual' akan menjadi responsnya. Respons sangat penting karena membantu kita mengukur dampak dari faktor yang dimanipulasi, sehingga memungkinkan pembahasan dan pengambilan keputusan yang objektif.

Tabel 10.1 Faktor-faktor, Tingkatnya, dan Respons Potensial dalam Eksperimen Pemasaran Hipotetis.

Faktor	Tingkat	Respons Potensial
Harga	Tinggi, rendah	Volume penjualan
Penempatan iklan	Atas, tengah, bawah	Rasio klik-tayang

Warna iklan	Merah, biru, hijau	Rasio keterlibatan

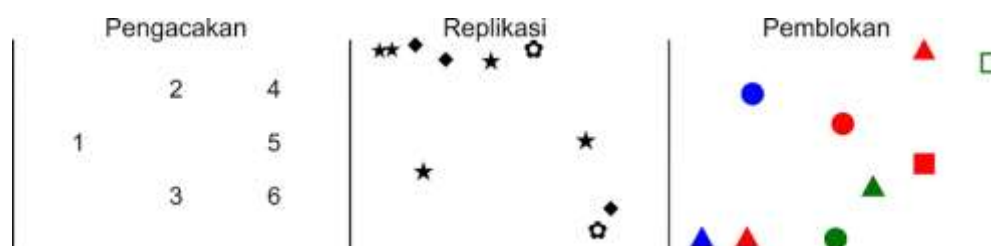
Pengacakan, Replikasi, dan Pemblokiran dalam Desain Eksperimental

Desain eksperimental yang efektif bukan hanya tentang memilih faktor dan level; tetapi juga tentang bagaimana kita menjalankan eksperimen. Tiga prinsip inti sangat penting untuk integritas dan validitas eksperimen: pengacakan, replikasi, dan pemblokiran.

Pengacakan melibatkan penugasan unit eksperimental secara acak ke kombinasi faktor-level yang berbeda. Dengan melakukan ini, kita memastikan bahwa efek faktor eksternal tersebar merata di semua kondisi eksperimental, membantu mengurangi potensi bias. Pendekatan ini memastikan bahwa kesimpulan yang ditarik valid dan dapat dikaitkan dengan faktor yang sedang dipelajari, bukan dengan pengaruh eksternal.

Replikasi mengacu pada pengulangan seluruh eksperimen atau perlakuan spesifik dalam eksperimen. Dengan mereplikasi, kita dapat memahami variabilitas yang melekat dalam pengukuran kita. Ini memberikan perkiraan efek faktor yang lebih tepat, meningkatkan keandalan dan kokohnya kesimpulan kita.

Blocking adalah teknik yang digunakan untuk memperhitungkan variabilitas yang dapat dikaitkan dengan sumber eksternal yang bukan merupakan kepentingan utama dalam eksperimen. Dengan mengelompokkan unit eksperimental ke dalam blok berdasarkan sumber eksternal ini, kita dapat mengontrol atau menghilangkan variabilitas yang disebabkan oleh sumber-sumber ini, sehingga menghasilkan penilaian yang lebih jelas tentang faktor-faktor utama yang menjadi perhatian (lihat Gambar 10.1).



Gambar 10.1 Pentingnya Pengacakan, Replikasi, dan Pemblokiran dengan Diagram Sederhana.

Studi Kasus: Penggunaan Desain Eksperimental yang Efektif dalam Pemasaran

Desain eksperimental yang efektif telah mendorong banyak inisiatif pemasaran yang sukses. Berikut adalah beberapa studi kasus yang menunjukkan kegunaannya:

- **Eksperimen strategi penetapan harga untuk situs e-commerce.** Sebuah pengecer online mencurigai bahwa strategi penetapan harganya tidak optimal. Mereka menyiapkan eksperimen di mana mereka secara acak menugaskan pengunjung ke salah satu dari tiga tingkat harga untuk suatu produk. Dengan pembahasan penjualan, mereka mengidentifikasi titik harga optimal yang memaksimalkan pendapatan.
- **Penempatan iklan di platform media sosial.** Sebuah perusahaan media sosial ingin menentukan penempatan terbaik untuk iklannya guna memaksimalkan klik tanpa

membuat pengguna merasa tidak nyaman. Mereka membuat beberapa penempatan iklan (faktor) dengan berbagai tingkat prominensi (level) dan mengukur rasio klik-tayang (CTR) (respons). Melalui desain eksperimental ini, perusahaan tersebut menentukan penempatan iklan optimal yang menyeimbangkan visibilitas dengan pengalaman pengguna.

- **Desain kemasan untuk produk konsumen.** Sebuah perusahaan barang konsumen berhipotesis bahwa kemasan produk memengaruhi keputusan pembelian. Mereka merancang eksperimen dengan dua desain kemasan dan menempatkannya di toko-toko tertentu. Data penjualan menunjukkan preferensi yang jelas untuk satu desain dibandingkan desain lainnya, yang menyebabkan peluncuran kemasan yang disukai secara menyeluruh di perusahaan.

Masing-masing studi kasus ini menggarisbawahi kekuatan desain eksperimental dalam mendorong keputusan pemasaran. Dengan menguji hipotesis secara sistematis dalam lingkungan yang terkontrol, bisnis dapat memperoleh wawasan yang dapat ditindaklanjuti dan secara langsung memengaruhi keuntungan.

Konsep Kunci dalam Desain Eksperimen

DoE melibatkan beberapa konsep kunci yang sangat penting untuk memahami dan menerapkan teknik ini dalam pemasaran atau bidang lain:

- **Faktor.** Faktor adalah variabel independen terkontrol yang tingkatnya ditetapkan oleh peneliti. Misalnya, dalam eksperimen pemasaran, faktor dapat berupa harga, strategi promosi, atau fitur produk.
- **Tingkat.** Tingkat adalah nilai spesifik dari faktor. Jika faktornya adalah harga, misalnya, tingkat yang berbeda dapat berupa Rp50.000, Rp100.000, Rp150.000, dan seterusnya.
- **Respons.** Respons adalah variabel dependen atau pengukuran hasil.
- **Perlakuan.** Perlakuan adalah kombinasi spesifik dari tingkat faktor.
- **Pengacakan.** Pengacakan adalah penugasan acak perlakuan ke unit eksperimental untuk mengurangi efek faktor eksternal.
- **Replikasi.** Replikasi melibatkan pengulangan eksperimen beberapa kali untuk mendapatkan perkiraan variabilitas data yang lebih akurat.
- **Pemblokiran.** Blocking adalah teknik yang digunakan untuk menghilangkan efek dari sumber variasi yang dapat diidentifikasi.
- **Interaksi.** Interaksi antar faktor terjadi ketika efek dari satu faktor bergantung pada tingkat faktor lain.

Memahami konsep-konsep ini sangat penting untuk merancang dan melaksanakan eksperimen yang efektif, menafsirkan hasil secara akurat, dan membuat keputusan yang tepat berdasarkan hasil tersebut.

Desain Eksperimen dalam Pemasaran Email

DoE dapat diterapkan secara praktis dalam berbagai aspek pemasaran, termasuk pemasaran email (lihat Tabel 10.2). Berikut contohnya. Misalkan sebuah perusahaan ingin mengoptimalkan strategi pemasaran emailnya untuk meningkatkan CTR. Perusahaan tersebut

mengidentifikasi empat faktor yang berpotensi memengaruhi tingkat ini: baris subjek, panjang email, waktu pengiriman, dan keberadaan penawaran. Setiap faktor ditetapkan pada dua level:

- **Judul subjek:** 'Berfokus pada produk' versus 'Berfokus pada pelanggan'
- **Panjang email:** 'Pendek' versus 'Panjang'
- **Waktu pengiriman:** 'Pagi' versus 'Sore'
- **Kehadiran penawaran:** 'Ya' versus 'Tidak'

Perusahaan menggunakan desain faktorial 2^4 , menghasilkan 16 kombinasi berbeda dari faktor-faktor ini. Setiap kombinasi (perlakuan) secara acak diberikan kepada sekelompok pelanggan, dan CTR (respons) diamati. Pembahasan statistik dari data yang dihasilkan memungkinkan perusahaan untuk menentukan efek individual dan interaktif dari faktor-faktor ini pada CTR.

Misalnya, perusahaan mungkin menemukan bahwa email dengan judul subjek yang berfokus pada pelanggan, dikirim pada malam hari, dengan penawaran yang disertakan, menghasilkan CTR tertinggi, terlepas dari panjang email. Hal ini akan memungkinkan perusahaan untuk mengoptimalkan strategi pemasaran emailnya berdasarkan hasil eksperimen.

Tabel 10.2 Membandingkan Hasil dari Desain Eksperimen yang Berbeda dalam Pemasaran Email.

Desain Eksperimental	Subjek Email	Tingkat Pembukaan (%)	CTR (%)	Tingkat Konversi (%)
Kelompok kontrol	Penawaran Standar	20	5	2
Kelompok perlakuan 1	Diskon 10%	30	10	3
Kelompok perlakuan 2	Beli 1 Gratis 1	25	7	2,5

10.3 DESAIN FAKTORIAL FRAKSIONAL

Gambaran Umum Desain Faktorial Fraksional

Desain faktorial fraksional merupakan subset dari desain eksperimen faktorial penuh. Meskipun desain faktorial penuh mempertimbangkan setiap kemungkinan kombinasi faktor dan levelnya, desain faktorial fraksional hanya mencakup "sebagian kecil" dari kombinasi ini, sehingga lebih efisien dan hemat biaya, terutama ketika berurusan dengan sejumlah besar faktor.

Dalam desain faktorial fraksional, tidak semua kemungkinan kombinasi level untuk setiap faktor diuji. Sebaliknya, subset kombinasi yang dipilih dengan cermat digunakan. Hal ini menghasilkan jumlah eksperimen total yang lebih sedikit, yang dapat menghemat waktu dan sumber daya secara signifikan. Namun, konsekuensinya adalah beberapa informasi tentang interaksi antar faktor mungkin hilang atau tercampur.

Sebagai contoh, pertimbangkan sebuah perusahaan yang tertarik untuk menguji efektivitas empat strategi pemasaran yang berbeda, masing-masing dengan dua level. Desain faktorial penuh akan melibatkan $2^4 = 16$ percobaan. Namun, desain setengah faktorial hanya

akan mencakup $2(4-1) = 8$ percobaan, menghemat waktu dan sumber daya perusahaan sekaligus memberikan wawasan berharga tentang efek strategi tersebut (lihat Gambar 10.2).

Penting untuk dicatat bahwa pemilihan kombinasi dalam desain faktorial fraksional harus dilakukan dengan hati-hati, idealnya dengan bantuan perangkat lunak statistik atau seorang ahli statistik, untuk memastikan bahwa desain tersebut seimbang dan memungkinkan analisis yang bermakna.

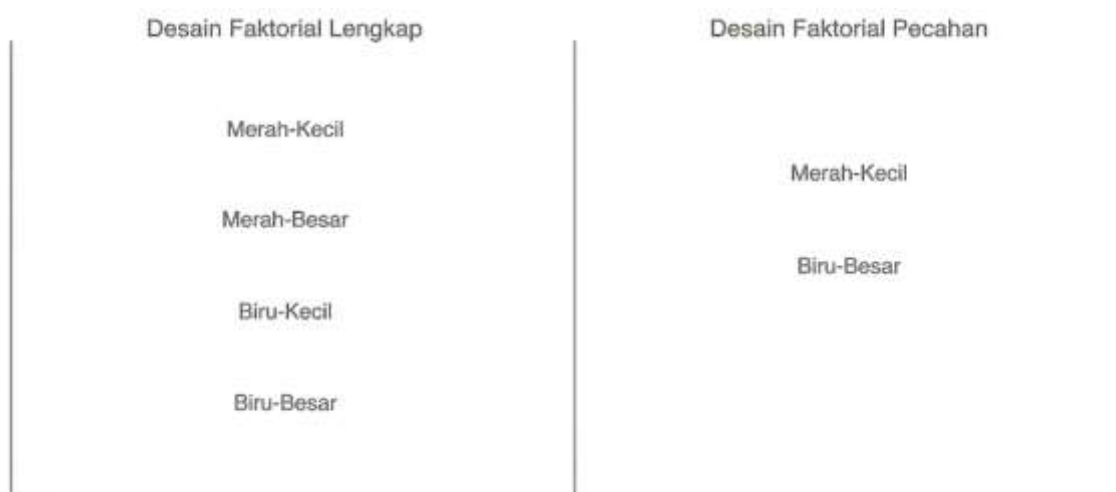
Manfaat Desain Fraksional: Efisiensi dan Efektivitas Biaya

Desain faktorial fraksional, sebagai bagian dari spektrum desain eksperimental yang lebih luas, menghadirkan proposisi nilai yang unik. Manfaat utamanya, seperti namanya, terletak pada kemampuannya untuk hanya menggunakan sebagian kecil dari keseluruhan kondisi eksperimental, sehingga terbukti efisien dan hemat biaya.

Efisiensi

Desain eksperimen faktorial penuh tradisional, yang mempelajari semua kemungkinan kombinasi faktor, dapat menjadi tidak praktis ketika jumlah faktor meningkat karena jumlah kondisi eksperimen yang dibutuhkan tumbuh secara eksponensial dengan jumlah faktor dan tingkatannya.

Desain faktorial fraksional mengatasi tantangan ini dengan secara cerdas memilih subset dari kondisi yang mungkin, memungkinkan studi tentang efek faktor utama dan interaksi tertentu tanpa perlu eksperimen skala penuh. Hal ini membuat proses lebih mudah dikelola, terutama dalam konteks di mana terdapat kendala waktu atau sumber daya.



Gambar 10.2 Desain Faktorial Penuh Versus Desain Faktorial Fraksional.

Efektivitas Biaya

Menjalankan eksperimen skala penuh, terutama dalam konteks dunia nyata seperti pengujian pasar atau pengaturan industri, dapat mahal. Dengan mengurangi jumlah percobaan, desain faktorial fraksional dapat secara signifikan menurunkan biaya. Hal ini sangat penting bagi usaha kecil dan menengah atau perusahaan rintisan dengan anggaran terbatas

karena memungkinkan mereka untuk memperoleh wawasan berharga tanpa investasi keuangan yang besar.

Pembuatan dan Analisis Desain Fraksional

Pembuatan dan analisis selanjutnya dari desain faktorial fraksional memerlukan pendekatan terstruktur.

Pembuatan

Desain fraksional dibangun menggunakan serangkaian generator yang mendefinisikan hubungan antar faktor. Generator membantu dalam menghasilkan fraksi dari desain faktorial penuh dengan menentukan kombinasi faktor mana yang akan disertakan. Biasanya, perangkat lunak statistik khusus atau panduan ahli dicari untuk menghasilkan desain ini, memastikan desain tersebut valid secara statistik dan seimbang secara efektif. Resolusi desain, yang menentukan tingkat kebingungan antar faktor, merupakan pertimbangan penting dalam proses pembuatan ini.

Analisis

Setelah eksperimen dilakukan, data yang dikumpulkan dianalisis menggunakan teknik statistik untuk memperkirakan efek faktor dan signifikansinya (lihat Tabel 10.3). Ini melibatkan penentuan apakah perubahan yang diamati pada variabel respons signifikan secara statistik dan dapat dikaitkan dengan faktor yang dimanipulasi. Penting untuk dicatat bahwa, karena sifat fraksional dari desain, mungkin ada interaksi faktor tertentu yang efeknya tidak dapat dipisahkan secara jelas, yang dikenal sebagai aliasing. Memahami dan menafsirkan efek ini membutuhkan pertimbangan yang cermat untuk memastikan kesimpulan yang valid.

Tabel 10.3 Bagaimana Desain Fraksional Dihasilkan Dan Bagaimana Menganalisisnya.

Percobaan	Faktor A	Faktor B	Faktor C	Generator	Dianalisis sebagai
1	+	+	+	ABC	Efek utama A
2	-	+	-	ABC	Efek utama B
3	+	-	-	A	Interaksi AB
4	-	-	+	B	Interaksi AC

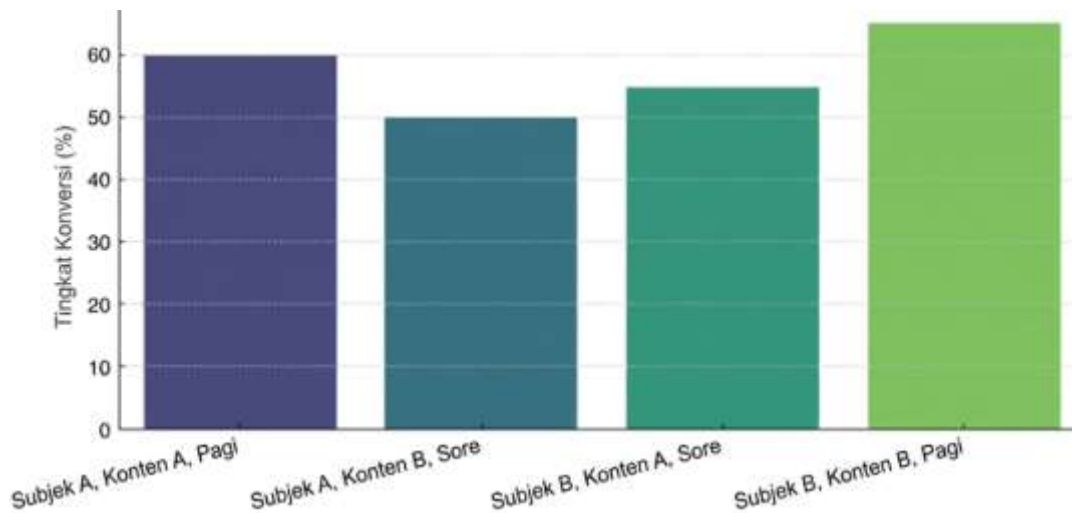
Aplikasi dalam Pemasaran: Menguji Berbagai Elemen Kampanye

Dalam dunia pemasaran yang dinamis, di mana berbagai elemen bergabung untuk menentukan keberhasilan kampanye, desain faktorial fraksional menemukan aplikasi yang sangat penting. Misalnya, pertimbangkan kampanye pemasaran digital di mana sebuah organisasi ingin menguji efek gabungan dari desain iklan, teks ajakan bertindak (call-to-action), penempatan iklan di halaman web, dan waktu penayangan iklan. Melakukan eksperimen faktorial penuh akan melibatkan pengujian setiap kemungkinan kombinasi faktor-faktor ini, sebuah proses yang dapat memakan waktu dan mahal.

Dengan menerapkan desain faktorial fraksional, organisasi dapat secara efisien menguji sebagian dari kombinasi ini, memungkinkan mereka untuk membedakan elemen yang paling berdampak dari kampanye mereka. Pendekatan ini akan memberikan wawasan tentang

efek utama dari setiap faktor dan beberapa interaksi penting tanpa perlu pengujian yang menyeluruh.

Dalam contoh lain, sebuah perusahaan yang meluncurkan produk baru mungkin ingin menentukan kombinasi terbaik dari penetapan harga, pengemasan, dan strategi promosi. Alih-alih menguji setiap kemungkinan kombinasi di pasar, desain fraksional dapat membantu mengidentifikasi campuran optimal dengan hanya menguji sebagian kombinasi, menghemat waktu dan uang sekaligus tetap memberikan wawasan pasar yang berharga (lihat Gambar 10.3). Pada intinya, desain faktorial fraksional menawarkan pemasar alat yang ampuh untuk mengoptimalkan kampanye dengan membuat proses eksperimental lebih efisien dan hemat biaya.



Gambar 10.3 Hasil Penggunaan Desain Fraksional dalam Uji Kampanye Pemasaran.

Konsep Kunci dalam Desain Faktorial Fraksional

Desain faktorial fraksional adalah alat yang ampuh dalam desain eksperimental, terutama ketika berurusan dengan banyak faktor dan sumber daya yang terbatas. Untuk memastikan pemahaman yang tepat tentang desain ini, penting untuk memahami beberapa konsep kunci:

- **Resolusi.** Resolusi desain faktorial fraksional mengacu pada tingkat confounding atau aliasing antara efek utama dan interaksi. Desain dengan resolusi lebih tinggi memiliki confounding yang lebih sedikit, tetapi juga membutuhkan lebih banyak eksperimen. Desain dengan resolusi lebih rendah lebih efisien dalam hal jumlah eksperimen tetapi mungkin memiliki confounding yang lebih banyak.
- **Confounding atau aliasing.** Dalam desain faktorial fraksional, efek faktor tertentu dapat di-confounding atau di-aliasing dengan yang lain, artinya efek individualnya tidak dapat dipisahkan. Hal ini terjadi ketika hanya sebagian kecil dari desain faktorial penuh yang digunakan, yang menyebabkan hilangnya beberapa informasi. Perencanaan yang cermat diperlukan untuk meminimalkan confounding dan memastikan hasil yang bermakna.
- **Ortogonalitas.** Desain ortogonal memastikan bahwa faktor-faktor tersebut independen satu sama lain, memungkinkan estimasi terpisah dari efek utama dan

interaksi tanpa gangguan. Ortogonalitas adalah sifat yang diinginkan dalam desain faktorial fraksional karena membantu meminimalkan dampak confounding.

- **Generator dan relasi definisi.** Generator digunakan untuk mendefinisikan desain faktorial fraksional dengan menentukan struktur confounding. Relasi definisi adalah ekspresi aljabar yang menghubungkan faktor-faktor dalam desain dan dapat digunakan untuk mengidentifikasi alias dan membangun matriks desain.
- **Blokir.** Blokir adalah teknik yang digunakan untuk memperhitungkan variabilitas dalam unit eksperimental, memastikan bahwa hasilnya tidak dipengaruhi oleh faktor-faktor pengganggu ini. Dalam desain faktorial fraksional, blokir dapat digunakan untuk mengontrol variabilitas sambil mempertahankan efisiensi desain.

Desain Faktorial Fraksional dalam Pengujian A/B

Desain faktorial fraksional banyak digunakan dalam pengujian A/B, terutama ketika ada banyak variabel yang akan diuji dan sumber daya yang terbatas. Desain ini memungkinkan pengujian beberapa faktor secara efisien secara bersamaan sambil mengontrol efek confounding. Pertimbangkan sebuah situs web e-commerce yang ingin mengoptimalkan halaman arahan (landing page) untuk meningkatkan rasio konversi. Perusahaan mungkin mempertimbangkan berbagai faktor seperti warna tombol 'Beli Sekarang', posisi ulasan pelanggan, ukuran font deskripsi produk, dan tata letak gambar produk.

Menguji semua kemungkinan kombinasi faktor-faktor ini akan membutuhkan desain faktorial penuh, yang dapat memakan waktu dan sumber daya yang intensif. Dengan menerapkan desain faktorial fraksional, perusahaan dapat secara signifikan mengurangi jumlah pengujian yang dibutuhkan. Katakanlah perusahaan memilih untuk menguji tiga faktor pada dua level masing-masing; desain faktorial fraksional 2^{3-1} hanya akan membutuhkan empat pengujian, bukan delapan seperti yang dibutuhkan oleh desain faktorial penuh. Desain ini memastikan bahwa efek utama dapat diestimasi secara independen, dengan asumsi bahwa interaksi tingkat tinggi dapat diabaikan.

Meskipun pendekatan ini menawarkan efisiensi, pendekatan ini juga menghadirkan kompromi. Perusahaan tidak akan dapat membedakan antara efek dari kombinasi faktor tertentu (dikenal sebagai confounding). Namun, dengan perencanaan yang cermat, interaksi penting dapat diperkirakan, dan keterbatasan ini dapat diatasi. Tabel 10.4 membandingkan kedua pendekatan pengujian ini di berbagai faktor seperti jumlah variasi yang diuji, durasi eksperimen, kompleksitas, dan resolusi, memberikan gambaran yang jelas tentang keunggulan dan keterbatasan masing-masing.

Tabel 10.4 Membandingkan Hasil dari Pengujian A/B Standar dengan Pengujian A/B Faktorial Fraksional.

Faktor Perbandingan	Tes A/B Standar	Tes A/B Faktorial Fraksional
Jumlah variasi yang diuji	Dua (A dan B)	Berganda
Durasi percobaan	Pendek hingga sedang	Sedang hingga panjang
Kompleksitas	Rendah	Sedang hingga tinggi
Resolusi	Tinggi untuk dua variasi	Tinggi untuk beberapa variasi

Desain faktorial fraksional menawarkan cara yang efisien untuk mengeksplorasi faktor-faktor yang paling berpengaruh dan interaksinya terhadap tingkat konversi. Dengan menerapkan analisis statistik pada hasilnya, perusahaan dapat mengidentifikasi kombinasi optimal dari faktor-faktor ini untuk memaksimalkan konversi.

10.4 MULTI-ARMED BANDIT

Gambaran Umum Multi-Armed Bandit

MAB adalah alat penting yang digunakan dalam desain eksperimen dan pengambilan keputusan, khususnya ketika ada kebutuhan untuk menyeimbangkan eksplorasi (mencoba semua opsi untuk mengumpulkan lebih banyak data) dan eksploitasi (tetap pada opsi terbaik yang diketahui).

Ini berasal dari bidang pembelajaran penguatan (reinforcement learning), yang merupakan area pembelajaran mesin di mana agen belajar membuat keputusan dengan berinteraksi dengan lingkungannya.

Istilah *multi-armed* bandit berasal dari skenario hipotetis yang melibatkan seorang penjudi di deretan mesin slot (sering disebut sebagai one-armed bandit karena mekanisme tuasnya dan kecenderungannya untuk mengosongkan kantong pemain), yang harus memutuskan mesin mana yang akan dimainkan, berapa kali memainkan setiap mesin, dan dalam urutan apa memainkannya untuk memaksimalkan total hadiah. Dalam pemasaran, skenario ini dapat dilihat sebagai metafora untuk berbagai masalah pengambilan keputusan, seperti memilih di antara berbagai strategi pemasaran, desain, iklan, atau model penetapan harga, ketika tidak jelas mana yang akan menghasilkan hasil terbaik.

Algoritma MAB dirancang untuk meminimalkan penyesalan, yaitu perbedaan antara total imbalan yang dapat dicapai dengan selalu memilih opsi terbaik dan total imbalan yang sebenarnya dicapai oleh algoritma. Algoritma ini memberikan alternatif yang lebih canggih daripada pengujian A/B tradisional dalam banyak situasi karena algoritma ini terus memperbarui pengetahuannya tentang setiap opsi dan menyesuaikan alokasi sumber daya sesuai dengan itu.

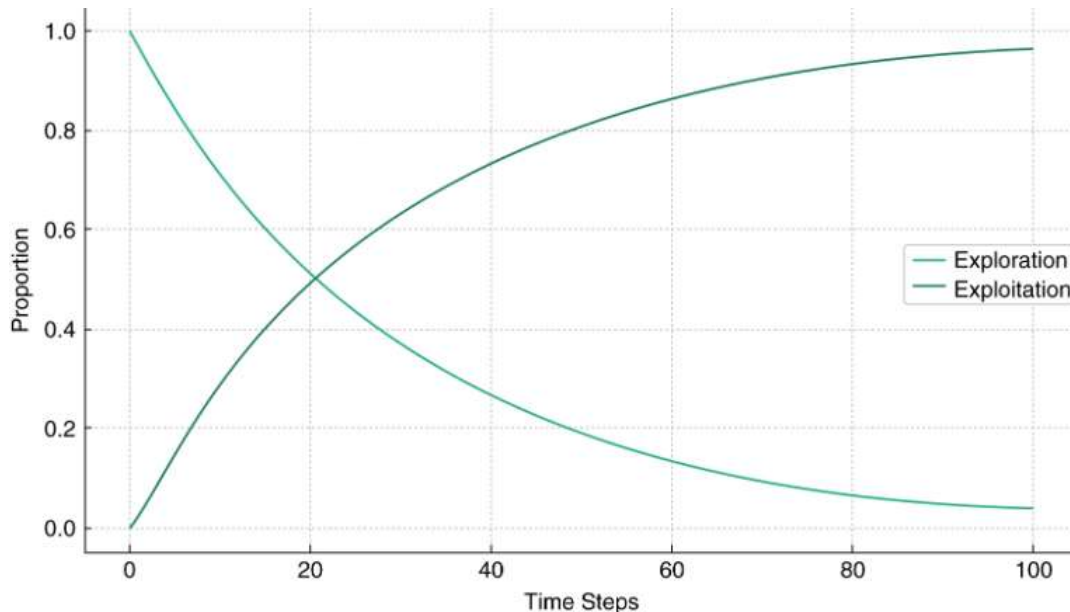
Dilema Eksplorasi Versus Eksploitasi

Salah satu tantangan utama dalam proses pengambilan keputusan, terutama dalam konteks MAB dan pembelajaran penguatan, adalah ketegangan antara eksplorasi dan eksploitasi. Dilema ini merupakan tema sentral dalam banyak skenario pembelajaran mesin dan optimasi.

Eksplorasi melibatkan penyelidikan hal yang tidak diketahui. Dalam skenario praktis, ini berarti mengalokasikan sumber daya ke area dengan hasil yang tidak pasti dengan harapan menemukan solusi baru atau yang lebih baik. Meskipun eksplorasi membawa potensi imbalan, ia juga disertai risiko karena tidak ada jaminan pengembalian yang positif.

Namun, **eksploitasi** melibatkan pemanfaatan informasi yang diketahui untuk membuat keputusan. Saat mengeksploitasi, suatu entitas bergantung pada pengetahuan yang diperoleh untuk memaksimalkan imbalan langsung, memilih opsi yang sebelumnya telah menunjukkan hasil positif. Namun, ketergantungan yang berlebihan pada eksploitasi dapat berarti

kehilangan peluang yang muncul dari solusi baru atau lingkungan yang berubah. Menyeimbangkan keduanya merupakan tantangan: terlalu condong ke eksplorasi dapat berarti pemborosan sumber daya pada pilihan yang kurang optimal, dan fokus pada eksploitasi dapat mencegah penemuan opsi yang lebih efisien atau bermanfaat (lihat Gambar 10.4).



Gambar 10.4 Pertukaran Antara Eksplorasi dan Eksploitasi Seiring Waktu.

Algoritma: ϵ -Greedy, Batas Kepercayaan Atas, dan Pengambilan Sampel Thompson

Berbagai algoritma telah dirancang untuk mengatasi tantangan eksplorasi versus eksploitasi, dengan tujuan mencapai keseimbangan yang efektif antara keduanya.

ϵ -Greedy (Epsilon-Greedy)

Ini adalah salah satu metode paling sederhana. Metode ini bekerja dengan memilih opsi terbaik yang diketahui sebagian besar waktu (eksploitasi) tetapi kadang-kadang (dengan probabilitas kecil ϵ) memilih opsi acak (eksplorasi). Parameter ϵ menentukan keseimbangan, dan ϵ yang lebih besar meningkatkan eksplorasi dengan mengorbankan eksploitasi.

Batas Kepercayaan Atas

Pendekatan UCB (batas kepercayaan atas) memperhitungkan ketidakpastian dalam nilai estimasi setiap opsi. Algoritma memilih opsi berdasarkan batas kepercayaan atas dari nilai estimasinya. Ini memastikan bahwa opsi dipilih tidak hanya berdasarkan imbalan rata-rata yang diamati tetapi juga berdasarkan ketidakpastian atau varians dalam imbalannya. Seiring waktu, seiring semakin banyak data yang dikumpulkan, ketidakpastian berkurang, dan strategi UCB cenderung lebih menyukai opsi dengan imbalan rata-rata yang diamati lebih tinggi.

Pengambilan Sampel Thompson

Ini adalah metode probabilistik yang memilih opsi berdasarkan distribusi posterior Bayesian dari imbalannya. Algoritma ini memodelkan ketidakpastian dalam distribusi imbalan setiap opsi dan mengambil sampel dari distribusi ini untuk menentukan opsi mana yang akan dipilih selanjutnya. Seiring waktu, seiring semakin banyak data yang dikumpulkan, distribusi posterior menjadi lebih halus, membimbing algoritma ke pilihan yang semakin optimal.

Untuk memberikan perbandingan komprehensif dari algoritma utama yang digunakan dalam pengambilan keputusan, termasuk ϵ -Greedy, UCB, dan Pengambilan Sampel Thompson, Tabel 10.5 merinci kelebihan, kekurangan, dan skenario kasus penggunaan idealnya, menawarkan wawasan tentang pemilihan algoritma yang paling tepat berdasarkan persyaratan dan konteks tertentu.

Tabel 10.5 Algoritma Utama (ϵ -Greedy, Batas Kepercayaan Atas, Pengambilan Sampel Thompson) beserta Kelebihan, Kekurangan, dan Skenario Penggunaan Terbaiknya.

Algoritma	Keunggulan	Kelemahan	Paling Cocok Digunakan Untuk
ϵ-Greedy	Sederhana Memungkinkan eksplorasi yang lebih teliti	Membutuhkan penyetelan ϵ secara manual	Skenario di mana tingkat eksplorasi dapat ditentukan sebelumnya
UCB	Menyeimbangkan eksplorasi dan eksploitasi berdasarkan interval kepercayaan.	Dapat memakan banyak sumber daya komputasi	Ketika tersedia sumber daya komputasi yang cukup
Pengambilan Sampel Thompson	Pendekatan probabilistik Beradaptasi berdasarkan data sebelumnya	Membutuhkan pengetahuan atau keyakinan sebelumnya tentang imbalan	Ketika keyakinan awal tentang imbalan tersedia

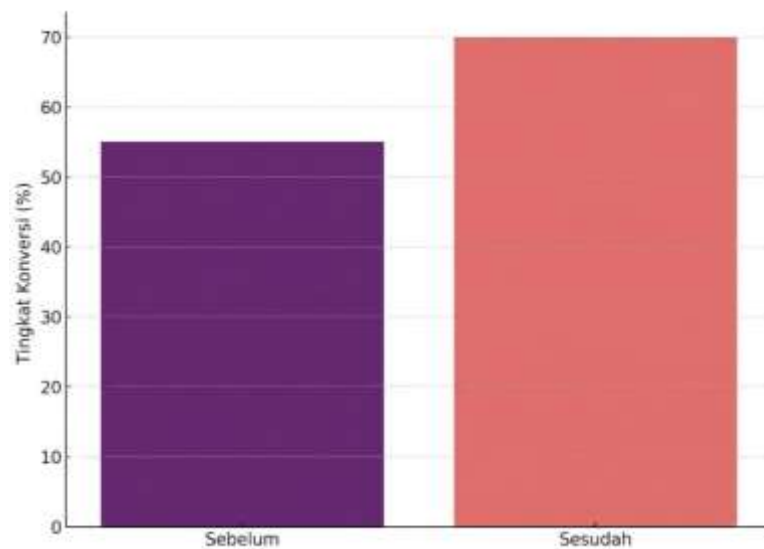
Kasus Penggunaan Praktis dalam Optimasi Pemasaran

Algoritma eksplorasi versus eksploitasi telah menemukan kegunaan yang signifikan dalam skenario optimasi pemasaran:

- **Optimasi kampanye iklan.** Sebuah platform online dapat menggunakan algoritma ϵ -greedy untuk menayangkan iklan berkinerja tertinggi kepada pengguna (eksploitasi) tetapi sesekali menampilkan iklan baru atau yang berkinerja lebih rendah (eksplorasi) untuk mengukur potensi efektivitasnya atau mengumpulkan lebih banyak data tentang kinerjanya.
- **Rekomendasi produk.** Platform e-commerce dapat menggunakan algoritma UCB atau Thompson sampling untuk merekomendasikan produk. Meskipun platform mungkin memiliki serangkaian produk yang dikenal populer (eksploitasi), platform tersebut juga dapat memperkenalkan produk yang lebih baru atau produk dengan interaksi pengguna yang lebih sedikit untuk memahami potensi daya tariknya bagi pelanggan (eksplorasi).
- **Pemasaran email.** Saat memperkenalkan format buletin atau strategi konten baru, perusahaan dapat memanfaatkan algoritma ini untuk mengirim versi baru ke sebagian kecil pelanggan mereka (eksplorasi) sambil terus mengirim format tradisional ke

sebagian besar (eksploitasi). Hal ini memungkinkan umpan balik secara real-time tanpa mempertaruhkan tingkat keterlibatan untuk seluruh basis pelanggan.

Untuk menggambarkan secara visual efektivitas MAB dalam meningkatkan optimasi pemasaran, Gambar 10.5 menyajikan grafik batang yang menunjukkan tingkat konversi, sebelum dan setelah penerapan algoritma MAB, yang mengilustrasikan manfaat nyata dari pendekatan strategis ini dalam skenario dunia nyata. Dalam semua skenario ini, kuncinya terletak pada penyempurnaan strategi yang berkelanjutan, memastikan bahwa seiring tersedianya lebih banyak data, keputusan pemasaran berkembang menjadi lebih efektif dan berpusat pada pelanggan.



Gambar 10.5 Metrik Kinerja (Tingkat Konversi) Sebelum dan Setelah Menggunakan MAB dalam Optimasi Pemasaran.

Konsep Kunci dalam Multi-Armed Bandit

Masalah mendasar dalam MAB adalah memutuskan lengan mana yang akan ditarik (yaitu, tindakan mana yang akan diambil), mengingat imbalan dari setiap tindakan tidak diketahui sampai tindakan tersebut dipilih. Masalah ini juga dikenal sebagai *dilema eksplorasi-eksploitasi*. Berikut adalah beberapa konsep kunci dalam MAB:

- **Lengan.** Dalam konteks MAB, sebuah lengan mewakili pilihan atau keputusan yang harus dibuat. Istilah ini berasal dari tuas (lengan) mesin slot (bandit satu lengan). Dalam pemasaran, sebuah lengan dapat mewakili kampanye iklan tertentu, strategi penetapan harga, atau desain produk.
- **Imbalan.** Ini adalah imbalan yang diterima setelah sebuah lengan ditarik. Imbalan dapat bervariasi setiap kali lengan yang sama ditarik. Dalam pemasaran, ini bisa berupa pendapatan yang dihasilkan, keterlibatan pelanggan, atau hasil yang diinginkan lainnya.

- **Penyesalan.** Penyesalan adalah konsep kunci dalam MAB. Ini mewakili biaya peluang karena tidak menarik lengan terbaik. Tujuan algoritma MAB adalah untuk meminimalkan penyesalan dari waktu ke waktu.
- **Eksplorasi versus eksploitasi.** Ini adalah pertukaran mendasar dalam MAB. Eksplorasi adalah tentang mengumpulkan lebih banyak informasi dengan menarik lengan yang berbeda, sedangkan eksploitasi adalah tentang menarik lengan dengan imbalan yang diharapkan tertinggi berdasarkan informasi yang tersedia saat ini.
- **Strategi ϵ -greedy.** Ini adalah strategi yang sederhana namun efektif dalam MAB. Dengan probabilitas ϵ , sebuah lengan dipilih secara acak (eksplorasi), dan dengan probabilitas $1-\epsilon$, lengan dengan imbalan estimasi tertinggi dipilih (eksploitasi).
- **Strategi UCB.** Ini adalah strategi MAB yang lebih canggih. Hal ini melibatkan penarikan lengan dengan batas kepercayaan atas tertinggi pada imbalan yang diharapkan. Strategi ini mencapai keseimbangan yang lebih baik antara eksplorasi dan eksploitasi.

Multi-Armed Bandit dalam Optimasi Situs Web

MAB telah menemukan berbagai aplikasi di bidang pemasaran. Salah satu aplikasi tersebut adalah dalam optimasi situs web. Mari kita pertimbangkan kasus perusahaan e-commerce yang ingin mengoptimalkan tata letak halaman berandanya untuk memaksimalkan keterlibatan pelanggan.

Misalkan perusahaan telah mendesain tiga tata letak berbeda (A, B, dan C) untuk halaman berandanya. Setiap tata letak adalah sebuah lengan dalam konteks MAB. Perusahaan ingin mengidentifikasi tata letak yang paling efektif, yaitu tata letak yang menghasilkan CTR tertinggi. Namun, perusahaan tidak mengetahui CTR sebenarnya dari tata letak tersebut. Oleh karena itu, perusahaan menghadapi dilema eksplorasi-eksploitasi: haruskah perusahaan mengumpulkan lebih banyak informasi tentang tata letak (eksplorasi) atau memilih tata letak yang saat ini tampak terbaik (eksploitasi)?

Perusahaan dapat menggunakan algoritma MAB, seperti ϵ -greedy atau UCB, untuk mengatasi dilema ini. Algoritma tersebut akan menyeimbangkan antara mengeksplorasi semua tata letak dan mengeksploitasi tata letak yang saat ini berkinerja terbaik. Seiring waktu, algoritma akan konvergen ke tata letak terbaik, meminimalkan penyesalan (yaitu, biaya peluang karena tidak memilih tata letak terbaik). Contoh aplikasi tersebut diberikan oleh Chapelle dan Li. Mereka melaporkan bahwa algoritma contextual bandit (varian MAB) menghasilkan peningkatan 12,5% dalam CTR untuk rekomendasi artikel berita di Yahoo!

10.5 DESAIN EKSPERIMENTAL

Gambaran Umum Eksperimen Online dan Offline

Desain eksperimen merupakan bagian penting dari analitik pemasaran, dan melibatkan pengujian berbagai strategi atau taktik untuk memahami dampaknya pada indikator kinerja utama. Eksperimen dapat diklasifikasikan secara luas menjadi dua kategori: online dan offline.

Eksperimen online, juga dikenal sebagai pengujian *A/B* atau pengujian terpisah, dilakukan di lingkungan digital langsung seperti situs web, aplikasi, atau kampanye pemasaran

email. Di sini, kinerja berbagai versi halaman web, antarmuka aplikasi, atau pesan pemasaran (seperti email atau iklan) diuji pada berbagai kelompok pengguna. Pendekatan ini memungkinkan pengumpulan data secara real-time dan wawasan yang cepat. Eksperimen online telah dipopulerkan oleh perusahaan yang mengutamakan digital seperti Google dan Amazon, dan telah berperan penting dalam membentuk desain pengalaman pengguna dan strategi pemasaran digital.

Namun, eksperimen offline dilakukan dalam lingkungan yang terkontrol, seringkali secara fisik. Ini dapat melibatkan kelompok fokus, wawancara tatap muka, survei, atau uji coba lapangan. Misalnya, sebuah perusahaan dapat menguji produk baru dengan merilisnya di sejumlah toko terpilih dan mengukur kinerjanya dibandingkan dengan kelompok kontrol toko. Eksperimen offline sangat berguna ketika eksperimen tersebut melibatkan faktor-faktor yang tidak dapat didigitalisasi, seperti atribut produk fisik atau pengalaman di dalam toko.

Eksperimen online dan offline memiliki kekuatan dan kelemahan masing-masing. Pilihan di antara keduanya bergantung pada faktor-faktor seperti sifat produk/layanan, target audiens, hipotesis yang diuji, dan sumber daya yang tersedia.

Tantangan dan Manfaat Pengujian A/B Online

Pengujian A/B online, ketika dua atau lebih variasi strategi digital dibandingkan, telah muncul sebagai alat fundamental dalam pemasaran digital dan optimasi pengalaman pengguna (lihat Tabel 10.6).

Tabel 10.6 Perbandingan Tantangan dan Manfaat Pengujian A/B Daring Versus Metode Offline.

Faktor Perbandingan	Pengujian A/B Online	Metode Offline
Kecepatan	Cepat	Lebih lambat
Biaya	Variabel (bisa rendah)	Bisa tinggi
Skalabilitas	Tinggi	Lebih rendah
Fleksibilitas	Tinggi	Variabel
Analisis waktu nyata	Mungkin	Tidak selalu memungkinkan
Kedalaman wawasan	Tergantung pada pelacakan	Bisa lebih mendalam dengan data kualitatif

Manfaat

- **Umpan balik waktu nyata.** Tes online menawarkan wawasan langsung, memungkinkan pemasar untuk dengan cepat menyesuaikan strategi berdasarkan perilaku pengguna.
- **Skalabilitas.** Dengan platform digital, dimungkinkan untuk menguji variasi dalam skala besar, mencakup berbagai wilayah geografis dan demografis.
- **Fleksibilitas.** Lingkungan online memungkinkan penyesuaian yang mudah. Jika strategi tertentu tidak berhasil, strategi tersebut dapat diganti atau diubah tanpa biaya yang signifikan.

Tantangan

- **Ketergantungan berlebihan pada data kuantitatif.** Tes online menyediakan data kuantitatif yang melimpah, tetapi informasi tersebut mungkin kurang memiliki wawasan kualitatif yang membantu menjelaskan perilaku pengguna.
- **Banyaknya variabel.** Lingkungan digital bersifat dinamis, dengan banyak variabel yang terjadi bersamaan. Hal ini terkadang dapat menyulitkan untuk memastikan kausalitas.
- **Kekhawatiran privasi.** Pengumpulan dan analisis data pengguna daring menimbulkan implikasi privasi, yang mengharuskan kepatuhan ketat terhadap peraturan perlindungan data.

Merancang dan Mengimplementasikan Eksperimen Offline

Meskipun eksperimen daring memiliki kemewahan alat otomatis dan aliran data yang luas, eksperimen offline di lingkungan fisik seringkali menimbulkan tantangan yang berbeda, yang terkait dengan variabel kompleks dan ketidakpastian perilaku manusia.

Desain

Dasar dari setiap eksperimen offline terletak pada desainnya yang cermat. Ini bukan hanya tentang mencantumkan variabel, tetapi juga memahami jalinan interaksi yang rumit dalam pengaturan dunia nyata. Pertimbangkan skenario ritel. Sebuah merek ingin mengetahui pengaruh tata letak toko baru terhadap penjualan. Selain membuat hipotesis tentang potensi dampaknya, mereka perlu menggali lebih dalam dan mempertimbangkan faktor-faktor ini:

- **Hipotesis.** Apakah tata letak baru meningkatkan penjualan?
- **Variabel.** Tata letak toko, penempatan produk, visibilitas papan nama, interaksi staf-pelanggan, dan bahkan faktor-faktor seperti pencahayaan toko atau musik latar.
- **Kriteria pengukuran.** Volume penjualan, jumlah pengunjung, waktu rata-rata pelanggan berada di toko, dan mungkin bahkan metrik seperti kepuasan pelanggan atau kunjungan berulang.

Namun, berikut adalah contoh nyata: Pada tahun 2018, sebuah jaringan kedai kopi global memutuskan untuk bereksperimen dengan desain "tanpa tempat duduk" di beberapa gerai untuk mempromosikan pesanan bawa pulang dan mengurangi kepadatan di dalam toko. Dengan mengukur tidak hanya penjualan tetapi juga waktu pemrosesan pesanan dan umpan balik pelanggan, mereka memperoleh wawasan yang mengarah pada perluasan desain ini ke lebih banyak gerai perkotaan.

Implementasi

Keajaiban terletak pada mengubah teori menjadi tindakan nyata. Dalam contoh ritel kita, toko akan memulai perubahan di gerai-gerai tertentu, bertindak sebagai kelompok perlakuan, sementara yang lain tetap tidak berubah, berfungsi sebagai kelompok kontrol. Namun, di sinilah kompleksitas muncul:

- **Kerangka waktu.** Berapa lama eksperimen harus dijalankan untuk memperhitungkan efek kebaruan awal?
- **Efek musiman.** Bagaimana jika musim liburan memengaruhi data penjualan?
- **Faktor yang tidak terduga.** Sebuah acara lokal tiba-tiba dapat meningkatkan jumlah pengunjung.

Singkatnya, eksperimen offline mirip dengan mengorkestrasi simfoni dengan banyak bagian yang bergerak. Meskipun tantangannya beragam, begitu pula imbalannya: menawarkan pemasar wawasan tentang perilaku dan preferensi konsumen di dunia nyata.

Menjembatani Data Online dan Offline untuk Wawasan Komprehensif

Di dunia yang saling terhubung saat ini, sangat penting untuk mengintegrasikan wawasan dari saluran online dan offline untuk pemahaman holistik tentang perilaku konsumen. Menggabungkan data dari kedua ranah dapat memberikan gambaran yang lebih komprehensif tentang perjalanan pelanggan. Misalnya, seorang konsumen mungkin pertama kali menemukan merek secara online melalui iklan media sosial tetapi melakukan pembelian di toko fisik. Dengan mengintegrasikan data online (seperti CTR atau kunjungan situs web) dengan data offline (seperti pembelian di toko atau umpan balik), merek dapat memperoleh wawasan yang lebih dalam tentang titik kontak yang memengaruhi keputusan pembelian.

Namun, tantangannya terletak pada pengintegrasian beragam kumpulan data ini secara kohesif. Alat analisis data canggih dan sistem manajemen hubungan pelanggan memainkan peran penting dalam integrasi ini, memungkinkan pemasar untuk melacak interaksi pelanggan di berbagai saluran dan titik kontak.

Konsep Kunci dalam Eksperimen Online dan Offline

Terdapat beberapa konsep kunci yang penting untuk memahami dan menerapkan eksperimen online dan offline dalam pemasaran:

- **Pengacakan.** Ini melibatkan penugasan subjek (pengguna, pelanggan, dan lain-lain.) secara acak ke kelompok yang berbeda, seperti kelompok kontrol dan satu atau lebih kelompok perlakuan. Pengacakan sangat penting untuk meminimalkan efek variabel pengganggu dan bias, sehingga memastikan hasil eksperimen valid dan dapat diandalkan.
- **Kelompok kontrol.** Kelompok kontrol menerima perlakuan "standar" atau "default", yang biasanya merupakan strategi atau penawaran saat ini. Kelompok ini berfungsi sebagai tolok ukur yang digunakan untuk membandingkan kelompok perlakuan.
- **Kelompok perlakuan.** Kelompok perlakuan menerima perlakuan baru atau berbeda yang sedang diuji. Misalnya, dalam uji A/B, satu kelompok mungkin melihat halaman web saat ini (kelompok kontrol), dan kelompok lain melihat versi dengan tata letak yang dimodifikasi (kelompok perlakuan).
- **Pengujian hipotesis.** Tujuan dari sebuah eksperimen seringkali adalah untuk menguji hipotesis tertentu. Misalnya, 'Mengubah warna tombol "Beli Sekarang" dari hijau menjadi merah akan meningkatkan CTR'. Hasil eksperimen digunakan untuk menolak atau gagal menolak hipotesis.
- **Signifikansi statistik.** Konsep ini berkaitan dengan kemungkinan bahwa hasil eksperimen terjadi secara kebetulan. Jika hasil eksperimen signifikan secara statistik, itu berarti kita dapat cukup yakin bahwa efek yang diamati adalah nyata dan bukan karena keacakan atau kebetulan.
- **Analisis daya.** Analisis daya digunakan untuk menentukan ukuran sampel yang diperlukan untuk mendeteksi efek dengan ukuran tertentu dengan tingkat kepercayaan

tertentu. Ini membantu memastikan bahwa eksperimen tidak terlalu kecil (berisiko kesalahan Tipe II, atau gagal mendeteksi efek nyata) maupun terlalu besar (membuang sumber daya).

Dalam ranah eksperimen online, konsep-konsep ini diterapkan dalam skala besar dan seringkali secara real-time. Namun, eksperimen offline mungkin melibatkan ukuran sampel yang lebih kecil dan jangka waktu yang lebih lama tetapi tetap mengikuti prinsip yang sama.

Eksperimen Online dan Offline dalam Pemasaran Multichannel

Eksperimen online dan offline dapat digunakan secara efektif dalam pemasaran multichannel untuk mengoptimalkan pengalaman pelanggan dan memaksimalkan hasil bisnis. Berikut adalah contoh praktis: Pertimbangkan sebuah perusahaan ritel besar yang mengoperasikan toko fisik dan situs web e-commerce. Perusahaan ingin menentukan strategi pemasaran yang paling efektif untuk mempromosikan lini produk baru. Untuk melakukan ini, mereka memutuskan untuk melakukan eksperimen online dan offline:

- **Eksperimen online.** Perusahaan dapat menjalankan uji A/B di situs web mereka, di mana setengah dari pengunjung situs web diperlihatkan iklan pop-up untuk lini produk baru (kelompok perlakuan), dan setengah lainnya melihat situs web seperti biasa tanpa pop-up (kelompok kontrol). Perusahaan kemudian mengukur CTR pada pop-up dan tingkat konversi selanjutnya di antara mereka yang mengklik.
- **Eksperimen offline.** Di toko fisik, perusahaan dapat melakukan eksperimen serupa dengan memasang pajangan promosi untuk lini produk baru di beberapa toko (kelompok perlakuan), dan di toko lain, produk ditempatkan di rak biasa (kelompok kontrol). Kemudian, perusahaan membandingkan penjualan lini produk baru di toko-toko yang diberi perlakuan dengan penjualan di toko-toko kontrol.

Dalam kedua eksperimen tersebut, perusahaan menggunakan pengujian hipotesis untuk menentukan apakah perlakuan (iklan pop-up atau tampilan promosi) menyebabkan peningkatan minat atau penjualan yang signifikan secara statistik untuk lini produk baru. Dengan membandingkan hasil eksperimen online dan offline, perusahaan dapat memperoleh wawasan tentang efektivitas berbagai strategi pemasaran di berbagai saluran (lihat Tabel 10.7).

Tabel 10.7 Hasil Strategi Pemasaran Multichannel Menggunakan Wawasan Eksperimental Online dan Offline.

Metrik	Sebelum Gabungan Wawasan (%)	Setelah Gabungan Wawasan (%)
Tingkat keterlibatan	60	75
Tingkat konversi	20	30
ROI	150	190

Integrasi eksperimen online dan offline dengan cara ini memungkinkan pemahaman yang lebih komprehensif tentang perilaku pelanggan dan efektivitas pemasaran. Hal ini juga menggarisbawahi nilai pendekatan multichannel untuk pemasaran, yang bertujuan untuk menjangkau pelanggan di mana pun mereka berada, online atau offline.

10.6 KESIMPULAN

Kesimpulannya, desain eksperimental dalam pemasaran adalah alat yang ampuh yang memungkinkan pemasar untuk secara sistematis menguji hipotesis, mengukur efektivitas berbagai strategi pemasaran, dan menghasilkan wawasan yang dapat ditindaklanjuti. Metodologi yang dibahas dalam bab ini, termasuk DoE, desain faktorial fraksional, MAB, dan eksperimen online dan offline, menawarkan pendekatan berbeda untuk memahami dan mengoptimalkan strategi pemasaran berdasarkan bukti empiris.

Desain DoE dan faktorial fraksional, misalnya, dapat membantu pemasar memahami pengaruh berbagai faktor terhadap hasil pemasaran dan mengoptimalkan strategi mereka berdasarkan wawasan ini. Namun, MAB memberikan cara efektif untuk menyeimbangkan pertukaran eksplorasi-eksploitasi dalam situasi ketika sumber daya terbatas dan keputusan perlu dibuat secara real-time.

Terakhir, eksperimen online dan offline memberikan cara untuk mengintegrasikan saluran pemasaran digital dan tradisional untuk memberikan pandangan holistik tentang perilaku pelanggan dan efektivitas pemasaran. Penggunaan metode ini telah terbukti menghasilkan peningkatan kinerja pemasaran, seperti CTR yang lebih tinggi, peningkatan tingkat konversi, dan peningkatan retensi pelanggan. Namun, seperti halnya metode apa pun, keberhasilan desain eksperimental dalam pemasaran bergantung pada desain dan pelaksanaan eksperimen yang cermat, serta pemahaman yang tepat tentang prinsip-prinsip statistik yang mendasarinya. Dengan memahami dan menerapkan konsep-konsep ini, pemasar dapat membuat keputusan berbasis data, mengoptimalkan strategi pemasaran mereka, dan pada akhirnya mendorong hasil bisnis yang lebih baik.

LATIHAN 10.1

MENGANALISIS UJI A/B SEDERHANA

Tujuan: Untuk mendemonstrasikan prinsip-prinsip dasar desain eksperimental menggunakan skenario uji A/B dalam pemasaran email.

Tugas:

Anda diberikan data dari kampanye pemasaran email di mana dua baris subjek yang berbeda diuji untuk melihat mana yang menghasilkan tingkat pembukaan yang lebih tinggi. Tugas Anda adalah menganalisis data untuk menentukan baris subjek mana yang berkinerja lebih baik.

1. **Uji Statistik:** Lakukan uji t untuk melihat apakah perbedaan tingkat pembukaan antara kedua kelompok signifikan secara statistik.
2. **Interpretasi Hasil:** Berdasarkan nilai p dari uji t, simpulkan baris subjek mana yang berkinerja lebih baik.

Langkah-langkah:

1. Impor Pustaka:

- ```
1. import scipy.stats as stats
2. import pandas as pd
```

Kami mengimpor dua pustaka: **scipy.stats** untuk pengujian statistik dan **pandas** untuk menangani data dalam bentuk terstruktur (DataFrame).

## 2. Muat Data:

```
3. email_marketing_data=pd.read_csv('/data/Email_Marketing_AB_Test_Data.csv')
```

Kami memuat data ke dalam DataFrame pandas. Data ini mensimulasikan tingkat pembukaan email untuk dua baris subjek yang berbeda (Grup A dan Grup B).

## 4. Pisahkan Data menjadi Dua Grup:

```
4. group_A=email_marketing_data[email_marketing_data['Group']=='A']['OpenRate'] ==
5. group_B=email_marketing_data[email_marketing_data['Group']=='B']['OpenRate'] ==
```

Di sini, kita memfilter DataFrame untuk membuat dua seri terpisah: satu untuk setiap grup. **group\_A** berisi tingkat pembukaan untuk baris subjek A, dan **group\_B** untuk baris subjek B.

## 6. Lakukan Uji-t:

```
6. t_stat, p_value = stats.ttest_ind(group_A, group_B)
```

Kami melakukan uji t independen (**ttest\_ind**) untuk membandingkan rata-rata tingkat pembukaan email dari kedua kelompok. Uji ini akan membantu kami menentukan apakah ada perbedaan yang signifikan secara statistik antara tingkat pembukaan email dari kedua baris subjek tersebut.

## LATIHAN 10.2

### DESAIN FAKTORIAL FRAKSIONAL DALAM OPTIMASI IKLAN

**Tujuan:** Untuk mengilustrasikan penerapan desain faktorial fraksional dalam mengoptimalkan kampanye iklan.

#### Tugas:

Anda diberikan dataset dari eksperimen iklan online dengan beberapa faktor (seperti warna iklan, penempatan, dan ukuran) dan levelnya. Tugas Anda adalah menganalisis data untuk menentukan kombinasi optimal dari faktor-faktor ini untuk tingkat klik-tayang maksimum.

1. **Analisis Faktorial:** Gunakan analisis regresi untuk memahami dampak setiap faktor dan interaksinya terhadap tingkat klik-tayang.
2. **Optimasi:** Identifikasi kombinasi faktor yang menghasilkan tingkat klik-tayang prediksi tertinggi.

#### Langkah-langkah:

##### 1. Impor Pustaka dan Muat Data:

```
1. import statsmodels.api as sm
2. import pandas as pd
3. file_path =
 'Ad_Optimization_Fractional_Factorial_Design_Data.csv'#Replace with
 the correct file path
4. ad_optimization_data = pd.read_csv(file_path)
```

Kita mengimpor **statsmodels** untuk analisis regresi dan **pandas** untuk manipulasi data. Kemudian, kita memuat data ke dalam DataFrame pandas.

##### 2. Membuat Variabel Dummy:

```
5. ad_data_dummies=pd.get_dummies(ad_optimization_data,drop_first=True
)
```

Karena data kita mengandung variabel kategorikal (AdColor, Placement, Size), kita mengubahnya menjadi variabel dummy untuk analisis regresi. Argumen **drop\_first=True** digunakan untuk menghindari multikolinearitas dengan menghilangkan level pertama dari setiap variabel kategorikal.

### **3. Mempersiapkan Data untuk Regresi:**

```
6. X = ad_data_dummies.drop('ClickThroughRate', axis=1)
7. y = ad_data_dummies['ClickThroughRate']
8. X = sm.add_constant(X)
```

Kita memisahkan variabel independen (X) dan variabel dependen ('ClickThroughRate', y). Kita juga menambahkan konstanta ke dalam model, yang bertindak sebagai intersep dalam persamaan regresi.

### **4. Sesuaikan Model Regresi:**

```
9. model = sm.OLS(y, X).fit()
```

Kita menggunakan regresi kuadrat terkecil biasa (OLS) untuk menyesuaikan model. Metode ini menemukan garis yang paling sesuai melalui data dengan meminimalkan jumlah kuadrat deviasi vertikal dari setiap titik data ke garis tersebut.



## **BAB 11**

### **TEKNOLOGI BIG DATA DAN ANALISIS REAL-TIME**

#### **11.1 PENDAHULUAN**

Seiring dengan terus berkembangnya lanskap digital, demikian pula jumlah data yang dihasilkan dan dikumpulkan oleh bisnis. Data ini, yang sering dicirikan oleh volume, kecepatan, dan variasinya, umumnya disebut sebagai big data. Munculnya big data telah secara drastis mengubah cara bisnis beroperasi dan mengambil keputusan. Dengan memanfaatkannya, bisnis dapat memperoleh wawasan yang lebih dalam tentang operasi mereka, memahami perilaku pelanggan, memprediksi tren masa depan, dan membuat keputusan berbasis data.

Bagi para profesional pemasaran, big data menghadirkan peluang untuk mendapatkan pandangan 360 derajat tentang pelanggan. Dengan menganalisis sejumlah besar data yang dihasilkan melalui interaksi pelanggan di berbagai titik kontak, pemasar dapat mengembangkan strategi pemasaran yang lebih personal dan efektif.

Namun, memanfaatkan kekuatan big data bukanlah tanpa tantangan. Ukuran dan kompleksitas big data yang sangat besar membutuhkan teknologi yang kuat dan terukur untuk penyimpanan, pemrosesan, dan analisis. Di sinilah kerangka kerja komputasi terdistribusi, seperti Hadoop dan Spark, berperan. Teknologi ini memungkinkan bisnis untuk memproses dan menganalisis big data secara efisien dan efektif.

Selain memproses dan menganalisis big data, terdapat juga kebutuhan yang semakin meningkat untuk analitik real-time. Seperti namanya, analitik real-time melibatkan analisis data saat data tersebut dihasilkan untuk memberikan wawasan secara real-time. Analitik real-time memainkan peran penting dalam berbagai aktivitas pemasaran, seperti penawaran real-time dalam periklanan digital dan personalisasi dalam e-commerce. Dalam bab ini, kita akan menyelami dunia big data, mengeksplorasi kerangka kerja komputasi terdistribusi, membahas alat dan teknik analitik real-time, dan memahami bagaimana elemen-elemen ini memungkinkan personalisasi dan pemasaran real-time.

#### **11.2 BIG DATA**

##### **Gambaran Umum Big Data**

Big data adalah istilah yang menggambarkan volume data yang besar, terstruktur dan tidak terstruktur, yang membanjiri bisnis setiap hari. Hal ini dicirikan oleh empat V—volume, variasi, kecepatan, dan kebenaran—untuk menyoroti pentingnya kualitas data:

- **Volume** mengacu pada skala data yang sangat besar. Dengan digitalisasi bisnis dan proliferasi aktivitas online, organisasi kini berurusan dengan jumlah data mulai dari terabyte hingga zettabyte.
- **Variasi** berkaitan dengan beragam bentuk data, yang dapat diklasifikasikan sebagai terstruktur (misalnya, basis data relasional), semi-terstruktur (misalnya, file XML), dan tidak terstruktur (misalnya, file teks, video, audio).

- **Kecepatan** mengacu pada kecepatan data dibuat, disimpan, dianalisis, dan divisualisasikan. Dalam banyak kasus, pemrosesan data secara real-time atau mendekati real-time diperlukan untuk mendapatkan wawasan tepat waktu dan membuat keputusan cepat.
- **Keakuratan** menyoroti keandalan dan kepercayaan data. Mengingat banyaknya data dan berbagai sumbernya, memastikan kualitas dan akurasi data menjadi perhatian yang sangat penting.

Big data memiliki implikasi signifikan bagi pemasaran karena memungkinkan bisnis untuk memperoleh pemahaman yang lebih dalam tentang perilaku pelanggan, meningkatkan keterlibatan pelanggan, memprediksi tren masa depan, dan mengoptimalkan strategi pemasaran. Namun, untuk sepenuhnya memanfaatkan potensi big data, bisnis perlu menggunakan teknologi yang kuat dan terukur untuk penyimpanan, pemrosesan, dan analisis data (lihat Tabel 11.1).

**Tabel 11.1** Tantangan dan Peluang dalam Menangani Big Data Beserta Solusi Potensial dan Kondisi Industri Saat Ini.

| Tantangan             | Peluang                                             | Solusi                                                | Kondisi Saat Ini                                 |
|-----------------------|-----------------------------------------------------|-------------------------------------------------------|--------------------------------------------------|
| <b>Volume data</b>    | Mendapatkan wawasan dari kumpulan data besar        | Solusi penyimpanan yang dapat diskalakan seperti HDFS | Diadopsi secara luas di industri                 |
| <b>Kecepatan data</b> | Analisis dan pengambilan keputusan secara real-time | Alat pemrosesan waktu nyata seperti Kafka dan Storm   | Adopsi yang meningkat untuk aplikasi waktu nyata |
| <b>Variasi data</b>   | Memanfaatkan beragam sumber data                    | Alat integrasi data dan basis data yang fleksibel     | Alat dalam tahap pengembangan dan adopsi         |
| <b>Keamanan data</b>  | Memastikan privasi dan kepercayaan                  | Teknik enkripsi dan penyamaran data tingkat lanjut    | Perhatian tinggi                                 |

### **Perspektif Historis: Dari Basis Data Tradisional ke Big Data**

Perjalanan dari basis data tradisional ke era big data merupakan bukti evolusi teknologi yang cepat dan dahaga informasi yang terus meningkat. Pada masa awal, bisnis mengandalkan sistem file sederhana dan kemudian, basis data relasional terstruktur, untuk menyimpan, mengambil, dan mengelola data. Sistem ini dirancang untuk tugas-tugas tertentu, dengan skema yang jelas dan infrastruktur tetap.

Namun, ledakan digital pada akhir abad ke-20 dan awal abad ke-21, yang didorong oleh internet, e-commerce, dan media sosial, memunculkan longsor data. Data ini berbeda dari format terstruktur tradisional, mencakup berbagai bentuk seperti teks, gambar, dan video. Sistem yang ada, betapapun canggihnya, tidak siap untuk menangani lonjakan volume dan variasi ini.

Kebutuhan untuk memproses data yang luas dan beragam ini melahirkan konsep big data dan pengembangan teknologi yang secara khusus dirancang untuknya. Basis data NoSQL muncul sebagai alternatif untuk basis data relasional tradisional, menekankan fleksibilitas, skalabilitas, dan kemampuan untuk mengelola data tidak terstruktur.

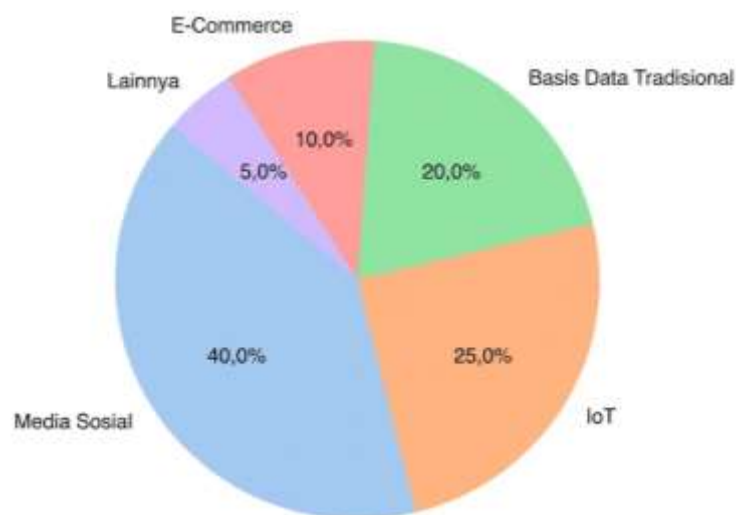
Secara paralel, kerangka kerja komputasi terdistribusi seperti Hadoop muncul, yang dirancang khusus untuk menyimpan dan memproses sejumlah besar data. Kerangka kerja ini memanfaatkan kekuatan banyak mesin yang bekerja secara bersamaan, melampaui keterbatasan sistem mesin tunggal.

Transisi dari basis data tradisional ke big data bukan hanya pergeseran teknologi. Ini mewakili perubahan yang lebih luas dalam pemahaman tentang potensi data. Dengan big data, bisnis dan peneliti tidak lagi terbatas pada melihat sampel. Mereka dapat menganalisis seluruh dataset, yang mengarah pada wawasan yang lebih kaya dan proses pengambilan keputusan yang lebih kuat.

### **Tantangan dan Peluang dalam Menangani Big Data**

Big data, terlepas dari potensinya, memiliki serangkaian tantangan. Volume data yang sangat besar dapat sangat membebani, dan tanpa alat dan strategi yang tepat, organisasi dapat dengan mudah menemukan diri mereka tenggelam dalam data tetapi kekurangan wawasan. Kualitas data adalah masalah lain. Dengan beragamnya jenis dan kecepatan data yang dihasilkan, memastikan keakuratan, konsistensi, dan keandalannya menjadi tantangan. Tanpa mengatasi masalah ini, wawasan yang diperoleh mungkin salah atau menyesatkan.

Keamanan dan privasi data merupakan masalah mendesak, terutama di era di mana pelanggaran data dan akses data tanpa izin sering terjadi. Melindungi kumpulan data besar, terutama yang berisi informasi sensitif atau pribadi, sangatlah penting.



**Gambar 11.1** Distribusi Sumber Data yang Berkontribusi pada Big Data

Di sisi positifnya, peluang yang ditawarkan oleh big data sangat besar. Bisnis dapat memperoleh wawasan mendalam tentang operasi dan perilaku pelanggan mereka, memungkinkan mereka untuk menyesuaikan strategi mereka dengan presisi yang sebelumnya tidak terbayangkan. Selain itu, sektor-sektor seperti perawatan kesehatan, perencanaan kota,

dan pembahasan telah memperoleh manfaat yang sangat besar, memanfaatkan big data untuk analitik prediktif, simulasi, dan solusi inovatif untuk masalah yang sudah lama ada.

### **Relevansi Big Data dalam Pemasaran Modern**

Di dunia digital yang sangat terhubung saat ini, pemasaran telah berubah dari fungsi yang sebagian besar bergantung pada intuisi menjadi fungsi yang sangat berakar pada pengambilan keputusan berbasis data. Big data memainkan peran penting dalam transformasi ini. Pertama, big data menawarkan pandangan komprehensif tentang pelanggan. Dengan menganalisis data dari berbagai titik kontak, bisnis dapat memahami preferensi, perilaku, dan masalah pelanggan dengan kedalaman yang luar biasa.

Segmentasi dan penargetan telah mencapai tingkat presisi baru. Alih-alih kampanye berbasis luas, pemasar dapat membuat pesan yang sangat disesuaikan, mengoptimalkannya untuk segmen mikro atau bahkan pelanggan individu. Analisis prediktif, yang didukung oleh big data, memungkinkan pemasar untuk memprediksi tren, memahami potensi pergeseran pasar, dan menyesuaikan strategi secara proaktif. Ini tidak hanya membantu dalam optimalisasi sumber daya tetapi juga dalam mendapatkan keunggulan kompetitif.

Terakhir, big data telah membuka jalan bagi pemasaran real-time. Merek sekarang dapat merespons secara instan terhadap tren atau peristiwa yang muncul, menciptakan pesan pemasaran yang sesuai dengan suasana hati atau situasi saat ini. Pada intinya, big data telah secara fundamental mengubah cara kerja pemasaran, menjadikannya lebih gesit, tepat, dan berpusat pada pelanggan daripada sebelumnya.

### **Konsep Kunci dalam Big Data**

Munculnya big data telah menghadirkan serangkaian konsep baru yang penting untuk memahami potensi dan tantangannya. Hal ini mencakup penyimpanan dan pengelolaan data, pemrosesan data, analitik data, serta privasi dan keamanan data:

- **Penyimpanan dan manajemen data.** Sistem penyimpanan dan manajemen data tradisional, seperti basis data relasional, seringkali kurang memadai dalam menangani skala dan kompleksitas big data. Untuk mengatasi tantangan ini, basis data NoSQL (misalnya, MongoDB, Cassandra) telah muncul dan menawarkan skema yang fleksibel dan skalabilitas horizontal. Sistem file terdistribusi, seperti sistem file terdistribusi Hadoop (HDFS), juga digunakan untuk menyimpan volume data yang besar di beberapa mesin.
- **Pemrosesan data.** Untuk memproses big data, khususnya secara real-time atau mendekati real-time, kerangka kerja komputasi terdistribusi (misalnya, Hadoop, Spark) umumnya digunakan. Kerangka kerja ini memungkinkan tugas pemrosesan data untuk dibagi dan dijalankan secara bersamaan di beberapa mesin, sehingga secara signifikan meningkatkan kecepatan pemrosesan.
- **Analisis data.** Analisis big data melibatkan penerapan teknik analitik canggih (misalnya, pembelajaran mesin, penambangan data) pada kumpulan data yang besar dan kompleks untuk menemukan pola, menghasilkan wawasan, dan mendukung pengambilan keputusan. Analisis ini dapat dikategorikan menjadi analitik deskriptif, diagnostik, prediktif, dan preskriptif.

- **Privasi dan keamanan data.** Dengan meningkatnya volume dan variasi data, isu-isu terkait privasi dan keamanan data menjadi semakin menonjol. Bisnis perlu menerapkan mekanisme yang kuat untuk melindungi data sensitif dan mematuhi peraturan perlindungan data yang relevan.

#### **Big Data dalam Pemasaran E-Commerce**

Dampak big data dalam industri e-commerce sangat signifikan, dengan Amazon sebagai contoh utamanya. Mereka memanfaatkan big data untuk memberikan pengalaman belanja yang dipersonalisasi, memperkirakan permintaan, dan mengoptimalkan operasi:

- **Pengalaman belanja yang dipersonalisasi.** Amazon menggunakan big data untuk melacak kebiasaan penelusuran, pembelian sebelumnya, dan preferensi penggunaannya untuk memberikan rekomendasi produk yang dipersonalisasi. Strategi ini meningkatkan pengalaman berbelanja dan membantu meningkatkan tingkat konversi.
- **Peramalan permintaan.** Amazon menggunakan big data untuk menganalisis pola pembelian historis, lalu lintas situs web, dan faktor eksternal lainnya (misalnya, hari libur, promosi) untuk secara akurat meramalkan permintaan dan mengelola inventaris. Pendekatan ini membantu meminimalkan kekurangan stok dan kelebihan stok, sehingga mengoptimalkan efisiensi operasional dan kepuasan pelanggan.
- **Optimasi operasional.** Amazon menggunakan big data untuk mengoptimalkan berbagai aspek operasinya, seperti manajemen gudang, logistik, dan layanan pelanggan. Misalnya, mereka menggunakan algoritma pembelajaran mesin untuk memprediksi lokasi optimal untuk menyimpan produk di gudang mereka, mengurangi waktu yang dibutuhkan untuk mengambil barang dan meningkatkan efisiensi operasional.

### **11.3 KERANGKA KERJA KOMPUTASI TERDISTRIBUSI**

#### **Gambaran Umum Kerangka Kerja Komputasi Terdistribusi**

Kerangka kerja komputasi terdistribusi, seperti Hadoop dan Spark, telah merevolusi cara kita memproses dan menganalisis volume data yang besar.

yang dikembangkan oleh Apache Software Foundation, adalah kerangka kerja sumber terbuka yang dirancang untuk menyimpan dan memproses sejumlah besar data di seluruh kluster komputer. Hadoop menggunakan sistem file terdistribusi (HDFS) yang memungkinkan data disimpan dalam format yang mudah diakses di seluruh jaringan sistem yang terhubung. Model pemrosesannya, MapReduce, memungkinkan pemrosesan dataset besar secara efisien dengan membagi beban kerja dan memproses data secara paralel.

juga dikembangkan oleh Apache, adalah sistem komputasi kluster yang cepat dan serbaguna. Meskipun juga mendukung pemrosesan data terdistribusi, Spark dikenal karena kemampuan komputasi dalam memorinya, yang secara signifikan mempercepat algoritma iteratif atau tugas penambangan data interaktif. Spark mendukung berbagai operasi seperti kueri SQL, streaming data, pembelajaran mesin, dan pemrosesan grafik, yang penting dalam menangani berbagai beban kerja big data.

Kerangka kerja komputasi terdistribusi ini telah menjadi penting dalam ekosistem big data karena kemampuannya untuk menangani volume data terstruktur dan tidak terstruktur yang sangat besar, skalabilitasnya, dan toleransi kesalahannya.

### **Dasar-Dasar Komputasi Terdistribusi dan Kebutuhannya**

Komputasi terdistribusi, pada intinya, adalah bidang ilmu komputer yang berfokus pada desain dan implementasi sistem yang membagi tugas dan memproses data pada banyak mesin atau node. Metodologi ini sangat kontras dengan pendekatan tradisional yang menggunakan satu komputer atau sistem untuk melakukan tugas atau menjalankan aplikasi.

Kebutuhan mendasar akan komputasi terdistribusi muncul dari berbagai tantangan:

- **Skala dan kompleksitas.** Dengan era digital yang membawa sejumlah besar data, sistem komputer tunggal tradisional sering mengalami hambatan pemrosesan dan penyimpanan. Namun, sistem terdistribusi dapat diskalakan, artinya mereka dapat menambahkan lebih banyak node ke sistem untuk mengelola peningkatan beban.
- **Toleransi kesalahan dan keandalan.** Sistem terdistribusi dirancang agar tangguh. Bahkan jika satu atau beberapa node gagal, sistem secara keseluruhan terus beroperasi, memastikan bahwa tidak ada satu titik kegagalan pun.
- **Berbagi sumber daya dan kolaborasi.** Sistem terdistribusi memungkinkan berbagi sumber daya yang kohesif dan efisien—baik itu daya komputasi, data, atau file—di wilayah geografis yang luas. Hal ini memfasilitasi kolaborasi antar entitas yang berlokasi di tempat yang berjauhan.

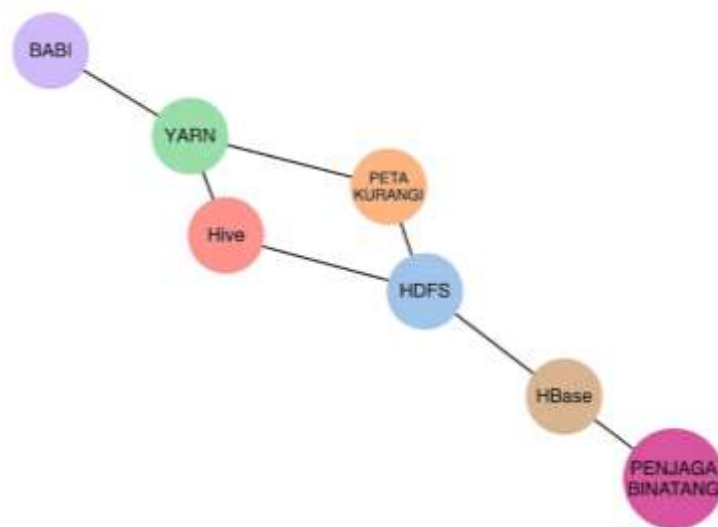
Dengan meningkatnya big data, adopsi komputasi terdistribusi telah menjadi sangat penting bagi bisnis dan peneliti.

### **Pengantar Hadoop: Hadoop Distributed File System dan MapReduce**

Hadoop, yang dikembangkan oleh Apache Software Foundation, adalah kerangka kerja penting yang dirancang untuk penyimpanan dan pemrosesan terdistribusi dari kumpulan data yang sangat besar. Ia memiliki dua komponen utama:

- **HDFS.** HDFS adalah komponen penyimpanan Hadoop, dan menyediakan struktur sistem file yang terdistribusi dan terukur. Data di HDFS dipecah menjadi blok, biasanya berukuran 128 MB atau 256 MB, yang kemudian direplikasi di beberapa node, memastikan ketahanan data dan toleransi kesalahan.
- **MapReduce.** MapReduce adalah paradigma komputasi Hadoop, yang memungkinkannya untuk memproses kumpulan data besar. Ia beroperasi dalam dua fase: fase Map dan fase Reduce. Fase Map memproses data dan menghasilkan pasangan kunci-nilai perantara. Pasangan ini kemudian diproses oleh fase Reduce untuk menghasilkan output yang diinginkan. Model ini memastikan pemrosesan paralel dan menawarkan skalabilitas.

Kombinasi HDFS dan MapReduce Hadoop menyediakan kerangka kerja yang efisien untuk menangani kumpulan data besar, memanfaatkan potensi komputasi terdistribusi (lihat Gambar 11.2).

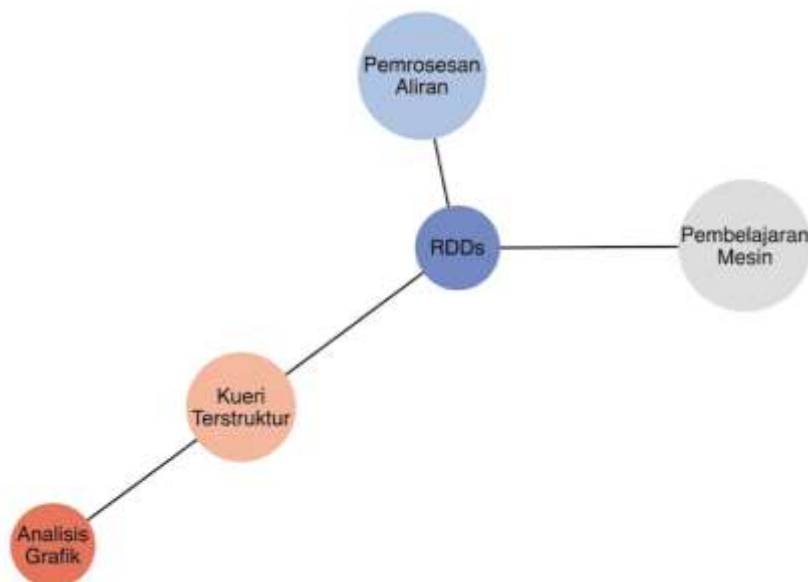


**Gambar 11.2** Ekosistem Hadoop, Termasuk Hadoop Distributed File System dan MapReduce.

### Pengantar Apache Spark dan Ekosistem Pemrosesan Data Utama

Apache Spark, proyek lain di bawah Apache Software Foundation, adalah sistem komputasi kluster yang cepat dan serbaguna. Sistem ini menawarkan solusi yang lebih canggih daripada Hadoop dalam konteks tertentu dan menghadirkan serangkaian pustaka dan fitur (lihat Gambar 11.3):

- **RDD (resilient distributed datasets).** RDD adalah struktur data dasar Spark. Ini adalah koleksi objek terdistribusi yang tidak dapat diubah yang dapat diproses secara paralel. RDD menawarkan toleransi kesalahan dengan mengingat urutan transformasi yang diterapkan pada dataset dasar, sehingga memungkinkan untuk membangun kembali data yang hilang.

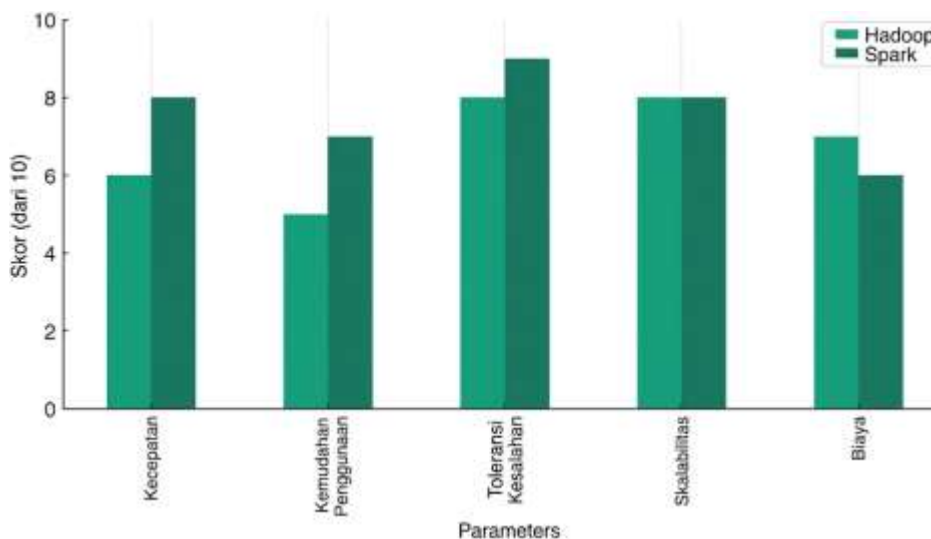


**Gambar 11.3** Apache Spark, Menyoroti Kumpulan Data Terdistribusi yang Tangguh, Spark Streaming, dan Pustaka Pembelajaran Mesin.

- **Spark streaming.** Ini adalah fitur yang memungkinkan pemrosesan aliran data langsung secara real-time. Data dapat dimasukkan dari berbagai sumber, diproses, dan kemudian didorong ke basis data, dasbor, atau sistem lain.
- **MLlib.** MLlib adalah singkatan dari pustaka pembelajaran mesin, yaitu pustaka bawaan Spark untuk tugas pembelajaran mesin. Pustaka ini menawarkan berbagai algoritma dan utilitas, sehingga memudahkan implementasi pembelajaran mesin pada kumpulan data besar.

### Membandingkan Hadoop dan Spark untuk Tugas yang Berbeda

Ketika memilih antara Hadoop dan Spark, keputusan seringkali bermuara pada persyaratan spesifik dari tugas yang ada (lihat Gambar 11.4):



**Gambar 11.4** Perbandingan Antara Hadoop dan Spark, Berfokus pada Kecepatan, Kemudahan Penggunaan, Toleransi Kesalahan, dan Parameter Lainnya.

- **Kecepatan pemrosesan data.** Meskipun Hadoop dan Spark menawarkan pemrosesan terdistribusi, Spark seringkali mengungguli Hadoop dalam hal kecepatan, terutama karena kemampuan komputasi dalam memorinya. Oleh karena itu, untuk tugas-tugas yang membutuhkan pemrosesan waktu nyata, Spark umumnya lebih cocok.
- **Kemudahan penggunaan.** Spark, dengan antarmuka pemrograman aplikasi tingkat tingginya, dianggap lebih ramah bagi pengembang daripada MapReduce Hadoop. Spark mendukung berbagai bahasa seperti Java, Scala, dan Python, menawarkan fleksibilitas dalam pengembangan.
- **Efisiensi biaya.** HDFS Hadoop adalah sistem penyimpanan berbasis disk, sehingga terkadang lebih hemat biaya untuk menyimpan dataset besar dibandingkan dengan pendekatan memori Spark.
- **Kompleksitas tugas.** Untuk tugas iteratif yang kompleks, terutama di bidang pembelajaran mesin, Spark, dengan MLlib-nya, menyediakan solusi yang lebih canggih



dan efisien. Namun, untuk tugas pemrosesan dan penyimpanan data yang lebih sederhana, Hadoop sudah cukup.

- **Integrasi dan kompatibilitas.** Hadoop dan Spark kompatibel satu sama lain. Spark dapat menggunakan HDFS untuk penyimpanan, dan Hadoop dapat memanfaatkan Spark untuk komputasi. Fleksibilitas ini memungkinkan organisasi untuk memilih yang terbaik dari kedua dunia, tergantung pada kebutuhan spesifik mereka.

Singkatnya, Hadoop dan Spark menawarkan keunggulan unik. Pilihan di antara keduanya harus didasarkan pada persyaratan spesifik dari tugas pemrosesan data, skala data, dan hasil yang diinginkan.

### **Konsep Kunci dalam Kerangka Kerja Komputasi Terdistribusi**

Beberapa konsep kunci mendasari operasi dan fungsionalitas kerangka kerja komputasi terdistribusi seperti Hadoop dan Spark:

- **Penyimpanan terdistribusi.** Ini adalah prinsip menyimpan data di beberapa node dalam jaringan, bukan di basis data terpusat. Di Hadoop, ini dicapai melalui HDFS. HDFS membagi file data besar menjadi blok yang lebih kecil, mendistribusikannya ke seluruh node dalam cluster, dan mempertahankan redundansi untuk toleransi kesalahan.
- **MapReduce.** Ini adalah model pemrograman untuk memproses kumpulan data besar dengan algoritma paralel dan terdistribusi pada cluster. Fungsi Map memproses blok data dan menghasilkan serangkaian pasangan kunci-nilai perantara. Fungsi Reduce menggabungkan semua nilai perantara yang terkait dengan kunci yang sama.
- **Komputasi dalam memori.** Spark menggunakan komputasi dalam memori, yang menyimpan data di RAM di seluruh jaringan terdistribusi, memungkinkan pengambilan dan analisis data yang lebih cepat dibandingkan dengan penyimpanan berbasis disk. Ini sangat efektif untuk algoritma iteratif dalam pembelajaran mesin dan tugas penambahan data interaktif.
- **Toleransi kesalahan.** Hadoop dan Spark dirancang untuk memiliki toleransi kesalahan. Ini berarti mereka dapat terus beroperasi bahkan jika node individual gagal. Hadoop mencapai hal ini melalui replikasi data di HDFS, dan Spark menggunakan RDD yang melacak informasi silsilah data untuk membangun kembali data yang hilang.
- **Skalabilitas.** Kerangka kerja komputasi terdistribusi dapat dengan mudah diskalakan untuk menangani lebih banyak data dengan menambahkan lebih banyak node ke jaringan. Kemampuan untuk memperluas kapasitas ini membuat kerangka kerja ini cocok untuk pemrosesan big data.

### **Teknologi Cloud dalam Big Data dan Analisis Real-time**

Munculnya dan evolusi teknologi cloud telah menjadi pengubah permainan dalam cara big data dikelola, diproses, dan dianalisis. Platform cloud seperti Amazon Web Services (AWS), Microsoft Azure, dan Google Cloud Platform (GCP) telah mendemokratisasi akses ke sumber daya komputasi berkinerja tinggi, memungkinkan bisnis dari semua ukuran untuk memanfaatkan analitik big data dan pemrosesan real-time tanpa investasi awal yang signifikan dalam infrastruktur.

menawarkan serangkaian layanan yang dirancang untuk solusi big data, seperti Amazon S3 untuk penyimpanan, Amazon Redshift untuk gudang data, dan Amazon Kinesis untuk streaming dan analitik data real-time. Layanan ini dapat diskalakan dan dapat menangani sejumlah besar data, memberikan fleksibilitas kepada bisnis untuk hanya membayar apa yang mereka gunakan.

menyediakan berbagai layanan serupa, dengan Azure Blob Storage untuk penyimpanan data, Azure Synapse Analytics untuk analisis big data, dan Azure Stream Analytics untuk pemrosesan peristiwa real-time. Lingkungan terintegrasi Azure memungkinkan analitik data hibrida yang mulus, menggabungkan data on-premises dan cloud.

terkenal dengan layanan BigQuery-nya, sebuah gudang data multi-cloud tanpa server, sangat skalabel, dan hemat biaya yang memungkinkan kueri SQL super cepat menggunakan daya pemrosesan infrastruktur Google. Selain itu, layanan Pub/Sub dan Dataflow GCP menawarkan kemampuan yang kuat untuk pipeline analitik real-time.

Teknologi cloud juga secara signifikan memengaruhi analitik real-time dengan menyediakan alat yang dapat memproses dan menganalisis aliran data dengan latensi minimal. Alat-alat ini memungkinkan bisnis untuk membuat keputusan berbasis data lebih cepat, seringkali secara real-time, sehingga membuka peluang untuk personalisasi instan dan strategi keterlibatan pelanggan yang dinamis.

Keuntungan teknologi cloud dalam big data dan analitik real-time meliputi hal-hal berikut:

- **Skalabilitas.** Layanan cloud dapat ditingkatkan atau diturunkan skalanya berdasarkan permintaan, yang ideal untuk aplikasi big data yang mungkin memiliki beban kerja yang bervariasi.
- **Efektivitas biaya.** Dengan model harga bayar sesuai penggunaan, perusahaan dapat mengoptimalkan biaya tanpa mengorbankan kemampuan komputasi.
- **Kinerja.** Penyedia cloud berinvestasi besar-besaran dalam teknologi dan infrastruktur terbaru, memastikan kinerja dan ketersediaan yang tinggi.
- **Aksesibilitas.** Platform cloud memudahkan bisnis untuk mengakses dan berkolaborasi dalam proyek big data secara global.

Singkatnya, integrasi teknologi cloud telah menjadi penting dalam lanskap big data dan analitik real-time, mendorong inovasi dan menyediakan solusi yang terukur dan hemat biaya bagi bisnis. Sifat elastis sumber daya cloud selaras dengan persyaratan dinamis pemrosesan big data, sedangkan layanan analitik canggih yang tersedia di cloud memberdayakan organisasi untuk mendapatkan wawasan yang dapat ditindaklanjuti lebih cepat dari sebelumnya.

Dengan memanfaatkan infrastruktur dan layanan yang kuat yang disediakan oleh penyedia cloud utama seperti AWS, Microsoft Azure, dan GCP, bisnis dapat fokus pada perolehan nilai dari data mereka daripada mengelola teknologi yang mendasarinya. Pergeseran ini tidak hanya meningkatkan efisiensi operasional tetapi juga mempercepat jalur dari data ke pengambilan keputusan, memungkinkan personalisasi waktu nyata dan strategi pemasaran yang responsif.

Evolusi menuju solusi big data berbasis cloud mewakili pergeseran paradigma dalam cara data diproses, disimpan, dan dianalisis, menandai era baru ketangkasan dan inovasi di

bidang pemasaran. Seiring dengan terus berkembangnya teknologi cloud, teknologi ini tidak diragukan lagi akan membentuk tulang punggung analitik pemasaran, mendorong wawasan yang lebih prediktif, dapat ditindaklanjuti, dan personal daripada sebelumnya.

### **Menggunakan Hadoop dan Spark untuk Analitik Pemasaran**

Hadoop dan Spark sering digunakan untuk menangani dan memproses sejumlah besar data dalam analitik pemasaran. Keduanya dapat memberikan wawasan berharga yang dapat membantu mengoptimalkan strategi pemasaran.

Misalnya, sebuah perusahaan ritel internasional mungkin mengumpulkan data pelanggan dari berbagai saluran, termasuk platform belanja online, pembelian di toko, keterlibatan media sosial, dan interaksi layanan pelanggan. Data ini dapat berupa berbagai format, seperti teks, gambar, dan data terstruktur. Volume data dapat dengan mudah mencapai beberapa terabyte atau bahkan petabyte, menjadikannya kasus penggunaan yang sempurna untuk kemampuan penyimpanan terdistribusi Hadoop.

Pertama, data mentah disimpan di HDFS, di mana data tersebut dipecah menjadi blok-blok yang mudah dikelola dan didistribusikan ke seluruh node dalam kluster Hadoop. Sistem ini menawarkan redundansi dan toleransi kesalahan, memastikan bahwa tidak ada data yang hilang bahkan jika satu atau lebih node gagal.

Selanjutnya, perusahaan menggunakan MapReduce di Hadoop untuk memproses dan membersihkan data. Misalnya, fungsi Map dapat digunakan untuk menyaring data yang tidak relevan, sementara fungsi Reduce dapat meringkas data ke dalam format yang lebih mudah dikelola, seperti total penjualan harian untuk setiap toko.

Setelah tahap pra-pemrosesan dan pembersihan, perusahaan menggunakan Spark untuk melakukan analisis yang lebih kompleks. Berkat kemampuan komputasi dalam memori Spark, ia dapat dengan cepat memproses kumpulan data besar untuk memberikan wawasan secara real-time. Sebagai contoh, perusahaan dapat menggunakan MLlib dari Spark untuk membangun model segmentasi pelanggan, mengelompokkan pelanggan berdasarkan perilaku pembelian, riwayat penelusuran, dan informasi demografis.

Dengan memanfaatkan Hadoop dan Spark, perusahaan dapat mengubah data pelanggan yang sangat besar menjadi wawasan yang dapat ditindaklanjuti, seperti mengidentifikasi segmen pelanggan yang paling berharga, menyesuaikan kampanye pemasaran untuk menargetkan kelompok tertentu, dan mengoptimalkan penawaran produk berdasarkan preferensi pelanggan.

## **11.4 PENGGUNAAN TEKNOLOGI DAN METODE CANGGIH**

### **Gambaran Umum Analitik Real-Time**

Analitik real-time adalah penggunaan teknologi dan metode canggih untuk menganalisis data segera setelah data tersebut dihasilkan atau dikumpulkan. Tidak seperti analitik tradisional, yang seringkali melibatkan analisis data historis, analitik real-time memungkinkan interpretasi dan tindakan langsung, menawarkan keunggulan kompetitif yang signifikan bagi bisnis.

Analitik real-time menjadi semakin penting karena volume, kecepatan, dan variasi data telah meledak dalam beberapa tahun terakhir. Dengan munculnya Internet of Things (IoT), media sosial, dan teknologi digital lainnya, bisnis sekarang dapat mengumpulkan sejumlah besar data dengan kecepatan yang belum pernah terjadi sebelumnya. Data ini, ketika dianalisis secara real-time, dapat memberikan wawasan berharga tentang perilaku pelanggan, tren pasar, dan efisiensi operasional.

Beberapa sektor, termasuk keuangan, perawatan kesehatan, manufaktur, dan pemasaran, telah menerapkan analitik real-time. Dalam pemasaran, misalnya, analitik waktu nyata memungkinkan pemasar untuk memantau perilaku dan keterlibatan pelanggan secara real-time, sehingga memungkinkan mereka untuk mempersonalisasi pesan pemasaran, mengoptimalkan kinerja kampanye, dan meningkatkan layanan pelanggan.

Analitik waktu nyata dapat difasilitasi oleh berbagai alat dan teknologi. Misalnya, platform pemrosesan aliran data seperti Apache Kafka dapat memproses data berkecepatan tinggi secara real-time, dan alat intelijen bisnis real-time dapat menyediakan analitik dan visualisasi secara real-time. Selain itu, teknik pembelajaran mesin dan kecerdasan buatan (AI) dapat digunakan untuk menganalisis pola data yang kompleks dan membuat prediksi secara real-time.

Terlepas dari manfaatnya, analitik waktu nyata juga menimbulkan beberapa tantangan, termasuk masalah privasi dan keamanan data, kebutuhan akan infrastruktur TI yang kuat dan terukur, dan persyaratan untuk keterampilan analitis tingkat lanjut.

### **Kafka, Storm, dan Elasticsearch**

Analisis waktu nyata memerlukan penggunaan alat dan platform canggih yang dapat menangani kecepatan, volume, dan kompleksitas aliran data. Di antara berbagai platform yang tersedia, Kafka, Storm, dan Elasticsearch telah muncul sebagai beberapa solusi terkemuka di bidang ini:

- **Kafka.** Dikembangkan oleh LinkedIn dan kemudian disumbangkan ke Apache Software Foundation, Kafka adalah platform streaming peristiwa terdistribusi yang dirancang untuk throughput tinggi, toleransi kesalahan, dan skalabilitas. Terutama digunakan untuk membangun pipeline data waktu nyata dan aplikasi streaming, Kafka bekerja dengan memungkinkan produsen untuk mengirim pesan ke topik tertentu, yang kemudian dapat di berlangganan oleh konsumen, memfasilitasi komunikasi waktu nyata dan streaming data antar sistem.
- **Storm.** Apache Storm adalah sistem komputasi waktu nyata terdistribusi yang memungkinkan pemrosesan volume besar data berkecepatan tinggi. Dengan kemampuannya memproses aliran data, Storm cocok untuk analitik waktu nyata, pemantauan, dan tugas ekstraksi, transformasi, dan pemuatan (ETL). Modelnya didasarkan pada spout (sumber data) dan bolt (prosesor data), memungkinkan pengembang untuk mengatur topologi untuk transformasi dan pemrosesan data.
- **Elasticsearch.** Sebagai mesin pencarian dan analitik sumber terbuka, Elasticsearch mahir dalam menangani tugas pengindeksan dan pencarian waktu nyata. Dibangun di atas pustaka Lucene, Elasticsearch umumnya digunakan untuk analisis data log atau

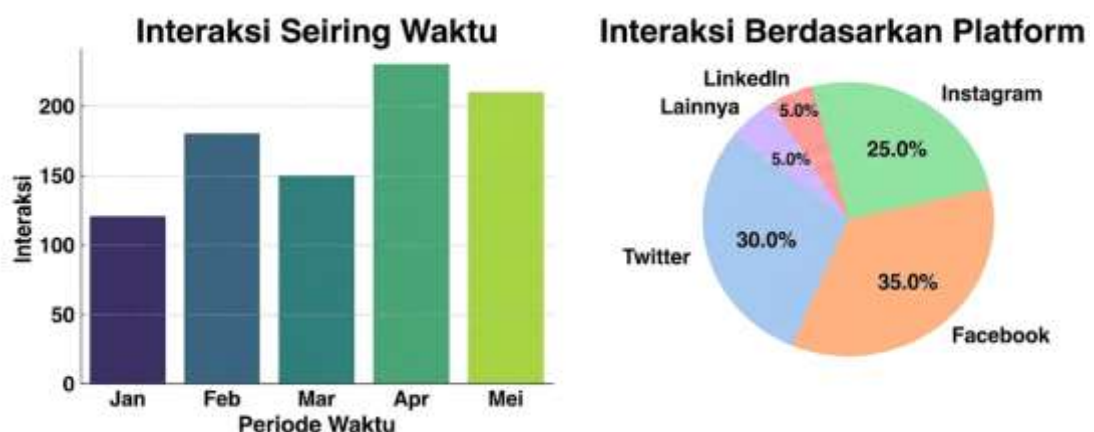
peristiwa karena kemampuannya untuk menanyakan sejumlah besar data terstruktur dan tidak terstruktur dengan cepat.

Kombinasi alat-alat ini menyediakan rangkaian yang ampuh untuk analitik waktu nyata. Misalnya, seseorang dapat menggunakan Kafka untuk mengalirkan data log, memprosesnya secara waktu nyata menggunakan Storm, dan kemudian mengindeks dan menanyakannya menggunakan Elasticsearch.

### **Menerapkan Dasbor dan Peringatan Waktu Nyata**

Dasbor dan peringatan waktu nyata merupakan komponen penting bagi bisnis yang membutuhkan wawasan instan dan respons tepat waktu. Mereka menawarkan representasi visual dari data streaming, memungkinkan pemangku kepentingan untuk membuat keputusan yang tepat secara instan.

- **Membangun dasbor waktu nyata.** Dengan memanfaatkan platform seperti Elasticsearch dengan Kibana, pengguna dapat mengimplementasikan dasbor waktu nyata. Dasbor ini dapat memvisualisasikan berbagai metrik, mulai dari indikator kinerja sistem hingga metrik keterlibatan pelanggan. Kunci dari dasbor yang efektif bukan hanya tampilan data waktu nyata tetapi juga kemudahan interpretasi, memastikan bahwa pemangku kepentingan dapat segera memahami signifikansi data.
- **Menyiapkan peringatan.** Bersamaan dengan analitik waktu nyata, kemampuan untuk menyiapkan peringatan langsung sangat penting. Misalnya, dengan menggunakan potensi komputasi waktu nyata Storm, sistem dapat memantau aliran data untuk anomali atau peristiwa tertentu. Setelah peristiwa tersebut terdeteksi, peringatan dapat dipicu, mengirimkan pemberitahuan kepada pemangku kepentingan yang relevan melalui SMS, email, atau metode komunikasi lainnya (lihat Gambar 11.5).



**Gambar 11.5** Dasbor Real-Time yang Menampilkan Analitik dari Pemasaran Media Sosial.

### **Tantangan dalam Pemrosesan dan Analisis Data Real-Time**

Meskipun pemrosesan dan analisis data real-time menawarkan keuntungan yang signifikan, hal ini juga menghadirkan tantangan:

- **Volume dan kecepatan data.** Salah satu tantangan utama adalah skala besar data yang mengalir secara real-time. Sistem harus dilengkapi untuk menangani aliran data yang

masif, memastikan bahwa tidak ada data yang hilang dan pemrosesan terjadi tanpa jeda.

- **Kualitas dan kebenaran data.** Aliran data real-time mungkin mengandung inkonsistensi, noise, atau kesalahan. Menerapkan mekanisme pembersihan dan validasi data yang efektif secara real-time menjadi sangat penting untuk memastikan analitik yang akurat.
- **Skalabilitas infrastruktur.** Infrastruktur harus dapat diskalakan secara dinamis dengan beban data yang berfluktuasi. Sistem statis mungkin kesulitan selama puncak masuknya data, yang berpotensi menyebabkan kehilangan data atau kerusakan sistem.
- **Pemrosesan peristiwa kompleks.** Mengidentifikasi pola atau peristiwa spesifik dalam aliran data, terutama ketika tersebar di beberapa aliran, adalah tugas yang kompleks. Sistem perlu dirancang untuk pemrosesan peristiwa kompleks tersebut untuk mengekstrak wawasan yang bermakna secara instan.
- **Kekhawatiran keamanan dan privasi.** Pemrosesan real-time berarti data sering ditransmisikan antar sistem atau melalui jaringan. Memastikan keamanan data ini dan mengatasi masalah privasi menjadi sangat penting, terutama dengan kerangka peraturan seperti GDPR yang menekankan perlindungan data.

Mengatasi tantangan ini membutuhkan kombinasi infrastruktur yang kuat, algoritma yang efisien, dan strategi manajemen data yang efektif (lihat Tabel 11.2).

**Tabel 11.2** Tantangan dalam Pemrosesan dan Analisis Data Real-Time, Implikasinya, dan Solusi atau Cara Mengatasi yang Diusulkan.

| Tantangan       | Implikasi                                               | Solusi/Alternatif                                        |
|-----------------|---------------------------------------------------------|----------------------------------------------------------|
| Latensi data    | Keterlambatan dalam memperoleh wawasan secara real-time | Pengoptimalan alur data<br>Pengurangan transformasi data |
| Integritas data | Kualitas data terganggu                                 | Pemeriksaan validasi data<br>Alat pembersihan data       |
| Skalabilitas    | Tidak mampu menangani volume data yang tinggi           | Sistem terdistribusi<br>Penskalaan horizontal            |
| Keamanan data   | Potensi pelanggaran data                                | Enkripsi data<br>Audit keamanan berkala                  |

### Studi Kasus dalam Analisis Real-Time

#### Deteksi Penipuan Real-Time di Bidang Keuangan

Industri keuangan merupakan target utama bagi para penipu, khususnya di bidang transaksi online. Seiring meningkatnya transaksi digital, taktik penipuan juga semakin canggih. Sistem deteksi penipuan tradisional yang bergantung pada data historis dapat tertinggal dan mungkin tidak mendeteksi pola penipuan baru dengan cukup cepat.

Sebuah bank terkemuka menerapkan sistem analitik real-time untuk mengidentifikasi dan menghentikan transaksi mencurigakan saat terjadi. Dengan menggunakan platform pemrosesan aliran data seperti Apache Kafka, bank tersebut memasukkan data transaksi secara real-time. Model pembelajaran mesin yang dilatih berdasarkan pola penipuan historis

menilai setiap transaksi. Jika suatu transaksi dianggap berpotensi curang, transaksi tersebut akan dihentikan untuk peninjauan manual atau pengguna akan segera diberitahu untuk verifikasi.

Dengan menerapkan sistem analitik waktu nyata ini, bank tersebut melihat pengurangan signifikan dalam aktivitas penipuan. Pelanggan menghargai umpan balik langsung, yang melindungi dana mereka dan meningkatkan kepercayaan mereka pada langkah-langkah keamanan bank. Selain itu, sistem ini memungkinkan bank untuk beradaptasi dengan cepat terhadap pola penipuan baru, menawarkan mekanisme pertahanan dinamis daripada yang statis.

### **Optimasi Rantai Pasokan Waktu Nyata di Manufaktur**

Di bidang manufaktur, keterlambatan atau inefisiensi dalam rantai pasokan dapat mengakibatkan kerugian finansial yang signifikan. Sistem manajemen rantai pasokan tradisional, yang seringkali bekerja berdasarkan data historis, mungkin tidak bereaksi cukup cepat terhadap perubahan atau gangguan mendadak.

Sebuah perusahaan manufaktur global mengintegrasikan analitik waktu nyata ke dalam manajemen rantai pasokannya. Sensor IoT ditempatkan pada peralatan dan di gudang untuk melacak tingkat produk, efisiensi peralatan, dan status pengiriman. Saat data mengalir masuk, alat analitik waktu nyata memproses informasi ini, memprediksi potensi hambatan, kebutuhan pemeliharaan, atau kekurangan stok. Peringatan dikirim ke manajer segera setelah terdeteksi adanya potensi masalah.

Perusahaan manufaktur tersebut mengalami beberapa manfaat dari sistem ini. Terjadi pengurangan waktu henti peralatan yang signifikan karena peringatan pemeliharaan prediktif. Tingkat stok dioptimalkan, mengurangi biaya penyimpanan dan mencegah penghentian produksi karena kekurangan material. Perusahaan juga lebih siap untuk bereaksi terhadap gangguan eksternal, seperti keterlambatan transportasi, dengan menyesuaikan jadwal produksi atau mencari pemasok alternatif secara real-time. Pendekatan proaktif ini menghasilkan operasi yang lebih lancar dan penghematan biaya.

### **Konsep Kunci dalam Analitik Real-Time**

Terdapat beberapa konsep dan komponen kunci yang mendasari fungsi analitik real-time:

- **Streaming data.** Streaming data melibatkan aliran data berkelanjutan dari berbagai sumber seperti media sosial, aliran klik situs web, sensor IoT, dan sebagainya. Data ini diproses secara berurutan dan bertahap berdasarkan catatan demi catatan atau melalui jendela waktu geser.
- **Pemrosesan data real-time.** Tujuan pemrosesan data real-time adalah untuk mengambil tindakan sebagai respons terhadap suatu peristiwa dalam kerangka waktu yang ditetapkan, seringkali dalam beberapa detik atau kurang. Hal ini dapat melibatkan pemrosesan peristiwa kompleks, di mana beberapa aliran data dari berbagai sumber dianalisis untuk mengidentifikasi peristiwa atau pola yang bermakna.
- **Analisis aliran data.** Analisis aliran data melibatkan analisis aliran data waktu nyata untuk tujuan pengambilan keputusan. Ini penting untuk keterlibatan pelanggan waktu nyata, deteksi penipuan, dan optimasi operasional.

- **Komputasi dalam memori.** Komputasi dalam memori menyimpan data di RAM di seluruh jaringan terdistribusi yang memungkinkan pemrosesan dan analisis waktu nyata yang cepat. Ini merupakan komponen penting dari analitik waktu nyata karena memungkinkan pemrosesan data berkecepatan tinggi.
- **Arsitektur berbasis peristiwa.** Arsitektur berbasis peristiwa adalah paradigma desain sistem di mana keputusan waktu nyata dibuat berdasarkan deteksi pola peristiwa kompleks.

Memahami konsep-konsep ini sangat penting untuk mengimplementasikan analitik waktu nyata dalam pemasaran dan aplikasi bisnis lainnya, memungkinkan organisasi untuk merespons dengan cepat terhadap perubahan kondisi pasar dan perilaku pelanggan.

#### **Analisis Waktu Nyata dalam Pemasaran Media Sosial**

Salah satu ilustrasi terbaik dari analisis waktu nyata dapat ditemukan di ranah pemasaran media sosial. Platform media sosial menghasilkan banyak data yang terus diperbarui, menjadikannya lingkungan yang ideal untuk analisis waktu nyata.

Sebagai contoh, pertimbangkan sebuah perusahaan global yang meluncurkan produk baru dan menggunakan kampanye pemasaran berbasis tagar di X (sebelumnya Twitter). Perusahaan tersebut dapat menggunakan analisis waktu nyata untuk melacak penggunaan dan penyebaran tagar tersebut. Mereka dapat memantau seberapa cepat tagar tersebut dibagikan, dari mana tagar tersebut dibagikan, dan oleh siapa. Hal ini memungkinkan mereka untuk menyesuaikan strategi pemasaran mereka secara waktu nyata, memanfaatkan apa yang berhasil dan mengatasi area yang menjadi perhatian.

Selain itu, analitik waktu nyata juga dapat digunakan untuk analisis sentimen. Perusahaan dapat memantau sentimen keseluruhan dari tweet yang berisi tagar kampanye. Jika sentimen mulai berubah menjadi negatif, mereka dapat dengan cepat mengidentifikasi masalah dan mengatasinya sebelum meningkat.

Selain itu, analitik waktu nyata dapat digunakan untuk mengidentifikasi pengguna media sosial berpengaruh yang berinteraksi dengan kampanye. Perusahaan kemudian dapat berinteraksi dengan para influencer ini secara waktu nyata, berpotensi memanfaatkan jangkauan mereka untuk penyebaran kampanye lebih lanjut.

Alat analitik waktu nyata untuk pemasaran media sosial dapat berkisar dari alat bawaan di platform media sosial seperti Facebook Insights dan Twitter Analytics hingga platform mandiri yang lebih canggih seperti Brandwatch dan Hootsuite, yang menyediakan analisis waktu nyata yang lebih komprehensif dan terperinci.

### **11.5 PERSONALISASI DAN PEMASARAN WAKTU NYATA**

#### **Gambaran Umum Personalisasi dan Pemasaran Waktu Nyata**

Personalisasi dan pemasaran waktu nyata telah muncul sebagai strategi penting dalam lanskap pemasaran digital. Hal ini mewakili pergeseran dari pendekatan satu ukuran untuk semua ke interaksi yang lebih personal dan real-time dengan pelanggan. Personalisasi mengacu pada strategi menyesuaikan produk, layanan, dan komunikasi kepada pelanggan individu berdasarkan preferensi, perilaku, dan informasi real-time mereka. Hal ini sering



didorong oleh data dan analitik prediktif dan dapat diterapkan pada berbagai aspek pemasaran, mulai dari rekomendasi produk yang dipersonalisasi dan iklan yang ditargetkan hingga email dan konten situs web yang dipersonalisasi.

Namun, pemasaran real-time melibatkan merek yang bereaksi terhadap peristiwa, interaksi pelanggan, atau tren secara real-time, seringkali melalui media sosial atau saluran digital. Bentuk pemasaran ini spontan dan langsung, dirancang untuk terhubung dengan pelanggan pada saat yang tepat dengan pesan yang tepat. Hal ini sering membutuhkan kemampuan untuk menganalisis dan bertindak berdasarkan data secara real-time. Integrasi personalisasi dan pemasaran real-time dapat menghasilkan strategi pemasaran yang sangat efektif. Misalnya, pelanggan yang menjelajahi situs web perusahaan dapat menerima rekomendasi produk yang dipersonalisasi berdasarkan perilaku penjelajahan mereka, dan rekomendasi ini dapat diperbarui secara real-time saat pelanggan berinteraksi dengan situs web.

Pendekatan ini dapat meningkatkan keterlibatan pelanggan, meningkatkan kepuasan pelanggan, dan pada akhirnya, meningkatkan tingkat konversi dan pendapatan. Selain itu, hal ini memungkinkan perusahaan untuk membedakan diri di pasar yang ramai dan membangun hubungan yang lebih kuat dengan pelanggan mereka.

#### **Kebutuhan Personalisasi Real-Time**

Di era digital saat ini, konsumen dikelilingi oleh banyak informasi dan pilihan. Rentang perhatian mereka telah memendek, dan harapan mereka akan relevansi telah meningkat. Personalisasi real-time mengatasi tantangan ini secara langsung. Ini adalah metode untuk menyesuaikan konten, produk, atau rekomendasi layanan kepada pengguna individu secara real-time berdasarkan perilaku mereka saat ini, interaksi masa lalu, dan informasi kontekstual lainnya.

Manfaat personalisasi real-time sangat beragam. Pertama, ini meningkatkan pengalaman pengguna. Ketika pengguna merasa bahwa konten tersebut disesuaikan secara khusus untuk mereka, hal itu mengarah pada interaksi yang lebih menarik dan bermakna. Selain itu, personalisasi real-time dapat secara signifikan memengaruhi keuntungan perusahaan. Dengan menyajikan produk atau layanan yang lebih mungkin diminati pengguna pada saat itu juga, perusahaan dapat meningkatkan rasio konversi, meningkatkan nilai transaksi rata-rata, dan memupuk loyalitas pelanggan.

Selain itu, dalam lingkungan di mana preferensi pelanggan dapat berubah dengan cepat, dan faktor eksternal seperti peristiwa atau tren terkini dapat memengaruhi perilaku, personalisasi waktu nyata memastikan bahwa bisnis tetap gesit, menyesuaikan penawaran mereka sesuai dengan lanskap konsumen yang terus berkembang.

#### **Teknik untuk Segmentasi dan Penargetan Secara Langsung**

Segmentasi dan penargetan secara langsung melibatkan pengklasifikasian pengguna ke dalam segmen spesifik berdasarkan data waktu nyata dan kemudian menyediakan konten atau penawaran yang disesuaikan untuk segmen tersebut secara langsung (lihat Tabel 11.3). Model pembelajaran mesin, khususnya algoritma pengelompokan, telah menjadi instrumen penting dalam mencapai hal ini. Misalnya, teknik pembelajaran mesin tanpa pengawasan dapat

menganalisis perilaku pengguna selama sesi dan menempatkan mereka ke dalam segmen yang telah menunjukkan perilaku serupa di masa lalu.

**Tabel 11.3** Teknik untuk Segmentasi dan Penargetan Secara Langsung, Beserta Kelebihan, Kekurangan, dan Kasus Penggunaan Idealnya.

| Teknik                            | Keunggulan                                                         | Kelemahan                                          | Kasus Penggunaan Ideal                         |
|-----------------------------------|--------------------------------------------------------------------|----------------------------------------------------|------------------------------------------------|
| <b>Pelacakan perilaku</b>         | Real-time user insights<br>Personalized content                    | Kekhawatiran privasi<br>Tantangan penyimpanan data | Situs web e-commerce<br>Platform konten        |
| <b>Analisis prediktif</b>         | Memprediksi perilaku pengguna<br>Peningkatan keterlibatan pengguna | Membutuhkan data historis<br>Ketidakakuratan model | Pemasaran email<br>Rekomendasi produk          |
| <b>Penargetan berbasis lokasi</b> | Konten yang relevan secara geografis<br>Peningkatan konversi       | Membutuhkan akses lokasi<br>Pelanggaran privasi    | Penawaran/promosi lokal<br>Aplikasi perjalanan |

Teknik lain melibatkan analisis perjalanan pengguna atau jalur klik secara real-time untuk menentukan niat mereka dan kemudian memberikan rekomendasi konten atau produk yang sesuai dengan niat tersebut. Misalnya, jika pengguna di situs e-commerce memeriksa beberapa ulasan produk dan kemudian beralih ke halaman perbandingan harga, mereka mungkin lebih dekat untuk membuat keputusan pembelian.

#### **Menggunakan Data Real-Time untuk Optimasi Konten Dinamis**

Optimasi konten dinamis mengacu pada penyesuaian konten situs web atau aplikasi secara real-time berdasarkan perilaku pengguna, preferensi, atau faktor eksternal. Salah satu metode yang menonjol melibatkan pengujian A/B di mana berbagai versi konten disajikan kepada pengguna, dan reaksi mereka diukur secara real-time. Versi yang menghasilkan keterlibatan pengguna atau konversi yang lebih baik menjadi pilihan yang disukai.

Pendekatan lain melibatkan penggunaan analitik prediktif untuk menentukan konten apa yang paling mungkin diinteraksikan oleh pengguna tertentu dan kemudian menyajikan konten tersebut secara dinamis. Misalnya, jika portal berita online memahami dari perilaku masa lalu bahwa pengguna tertarik pada berita teknologi dan olahraga, halaman depan untuk pengguna tersebut mungkin memprioritaskan artikel dari kategori ini.

Siklus umpan balik waktu nyata, di mana interaksi pengguna dengan konten segera dianalisis untuk menyempurnakan dan menyesuaikan strategi konten, juga sangat penting. Misalnya, jika suatu konten menghasilkan tingkat keterlibatan yang tinggi, konten tersebut dapat dipromosikan lebih menonjol.

#### **Studi Kasus: Kampanye Pemasaran Waktu Nyata yang Sukses**

- **Tweet Oreo di Super Bowl.** Selama Super Bowl 2013, terjadi pemadaman listrik yang tidak terduga. Memanfaatkan momen tersebut, Oreo men-tweet, “Anda masih bisa mencelupkan dalam gelap,” yang menjadi sensasi. Reaksi pemasaran waktu nyata ini

menunjukkan ketangkasan merek dan kemampuan untuk terhubung dengan audiens besar selama acara langsung.

- **Thumbnail yang dipersonalisasi Netflix.** Mesin rekomendasi Netflix sangat terkenal, tetapi mereka membawa personalisasi selangkah lebih maju dengan optimasi thumbnail dinamis. Dengan menganalisis riwayat tontonan, Netflix menyesuaikan thumbnail acara dan film untuk memenuhi preferensi pengguna individu, sehingga meningkatkan rasio klik-tayang.
- **Kampanye “Tahun dalam Musik” Spotify.** Dengan memanfaatkan data pengguna, Spotify memberikan penggunanya rangkuman “Tahun dalam Musik” yang dipersonalisasi, yang merinci lagu yang paling sering diputar, genre, dan total waktu mendengarkan. Sentuhan personal ini menumbuhkan hubungan yang lebih dalam antara pengguna dan platform.

Studi kasus ini menggarisbawahi potensi pemasaran waktu nyata dan personalisasi serta bagaimana keduanya dapat menumbuhkan koneksi yang tulus, meningkatkan loyalitas merek, dan mendorong keterlibatan pengguna.

#### **Konsep Kunci dalam Personalisasi dan Pemasaran Waktu Nyata**

Beberapa konsep kunci mendasari pelaksanaan strategi personalisasi dan pemasaran waktu nyata yang efektif:

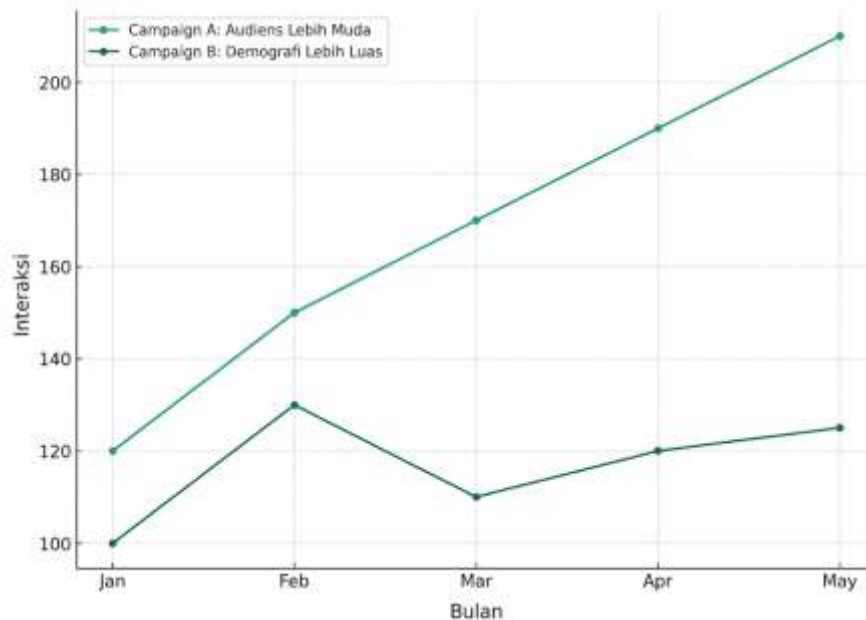
- **Segmentasi pelanggan.** Ini melibatkan pengelompokan pelanggan berdasarkan berbagai faktor seperti demografi, perilaku, minat, dan banyak lagi. Ini adalah aspek mendasar dari personalisasi karena memungkinkan upaya pemasaran yang ditargetkan.
- **Pelacakan perilaku.** Ini termasuk memantau perilaku pelanggan di berbagai saluran, seperti situs web, media sosial, dan email. Data yang dikumpulkan memberikan wawasan berharga untuk personalisasi.
- **Konten dinamis.** Ini adalah konten yang berubah berdasarkan perilaku, preferensi, atau faktor waktu nyata pengguna. Ini sangat penting untuk personalisasi dan pemasaran waktu nyata.
- **Analisis prediktif.** Ini adalah penggunaan data, algoritma statistik, dan teknik pembelajaran mesin untuk mengidentifikasi kemungkinan hasil di masa mendatang. Ini dapat digunakan untuk memprediksi perilaku pelanggan dan memungkinkan personalisasi yang lebih efektif.
- **Analisis data waktu nyata.** Ini melibatkan pemrosesan dan analisis data saat dihasilkan atau diterima. Ini memungkinkan bisnis untuk segera menanggapi tren yang muncul atau tindakan pelanggan, yang sangat penting untuk pemasaran waktu nyata.
- **Pemasaran berbasis pemicu.** Ini mengacu pada tindakan pemasaran yang dipicu oleh perilaku atau peristiwa pelanggan tertentu. Ini merupakan komponen penting dari pemasaran waktu nyata.
- **Pemasaran omnichannel.** Ini adalah praktik mengintegrasikan dan mengoordinasikan upaya pemasaran di berbagai saluran. Ini sangat penting untuk memastikan pengalaman pelanggan yang konsisten dan personal.

- **Pengujian A/B.** Ini adalah metode untuk membandingkan dua versi halaman web, iklan, atau materi pemasaran lainnya untuk melihat mana yang berkinerja lebih baik. Hal ini penting untuk menyempurnakan dan mengoptimalkan personalisasi dan strategi pemasaran waktu nyata.

#### **Personalisasi dan Pemasaran Real-Time untuk Peritel Online**

Pertimbangkan peritel online besar, mirip dengan Amazon, yang menggunakan strategi personalisasi dan pemasaran real-time untuk meningkatkan pengalaman pelanggan dan meningkatkan penjualan dengan meningkatkan pengalaman pengguna dari segmen audiens yang lebih muda (Kampanye A) (lihat Gambar 11.6):

- **Personalisasi.** Peritel menggunakan berbagai data, termasuk pembelian sebelumnya, riwayat penelusuran, dan demografi pelanggan, untuk menyesuaikan pengalaman berbelanja untuk Kampanye A-nya. Misalnya, ketika pelanggan masuk ke akun mereka, mereka disambut dengan nama dan disajikan dengan halaman beranda yang berisi rekomendasi produk yang sesuai dengan selera dan preferensi mereka. Personalisasi ini meluas ke komunikasi pemasaran, di mana pelanggan menerima email atau pemberitahuan push yang menampilkan produk yang mungkin mereka minati berdasarkan perilaku mereka sebelumnya.
- **Pemasaran real-time.** Peritel memanfaatkan data waktu nyata untuk bereaksi secara instan terhadap perilaku pelanggan dalam Kampanye A. Misalnya, jika pelanggan menambahkan barang ke keranjang belanja tetapi tidak menyelesaikan pembelian, peritel dapat mengirimkan pengingat melalui email atau menawarkan diskon terbatas waktu untuk mendorong pembelian. Selama acara penting atau hari libur, peritel dapat menyesuaikan halaman beranda dan promosi mereka secara waktu nyata untuk mencerminkan kesempatan tersebut dan memanfaatkan peningkatan minat pelanggan.
- **Implementasi.** Untuk mendukung strategi ini, peritel menggunakan platform analitik data canggih dan algoritma pembelajaran mesin. Sistem ini memproses sejumlah besar data secara real-time, memungkinkan personalisasi instan dan respons pemasaran. Pengujian A/B terus digunakan untuk menyempurnakan strategi, memastikan personalisasi yang paling efektif dan taktik pemasaran real-time diterapkan.
- **Hasil.** Sebagai hasil dari strategi ini, peritel mengalami peningkatan keterlibatan pelanggan, interaksi, tingkat konversi yang lebih tinggi, dan peningkatan loyalitas pelanggan dari pelanggan di Kampanye A dibandingkan dengan demografi yang lebih luas (Kampanye B).



**Gambar 11.6** Interaksi Audiens Muda (Kampanye A) Dibandingkan dengan Demografi yang Lebih Luas (Kampanye B).

## 11.6 KESIMPULAN

Integrasi big data dan analitik real-time merupakan salah satu kemajuan paling transformatif di era digital modern. Bab ini membahas secara mendalam pemahaman tentang teknologi, alat, dan metode dasar yang mendukung integrasi ini, mengungkap potensi dan tantangan yang luas di bidang ilmu data pemasaran.

Salah satu kesimpulan utama dari eksplorasi ini adalah besarnya dan kompleksitas big data. Seperti namanya, big data bukan hanya tentang volume; ia mencakup beragam sumber data, struktur, dan kecepatan. Keberadaannya yang meluas di dunia kita yang saling terhubung menggarisbawahi esensi bisnis modern: setiap interaksi, setiap transaksi, dan setiap titik kontak adalah peluang untuk mendapatkan wawasan berbasis data. Teknologi seperti Hadoop dan Spark telah muncul sebagai landasan dalam mengelola dan memproses banjir data ini, memungkinkan bisnis untuk memahami hal yang tampaknya tak teratasi.

Namun, bukan hanya akumulasi data yang transformatif; kemampuan untuk menganalisis data ini secara real-time-lah yang merevolusi industri. Analisis real-time, yang difasilitasi oleh alat-alat seperti Kafka dan Storm, memberdayakan bisnis untuk beralih dari sikap reaktif ke sikap proaktif. Dalam dunia pemasaran, ini berarti berinteraksi dengan konsumen pada saat yang tepat, dengan pesan yang tepat, dalam konteks yang tepat. Kecepatan seperti itu dulunya merupakan kemewahan; Saat ini, hal itu merupakan kebutuhan untuk tetap kompetitif dan relevan.

Namun, kemajuan ini membawa tantangan yang signifikan. Skala big data menuntut tata kelola data yang ketat, memastikan akurasi, keamanan, dan penggunaan yang etis. Selain itu, analitik waktu nyata, meskipun ampuh, membutuhkan infrastruktur yang kuat dan tenaga kerja terampil yang mahir dalam aspek teknis dan bisnis. Kita juga harus mengakui sifat dinamis bidang ini. Teknologi dan alat yang telah kita bedah dalam bab ini terus berkembang, didorong

oleh laju inovasi yang tiada henti. Sebagai pemasar, ilmuwan data, dan pemimpin bisnis, ada kebutuhan terus-menerus untuk belajar, beradaptasi, dan melakukan iterasi.

Singkatnya, teknologi big data dan analitik waktu nyata lebih dari sekadar kata kunci atau tren yang cepat berlalu; keduanya mewakili titik temu strategi bisnis modern dan kehebatan teknologi. Menerapkannya bukanlah pilihan, tetapi suatu keharusan bagi setiap perusahaan yang bertujuan untuk pertumbuhan berkelanjutan dan berpusat pada pelanggan di era yang digerakkan oleh data ini. Masa depan menghadirkan lanskap di mana alat-alat ini bukan sekadar fasilitator, tetapi komponen integral dari strategi bisnis, yang mengarahkan jalannya upaya pemasaran di seluruh dunia.

## **BAB 12**

# **KECERDASAN BUATAN GENERATIF DAN APLIKASINYA DALAM PEMASARAN**

### **12.1 PENDAHULUAN**

Kecerdasan buatan (AI) adalah bidang luas yang mencakup berbagai subbidang, seperti pembelajaran mesin, visi komputer, pemrosesan bahasa alami, dan banyak lagi. Di antara subbidang ini, AI generatif adalah salah satu bidang pembahasan dan pengembangan yang paling menarik dan inovatif. AI generatif mengacu pada algoritma yang dapat menghasilkan sampel data baru berdasarkan kumpulan data yang diberikan, seperti gambar, teks, audio, atau video. Algoritma ini dapat belajar dari data yang ada dan menciptakan konten baru dan realistis yang meniru distribusi data asli.

AI generatif memiliki banyak potensi aplikasi di berbagai domain dan industri, seperti seni, hiburan, pendidikan, perawatan kesehatan, dan banyak lagi. Salah satu industri yang dapat memperoleh manfaat besar dari AI generatif adalah pemasaran. Seperti yang dibahas dalam teks ini, pemasaran adalah proses menciptakan, mengkomunikasikan, dan memberikan nilai kepada pelanggan dan pemangku kepentingan.

Hal ini melibatkan pemahaman kebutuhan dan preferensi pelanggan, perancangan dan pengembangan produk dan layanan, pembuatan dan distribusi konten, serta pengukuran dan optimalisasi kinerja pemasaran. Algoritma ini telah berevolusi secara signifikan sejak terobosan awal dengan jaringan adversarial generatif (GAN) dan autoencoder variasional (VAE), bergerak menuju arsitektur yang lebih canggih yang dapat menghasilkan keluaran yang semakin kompleks dan beresolusi tinggi.

AI generatif dapat membantu pemasar dalam berbagai aspek pekerjaan mereka, seperti pembuatan konten, personalisasi, segmentasi, prediksi, dan optimasi. Dengan menggunakan AI generatif, pemasar tidak hanya dapat mengotomatiskan beberapa tugas yang membosankan dan berulang, tetapi juga meningkatkan kreativitas dan inovasi mereka. Selain itu, AI generatif dapat memungkinkan pemasar untuk menghasilkan konten yang lebih relevan dan menarik bagi pelanggan individu, serta untuk memodelkan dan mengantisipasi perilaku pelanggan dengan lebih akurat. Hal ini dapat meningkatkan kepuasan pelanggan, loyalitas, retensi, dan nilai seumur hidup.

Perkembangan terkini dalam model transformer, terutama GPT-4 dan pendahulunya, telah secara signifikan memengaruhi AI generatif, yang mengarah pada kemampuan revolusioner dalam pembuatan dan pemahaman teks. Kemajuan ini telah membuka jalan bagi AI generatif untuk memainkan peran transformatif dalam pemasaran, yang pada intinya, berupaya menciptakan, mengkomunikasikan, dan memberikan nilai kepada pelanggan. Namun, AI generatif juga menimbulkan beberapa tantangan dan masalah etika yang perlu ditangani oleh pemasar dan bisnis. Beberapa masalah ini meliputi kualitas data, privasi, persetujuan, kepemilikan, keaslian, akuntabilitas, keadilan, transparansi, dan dampak sosial. Misalnya, AI generatif dapat membuat konten palsu atau menyesatkan yang dapat merusak

reputasi atau kredibilitas merek atau seseorang. AI generatif juga dapat melanggar privasi atau persetujuan pelanggan atau pengguna dengan menggunakan data pribadi mereka tanpa sepengetahuan atau izin mereka. Selain itu, AI generatif dapat memperkenalkan bias atau diskriminasi dalam data atau algoritma yang dapat memengaruhi hasil atau keputusan aktivitas pemasaran. Pemasar dan bisnis harus menavigasi isu-isu otentisitas, kepemilikan, dan dampak sosial dengan cermat, mematuhi pedoman dan kerangka kerja yang muncul yang dikeluarkan oleh badan pengatur.

Oleh karena itu, sangat penting bagi pemasar dan bisnis untuk menyadari manfaat dan risiko AI generatif dan menggunakannya secara bertanggung jawab dan etis. Hal ini membutuhkan penerapan beberapa praktik terbaik dan pedoman yang dapat memastikan kualitas, keandalan, keamanan, keadilan, transparansi, dan akuntabilitas aplikasi AI generatif dalam pemasaran. Dalam bab ini, kita akan mengeksplorasi dasar-dasar dan prinsip-prinsip AI generatif, potensi aplikasinya dalam pemasaran dengan beberapa contoh dan studi kasus dari berbagai sektor dan wilayah, serta pertimbangan etis yang menyertai penerapannya. Seiring dengan terus berkembang dan meningkatnya teknologi secara pesat (seperti ChatGPT-4), penting bagi pemasar dan bisnis untuk selalu mengikuti perkembangan terbaru dan memahami bagaimana mereka dapat memanfaatkannya untuk meningkatkan hasil pemasaran sambil tetap mematuhi praktik yang bertanggung jawab.

## **12.2 MEMAHAMI KECERDASAN BUATAN GENERATIF**

### **Pengantar Model Generatif**

AI Generatif mencakup teknik pembelajaran mesin yang bertujuan untuk menghasilkan sampel data baru yang mencerminkan distribusi inheren data pelatihan. Tujuannya melampaui replikasi; yaitu untuk memahami seluk-beluk distribusi set pelatihan untuk menghasilkan titik data baru yang kaya varian. Berbeda dengan model diskriminatif, yang membedakan batas kelas, model generatif menginternalisasi distribusi kelas. Kemampuan mereka dalam menghasilkan titik data baru mendasari penamaan mereka. GAN, subkelas model generatif yang menonjol, mewujudkan arsitektur dualistik—generator yang membentuk instance data dan diskriminator yang membedakan antara instance yang dihasilkan dan instance otentik, memungkinkan generator untuk memperbaiki melalui umpan balik iteratif.

Model generatif telah berkembang pesat melampaui GAN. VAE telah mengalami penyempurnaan yang signifikan, memungkinkan proses generasi yang lebih stabil dan terkontrol. Lebih lanjut, model berbasis transformer telah membawa kebangkitan kembali dalam tugas generatif, terutama dalam generasi teks, dengan model seperti GPT-4 dari OpenAI yang menunjukkan fleksibilitas dan koherensi yang belum pernah terjadi sebelumnya di berbagai domain.

Dalam pemasaran, potensi AI generatif telah meningkat pesat. AI generatif telah merambah ke pembuatan konten visual hiper-realistis untuk periklanan, alat penulisan iklan berbasis AI yang menyusun teks pemasaran dengan berbagai gaya dan nada, serta model prediksi canggih untuk peramalan perilaku konsumen. Alat-alat ini sekarang menjadi dasar



dalam mensintesis persona pelanggan yang realistis, sehingga menyempurnakan strategi segmentasi, penargetan, dan pemosisian.

Tren terkini menunjukkan munculnya platform desain berbasis AI yang terintegrasi dengan model generatif, merevolusi desain produk, pengemasan, dan materi pemasaran visual. Perusahaan memanfaatkan platform ini untuk mengulang desain dalam waktu dan biaya yang jauh lebih singkat dibandingkan metode tradisional. Bidang pembelajaran penguatan yang berkembang pesat juga menemukan sinergi dengan model generatif, khususnya dalam penetapan harga dinamis dan manajemen inventaris, dengan beradaptasi terhadap perubahan kondisi pasar dan respons konsumen secara real-time.

Mengingat evolusi AI generatif yang pesat, pemasar harus tetap waspada terhadap implikasi etisnya. Seiring dengan semakin sulitnya membedakan konten yang dihasilkan AI dari konten yang dihasilkan manusia, permintaan akan transparansi dan keaslian pun meningkat. Kebutuhan akan kerangka kerja yang mengatur penggunaan AI generatif secara bertanggung jawab dalam pemasaran bukan hanya bijaksana tetapi juga diperlukan untuk mempertahankan kepercayaan konsumen dan kepatuhan terhadap peraturan.

#### **Jaringan Adversarial Generatif, Autoencoder Variasional, dan Lainnya**

AI generatif mencakup serangkaian teknologi AI yang mensintesis data yang secara statistik mirip dengan set pelatihan. Yang paling dominan di antaranya adalah jaringan adversarial generatif (GAN) dan autoencoder variasional (VAE), tetapi kemajuan terbaru telah memperluas spektrumnya:

- **GAN.** GAN, yang diperkenalkan oleh Goodfellow dan rekan-rekan. pada tahun 2014, terus berkembang, dengan arsitektur yang lebih baru seperti GAN progresif dan StyleGAN yang meningkatkan kualitas dan resolusi gambar yang dihasilkan, serta meluas ke domain seperti pembuatan video dan objek tiga dimensi. Inovasi-inovasi ini semakin menyempurnakan penerapan GAN dalam pemasaran untuk menciptakan maket produk hiper-realistik dan konten visual untuk periklanan.
- **VAE.** Meskipun VAE mahir dalam menghasilkan distribusi data yang kompleks, perkembangan terkini telah berfokus pada peningkatan stabilitas dan kualitas sampelnya. Pengenalan VAE hierarkis dan penggunaan mekanisme inferensi posterior yang lebih kuat telah membuat mereka semakin tangguh dalam menghasilkan keluaran yang koheren dan beragam.
- **Model berbasis Transformer.** Arsitektur Transformer, khususnya dalam tugas pembangkitan bahasa, telah sangat memengaruhi AI generatif dengan kemampuannya untuk menghasilkan teks yang relevan secara kontekstual dan koheren secara sintaksis. Aplikasi pemasaran telah sangat diuntungkan dari hal ini untuk membuat deskripsi produk, chatbot, dan penulisan iklan AI.
- **Aplikasi AI generatif dalam pemasaran.** AI generatif dapat digunakan dalam berbagai aplikasi pemasaran, termasuk pembuatan konten, personalisasi, analitik prediktif, dan pemodelan perilaku pelanggan. Misalnya, ia dapat menghasilkan email yang dipersonalisasi, unggahan media sosial, atau iklan berdasarkan pola yang telah dipelajarinya dari kampanye pemasaran sebelumnya. Di luar pembuatan dan

personalisasi konten, peran AI generatif dalam pemasaran kini meluas ke pembuatan dataset sintesis untuk riset pasar, simulasi skenario ekonomi dan pasar, serta menghasilkan pengalaman pelanggan interaktif melalui aplikasi realitas virtual dan realitas tertambah.

- **Evaluasi model AI generatif.** Kinerja model AI generatif biasanya dievaluasi menggunakan metrik seperti skor inception (IS) dan jarak inception fr chet (FID). Metrik ini menilai keragaman dan kualitas sampel yang dihasilkan dibandingkan dengan data asli.
- **Evaluasi model AI generatif.** Meskipun IS dan FID, yang merupakan metrik yang dapat menilai keragaman dan kualitas sampel yang dihasilkan dibandingkan dengan data asli, tetap menjadi metrik standar, terdapat preferensi yang berkembang untuk metrik yang lebih bernuansa dan spesifik aplikasi yang dapat mengevaluasi kegunaan sampel yang dihasilkan dengan lebih baik dalam pengaturan realistis, seperti metrik presisi dan recall, dan studi pengguna untuk penilaian subjektif.

#### **Kekuatan dan Keterbatasan Kecerdasan Buatan Generatif**

Model generatif dalam AI, seperti GAN dan VAE, telah menjadi benteng inovasi, menawarkan banyak aplikasi kreatif di berbagai sektor, termasuk bidang pemasaran yang dinamis. Model-model ini unggul dalam mensintesis data dengan fidelitas tinggi, yang telah terbukti sangat berharga dalam pembuatan seni digital, musik, dan terutama dalam pembuatan konten pemasaran yang realistis. Kemampuan mereka untuk menambah data memperkaya kumpulan data yang tersedia untuk melatih model pembelajaran mesin lainnya, sebuah keuntungan signifikan dalam skenario di mana kelangkaan data menjadi hambatan. Pembuatan konten secara real-time adalah keunggulan lain dari model-model ini, memungkinkan pemasar untuk menyesuaikan strategi periklanan dan konten dinamis yang merespons secara instan terhadap perubahan lanskap preferensi konsumen. Evolusi kemampuan personalisasi melalui AI generatif telah transformatif, mengarah ke era baru pemasaran yang ditargetkan yang melayani pengalaman konsumen individual dalam skala besar.

Namun, penerapan model generatif bukannya tanpa tantangan. Kebutuhan akan sumber daya komputasi yang ekstensif dapat menimbulkan hambatan yang signifikan, terutama bagi organisasi dengan kapasitas terbatas, yang berpotensi menyebabkan kesenjangan digital dalam pemanfaatan teknologi pemasaran. Ketergantungan pada kualitas data pelatihan adalah aspek penting lainnya, di mana bias atau ketidakcukupan dalam dataset dapat menyebabkan output yang kurang realistis atau bahkan bermasalah secara etis.

Menilai kualitas konten yang dihasilkan tetap menjadi masalah yang kompleks karena metrik kinerja konvensional gagal menangkap aspek-aspek nuansa kualitas dan keragaman dalam model generatif, sehingga memerlukan kerangka kerja evaluatif baru. Potensi penyalahgunaan AI generatif, seperti dalam pembuatan deepfake atau penyebaran informasi yang salah, menghadirkan kekhawatiran etis yang signifikan, menimbulkan pertanyaan tentang tata kelola dan pengawasan teknologi yang ampuh ini.

**Tabel 12.1** Kekuatan dan Keterbatasan Kecerdasan Buatan Generatif.

| Keunggulan                                             | Keterbatasan                                                    |
|--------------------------------------------------------|-----------------------------------------------------------------|
| Kemampuan adaptasi tinggi untuk aplikasi kreatif       | Kebutuhan komputasi dan data yang tinggi                        |
| Kapasitas untuk peningkatan data yang signifikan       | Hasil yang mungkin tidak realistis tanpa pelatihan yang memadai |
| Kemampuan pembuatan konten secara real-time            | Metrik evaluasi kualitas yang kompleks                          |
| Personalisasi yang ditingkatkan untuk konten pemasaran | Kekhawatiran etis termasuk potensi penyalahgunaan               |

### Sekilas tentang Pembahasan Mutakhir dalam Kecerdasan Buatan Generatif

Bidang AI generatif terus berkembang dengan kecepatan yang belum pernah terjadi sebelumnya, dengan kemajuan pembahasan yang secara konsisten membuka jalan baru. Pembaruan dalam lanskap pembahasan yang dinamis ini seringkali berfokus pada peningkatan kecanggihan, efisiensi, dan aksesibilitas model generatif. Berikut adalah beberapa perkembangan yang dapat diintegrasikan ke dalam bagian ini untuk mencerminkan keadaan terkini di bidang ini:

- **Perkembangan StyleGAN tingkat lanjut.** Perkembangan dari StyleGAN ke StyleGAN2 dan StyleGAN3 oleh Karras dan rekan-rekan. telah memberikan peningkatan kualitas gambar, menghilangkan artefak, dan memperkenalkan konsistensi temporal untuk pembuatan video. Kemajuan ini telah mendorong batas-batas apa yang dapat dianggap berbeda secara jelas dari foto atau video sebenarnya.
- **Model difusi.** Kemajuan penting lainnya dalam AI generatif adalah munculnya model difusi, seperti model probabilistik difusi denoising. Model-model ini, yang secara bertahap mengubah noise menjadi data terstruktur, telah diterapkan untuk menghasilkan gambar dengan fidelitas yang sebanding dengan yang dihasilkan oleh GAN, seringkali dengan stabilitas pelatihan dan keragaman sampel yang lebih baik.
- **AI generatif multimodal.** Pembahasan juga telah meluas ke model generatif multimodal, yang dapat memahami dan menghasilkan konten di berbagai modalitas, seperti DALL-E oleh OpenAI untuk generasi teks-ke-gambar dan RAG (retrieval-augmented generation) untuk pemahaman dan generasi bahasa alami yang ditingkatkan dengan menggabungkan pengambilan dokumen ke dalam proses generatif.
- **Peningkatan skalabilitas dan efisiensi.** Ada dorongan terus-menerus untuk membuat model generatif lebih skalabel dan efisien. Inovasi seperti transformator sparse dan teknik distilasi memungkinkan pelatihan model skala besar dengan miliaran parameter secara lebih mudah, meningkatkan kinerja dan mengurangi kebutuhan komputasi.
- **Kemajuan AI pada perangkat dan di edge computing.** Di luar upaya Qualcomm, terdapat kemajuan signifikan di bidang AI pada perangkat dan di edge computing, yang memungkinkan model generatif yang canggih untuk dijalankan pada perangkat keras kelas konsumen. Ini termasuk kemajuan dalam kuantisasi model dan pemangkasan

jaringan, yang memungkinkan model kompleks dikompresi menjadi bentuk yang lebih efisien tanpa kehilangan kinerja yang substansial.

- **Aplikasi dalam pemasaran.** Kegunaan AI generatif dalam pemasaran terus meluas, dengan AI tidak hanya mempersonalisasi konten tetapi juga mengoptimalkan saluran pemasaran, meningkatkan simulasi perjalanan pelanggan, dan membantu dalam strategi penetapan harga dinamis. Konten yang dihasilkan AI semakin tidak dapat dibedakan dari konten yang dibuat manusia, memungkinkan interaksi pelanggan yang lebih otomatis dan personal.

Singkatnya, AI generatif bukan hanya alat untuk membuat gambar atau teks realistis tetapi dengan cepat menjadi aset multifungsi yang dapat beroperasi di berbagai media, meningkatkan efisiensi, dan bahkan berjalan di perangkat konsumen. Model-model ini berperan penting dalam menciptakan pengalaman imersif yang memiliki dampak yang semakin besar pada pemasaran, desain, dan industri kreatif.

#### **Kecerdasan Buatan Generatif dalam Pembuatan dan Personalisasi Konten**

Pertimbangkan sebuah perusahaan bernama RetailX, pengecer online besar yang memiliki berbagai macam produk di berbagai kategori. Mereka menghadapi tugas yang menantang untuk membuat dan mempersonalisasi konten pemasaran untuk jutaan pelanggan unik, dengan beragam preferensi dan perilaku belanja. Untuk mengatasi tantangan ini, RetailX memutuskan untuk menerapkan teknik AI generatif.

Mereka pertama-tama mengumpulkan sejumlah besar data tentang pelanggan mereka, termasuk riwayat pembelian sebelumnya, perilaku penelusuran, dan detail pribadi yang diberikan oleh pelanggan itu sendiri. Data ini membentuk dasar untuk model AI generatif. Tim memutuskan untuk menggunakan model GAN, yang telah menunjukkan hasil yang menjanjikan dalam menciptakan konten yang dipersonalisasi. Mereka melatih GAN pada data pelanggan yang dikumpulkan. Bagian generator dari GAN menggunakan data pelatihan ini untuk membuat profil pelanggan hipotetis baru dan perilaku belanja terkaitnya. Bagian diskriminator kemudian mengevaluasi profil yang dihasilkan ini untuk keasliannya, membantu generator meningkatkan outputnya dari waktu ke waktu.

Setelah melatih GAN, RetailX menggunakannya untuk menghasilkan konten pemasaran yang dipersonalisasi untuk pelanggan mereka. Misalnya, mereka menghasilkan deskripsi produk dan rekomendasi unik untuk setiap pelanggan berdasarkan profil hipotetis mereka. Mereka juga menggunakan GAN untuk membuat kampanye email yang dipersonalisasi, di mana baris subjek, konten, dan rekomendasi produk disesuaikan dengan setiap pelanggan individu.

Seiring waktu, mereka mengamati peningkatan signifikan dalam keterlibatan pelanggan dan tingkat konversi. Konten pemasaran yang dipersonalisasi yang dibuat oleh model AI generatif beresonansi dengan pelanggan, yang mengarah pada tingkat klik-tayang yang lebih tinggi dan pada akhirnya penjualan yang lebih tinggi.

Dengan cara ini, RetailX memanfaatkan kekuatan AI generatif untuk mengotomatisasi dan mengoptimalkan proses pembuatan dan personalisasi konten mereka, mencapai hasil

pemasaran yang lebih baik dan pengalaman belanja yang lebih personal bagi pelanggan mereka.

### **12.3 GENERATIVE AI UNTUK PERSONALISASI KONTEN**

#### **Kecerdasan Buatan Generatif untuk Pembuatan Konten: Teks, Gambar, dan Video**

AI generatif adalah bidang pembelajaran mesin yang menarik dan berkembang pesat. AI generatif menggunakan algoritma untuk menghasilkan konten seperti teks, gambar, dan musik berdasarkan pola yang dipelajari dari data. Penggunaan AI yang inovatif ini telah mulai merevolusi banyak bidang, termasuk pembuatan dan personalisasi konten dalam pemasaran.

Dalam pemasaran, konten adalah raja. Membuat konten yang dipersonalisasi yang menarik bagi berbagai segmen pelanggan dapat menjadi tugas yang melelahkan dan kompleks. Tetapi dengan AI generatif, kita dapat mengotomatiskan proses ini dan menghasilkan konten yang dipersonalisasi dalam skala besar, sehingga meningkatkan upaya pemasaran.

AI generatif telah digunakan untuk menulis deskripsi produk, konten email, unggahan media sosial, dan bahkan artikel. Selain itu, AI generatif dapat membuat rekomendasi produk yang dipersonalisasi untuk pengguna individu berdasarkan preferensi dan perilaku masa lalu mereka. Hal ini meningkatkan keterlibatan dan rasio klik-tayang sekaligus meningkatkan pengalaman pelanggan secara keseluruhan. Misalnya, alat bantu penulisan berbasis AI dapat membantu pemasar menyusun konten yang lebih efektif dengan menyarankan cara untuk meningkatkan kejelasan, keringkasan, dan nada. Alat ini juga dapat menghasilkan salinan yang unik dan kreatif untuk kampanye iklan, yang sesuai dengan berbagai persona pelanggan.

Meskipun teknologi ini memiliki potensi yang sangat besar, ada tantangan yang perlu diatasi. Kualitas konten yang dihasilkan sangat bergantung pada kualitas dan jumlah data yang dimasukkan ke dalam sistem AI. Sistem ini terkadang dapat menghasilkan konten yang tidak pantas, dan ada kekhawatiran etis seputar penggunaan AI dalam pembuatan konten. Namun, dengan pembahasan dan pengembangan lebih lanjut, AI generatif siap menjadi pengubah permainan di bidang pemasaran konten.

#### **Mempersonalisasi Pengalaman Pengguna dengan Model Generatif**

Dunia pemasaran berkembang pesat berkat personalisasi. AI generatif memperluas sentuhan pribadi ini, menawarkan pengalaman khusus kepada pengguna:

- **Konten yang sangat personal.** Model generatif dapat menghasilkan konten yang disesuaikan dengan preferensi pengguna individu, mulai dari email yang dipersonalisasi hingga saran produk yang dikurasi, meningkatkan pengalaman pengguna dan berpotensi meningkatkan konversi.
- **Interaksi dan umpan balik pengguna.** Beberapa sistem canggih dapat beradaptasi secara real-time berdasarkan interaksi pengguna, memastikan konten tetap relevan dan menarik, sehingga semakin meningkatkan tingkat keterlibatan.
- **Pertimbangan etis.** Seiring dengan semakin dalamnya personalisasi, garis tipis antara pengalaman yang dikurasi dan pemasaran yang invasif menjadi kabur. Sangat penting

untuk mendekati personalisasi berbasis AI dengan menghormati norma privasi dan kenyamanan pengguna.

### **Kampanye Pemasaran Kecerdasan Buatan Generatif yang Sukses**

Beberapa merek telah menggunakan AI generatif untuk kampanye perintis:

- **Desain fesyen khusus.** Merek seperti StitchFix menggunakan model generatif untuk memprediksi tren fesyen, menghasilkan desain yang sesuai dengan preferensi pengguna saat ini.
- **Kreasi musik.** Perusahaan seperti Amper Music menggunakan AI untuk menghasilkan trek musik unik untuk kampanye iklan, memastikan mereka selalu memiliki soundtrack yang segar.

### **Alat dan Platform untuk Pembuatan Konten Generatif**

Bagi bisnis yang tertarik untuk mengeksplorasi AI generatif, berbagai alat dan platform menyederhanakan perjalanan:

- **Seri GPT OpenAI.** Rangkaian ini, terutama versi terbaru, memfasilitasi pembuatan teks berkualitas tinggi, cocok untuk berbagai kebutuhan konten.
- **Runway ML.** Ini adalah platform intuitif yang menghadirkan kemampuan model generatif untuk pembuatan konten visual, dari gambar hingga video.
- **DeepArt.io.** Dengan memanfaatkan transfer gaya neural, alat ini memungkinkan merek untuk membuat konten visual unik yang terinspirasi oleh gaya seni ikonik, menambahkan sentuhan kelas pada kampanye.

### **Konsep Kunci**

Saat mempertimbangkan implementasi AI generatif dalam pembuatan dan personalisasi konten, beberapa konsep kunci perlu dipahami:

- **Model generatif.** Model generatif, seperti GAN atau VAE, adalah jenis algoritma AI yang menghasilkan instance data baru yang mirip dengan data input. Model ini digunakan untuk menghasilkan konten yang beragam dan kreatif, mulai dari gambar sintetis hingga teks.
- **Personalisasi.** Dalam konteks pembuatan konten, personalisasi mengacu pada penyesuaian konten agar sesuai dengan kebutuhan dan preferensi individu setiap pengguna. Proses ini meningkatkan keterlibatan dan kepuasan pengguna. Personalisasi sering dilakukan berdasarkan profil pengguna, perilaku, dan umpan balik, dan dengan AI generatif, dimungkinkan untuk menghasilkan konten yang sangat personal dalam skala besar.
- **Kreativitas AI.** Konsep ini mengacu pada kemampuan sistem AI untuk menghasilkan konten yang tidak hanya baru tetapi juga kreatif dan inovatif. Meskipun ini adalah bidang yang kompleks dan subjektif, kemajuan terbaru dalam AI generatif telah menunjukkan hasil yang menjanjikan dalam menghasilkan musik, seni, dan teks yang dibuat oleh AI yang dianggap kreatif oleh manusia.
- **Privasi data.** Dalam konteks personalisasi, privasi data merupakan perhatian utama. Regulasi perlindungan data seperti GDPR memberlakukan aturan ketat tentang

bagaimana data pribadi dapat digunakan, dan penting untuk memastikan bahwa aplikasi AI generatif menghormati regulasi ini.

### **Contoh Praktis**

Pertimbangkan sebuah merek ritel yang ingin meningkatkan strategi pembuatan konten dan personalisasi. Mari kita sebut merek tersebut RetailXYZ. Untuk mengikuti perkembangan industri ritel yang serba cepat, RetailXYZ membutuhkan cara untuk membuat sejumlah besar konten yang dipersonalisasi dengan cepat dan efisien. Mereka beralih ke AI generatif untuk membantu mencapai tujuan ini.

Pertama, RetailXYZ menggunakan model AI generatif untuk membuat deskripsi produk yang unik. Dengan memasukkan data tentang kategori produk, fitur, dan deskripsi produk yang sudah ada ke dalam model, AI mampu menghasilkan ribuan deskripsi unik dalam waktu yang jauh lebih singkat daripada yang dibutuhkan manusia. Hal ini membantu menyederhanakan proses dan meningkatkan konsistensi serta kualitas deskripsi di seluruh katalog mereka.

Selanjutnya, RetailXYZ menggunakan AI generatif untuk membuat kampanye pemasaran email yang dipersonalisasi. Model tersebut dilatih berdasarkan kampanye email sebelumnya dan data pelanggan untuk menghasilkan konten email yang disesuaikan dengan preferensi dan perilaku setiap pelanggan. Hal ini menghasilkan tingkat pembukaan dan klik yang lebih tinggi karena konten tersebut sangat relevan bagi setiap penerima.

Akhirnya, RetailXYZ memanfaatkan kekuatan AI generatif untuk pemasaran media sosial. Mereka menggunakan model tersebut untuk menghasilkan postingan yang kreatif dan menarik bagi berbagai persona pelanggan berdasarkan minat dan interaksi mereka dengan merek di media sosial.

Terlepas dari tantangan awal dalam menyempurnakan model AI agar selaras dengan suara merek dan memastikan model tersebut menciptakan konten yang tepat, RetailXYZ menemukan kesuksesan yang signifikan dengan AI generatif. Hal ini tidak hanya meningkatkan efisiensi tetapi juga memungkinkan tingkat personalisasi yang lebih tinggi dalam upaya pemasaran mereka. Studi kasus ini menggambarkan kekuatan AI generatif dalam pemasaran. Dengan mengotomatiskan pembuatan konten yang dipersonalisasi, perusahaan seperti RetailXYZ dapat lebih efektif melibatkan pelanggan mereka dan meningkatkan strategi pemasaran mereka.

## **12.4 PREDIKSI PERILAKU PELANGGAN LEWAT GENERATIVE AI**

### **Gambaran Umum**

Analitik prediktif dan pemodelan perilaku pelanggan telah lama menjadi komponen penting dari strategi pemasaran, memungkinkan pemasar untuk mengantisipasi tren masa depan, memahami perilaku pelanggan, dan membuat keputusan yang tepat. Dengan munculnya AI generatif, bidang-bidang ini mengalami pergeseran paradigma.

Model AI generatif dapat menghasilkan instance data baru yang menyerupai data pelatihan. Hal ini memiliki implikasi mendalam bagi analitik prediktif dan pemodelan perilaku pelanggan. Pada intinya, AI generatif dapat menciptakan kumpulan data sintesis yang mencerminkan skenario dunia nyata, memungkinkan pemasar untuk mensimulasikan

berbagai strategi pemasaran dan mengukur respons pelanggan tanpa harus menerapkannya di dunia nyata.

Misalnya, AI generatif dapat mensimulasikan reaksi pelanggan terhadap peluncuran produk baru atau perubahan harga, memberikan wawasan yang sangat berharga sebelum implementasi sebenarnya. Hal ini dapat membantu pemasar menyempurnakan strategi mereka, mengantisipasi potensi jebakan, dan mengoptimalkan kepuasan pelanggan dan pertumbuhan pendapatan secara maksimal.

Selain itu, AI generatif juga dapat membantu dalam memahami dan memvisualisasikan perilaku pelanggan yang kompleks. Misalnya, model generatif seperti GAN dapat mempelajari distribusi perilaku pelanggan dan menghasilkan instance baru yang membantu dalam memahami pola dan tren yang mendasarinya.

Yang perlu diperhatikan, AI generatif adalah alat yang ampuh di era big data, di mana teknik analitik prediktif tradisional mungkin gagal karena volume dan kompleksitas data yang sangat besar. Dengan kemampuannya untuk bekerja dengan kumpulan data yang besar dan kompleks, AI generatif menyediakan cara yang efektif untuk memanfaatkan big data untuk analitik prediktif dan pemodelan perilaku pelanggan.

### **Meningkatkan Model Prediktif Menggunakan Teknik Generatif**

Teknik AI generatif, terutama yang berlandaskan pembelajaran mendalam, telah membuka dimensi baru dalam pemodelan prediktif. Alih-alih hanya memanfaatkan kumpulan data yang ada untuk membuat prediksi, teknik generatif memungkinkan penambahan dan pengayaan kumpulan data ini, memfasilitasi pelatihan model yang lebih baik dan prediksi yang lebih bernuansa. Bagian ini mengeksplorasi bagaimana teknik generatif dapat meningkatkan model prediktif tradisional, yang mengarah pada pengambilan keputusan yang lebih baik di bidang pemasaran.

### **Meningkatkan Data untuk Pelatihan yang Lebih Baik**

Salah satu tantangan mendasar dalam pemodelan prediktif adalah kelangkaan data. Misalnya, dalam analisis perilaku pelanggan, perilaku tertentu mungkin kurang terwakili karena kelangkaannya. Model generatif, khususnya GAN, dapat digunakan untuk menghasilkan sampel data sintetis, memastikan bahwa model mendapatkan pengalaman pelatihan holistik.

### **Simulasi Skenario**

Model generatif, berkat kapasitas penghasil datanya, dapat mensimulasikan berbagai skenario. Seorang pemasar yang ingin mengetahui dampak potensial dari strategi penetapan harga baru dapat menggunakan model ini untuk mensimulasikan respons pelanggan, sehingga mendapatkan wawasan sebelum implementasi sebenarnya.

### **Imputasi dan Penghilangan Derau Data**

Dalam banyak dataset, nilai yang hilang atau data yang bising dapat mengganggu daya prediksi model. Teknik generatif, khususnya VAE, telah menunjukkan potensi dalam mengisi nilai yang hilang berdasarkan distribusi data yang dipelajari, sehingga meningkatkan kualitas dataset dan, selanjutnya, hasil prediksi.



### **Generasi Fitur**

Teknik generatif juga dapat berperan dalam menciptakan fitur baru yang dapat meningkatkan kemampuan prediksi model. Dengan mempelajari pola kompleks dalam data, model generatif dapat mengidentifikasi dan menghasilkan fitur yang mungkin tidak jelas tetapi signifikan untuk tugas prediksi.

### **Ketahanan Model**

Model prediktif terkadang rentan terhadap gangguan kecil pada data input, yang menyebabkan prediksi yang tidak menentu. Model generatif dapat digunakan untuk menghasilkan versi data pelatihan yang terganggu atau ditambah, sehingga melatih model prediktif agar lebih kuat dan konsisten dalam prediksinya.

### **Memahami Interaksi Kompleks**

Banyak perilaku pelanggan dihasilkan dari interaksi rumit antara berbagai variabel. Model prediktif tradisional mungkin kesulitan untuk merangkum interaksi ini, tetapi model generatif, dengan kemampuannya untuk memahami distribusi data, dapat menangkap dan memodelkan interaksi kompleks ini, yang mengarah pada prediksi yang lebih bernuansa.

Kesimpulannya, integrasi teknik generatif dengan pemodelan prediktif menandai era baru dalam analitik prediktif. Dengan memungkinkan augmentasi data, simulasi skenario, dan pemahaman mendalam tentang pola yang rumit, teknik generatif siap untuk secara signifikan meningkatkan akurasi dan penerapan model prediktif dalam konteks pemasaran.

### **Mensimulasikan Perjalanan dan Perilaku Pelanggan**

Penggabungan AI generatif ke dalam simulasi perjalanan dan perilaku pelanggan membuka ranah kemungkinan baru. Dengan menciptakan perjalanan pelanggan hipotetis tetapi realistis, bisnis dapat memperoleh pemahaman mendalam tentang jalur potensial yang mungkin diambil pelanggan dan membuat keputusan proaktif.

### **Pengujian A/B Virtual**

Model generatif dapat mensimulasikan hasil dari berbagai strategi pemasaran pada profil pelanggan sintetis. Alih-alih menerapkan strategi secara real-time dan menunggu untuk mengumpulkan hasil, bisnis dapat secara virtual menguji beberapa kampanye pada profil yang dihasilkan ini, memungkinkan mereka untuk memilih kampanye yang paling efektif sebelumnya.

### **Persona Pelanggan Dinamis**

Persona pelanggan tradisional bersifat statis. Dengan AI generatif, pemasar dapat memiliki persona dinamis yang berkembang seiring waktu, memberikan wawasan real-time tentang perubahan preferensi pelanggan. Misalnya, persona dinamis dapat mencerminkan bagaimana preferensi pelanggan berubah setelah peristiwa penting dalam hidup, seperti pernikahan atau kelahiran anak.

### **Perencanaan Skenario**

Perusahaan dapat menggunakan AI generatif untuk mensimulasikan berbagai skenario, seperti peluncuran produk baru atau perubahan dinamika pasar, untuk mengukur potensi reaksi pelanggan. Hal ini membantu dalam mitigasi risiko, memastikan bisnis siap menghadapi berbagai hasil.

### Potensi Jebakan dan Penyalahgunaan Prediksi Generatif

Meskipun AI generatif menawarkan banyak manfaat, ia tidak terlepas dari potensi jebakan (lihat Tabel 12.2):

- **Ketergantungan Berlebihan.** Risiko potensial terletak pada ketergantungan berlebihan pada simulasi dan prediksi yang dihasilkan AI tanpa mempertimbangkan ketidakpastian dunia nyata. Model generatif, bagaimanapun, mendasarkan prediksinya pada data historis, yang mungkin tidak selalu memperhitungkan perubahan mendadak dalam dinamika pasar.
- **Kekhawatiran etis.** Membuat profil pelanggan yang sangat realistis mungkin hampir melanggar privasi pelanggan. Terlebih lagi, jika data yang dihasilkan disalahpahami sebagai data nyata, hal itu dapat menyebabkan strategi pemasaran yang salah arah atau bahkan implikasi hukum.
- **Bias data.** Jika data pelatihan memiliki bias bawaan, model generatif dapat mereproduksi dan bahkan memperkuat bias tersebut, yang menyebabkan prediksi yang menyimpang dan berpotensi berbahaya.

**Tabel 12.2** Potensi Jebakan dan Penyalahgunaan Prediksi Generatif dengan Contoh dan Konsekuensinya.

| Kesalahan/Penyalahgunaan                            | Contoh                                                | Konsekuensi                                      |
|-----------------------------------------------------|-------------------------------------------------------|--------------------------------------------------|
| Menghasilkan konten yang menyesatkan                | Membuat ulasan produk palsu                           | Menyesatkan pelanggan<br>Legal actions           |
| Membuat deepfake untuk disinformasi                 | Membuat video dengan dukungan palsu                   | Menyebarkan informasi palsu<br>Merusak reputasi  |
| Terlalu bergantung pada prediksi yang dihasilkan AI | Mengabaikan penilaian manusia dalam keputusan penting | Potensi kerugian bisnis                          |
| Mengabaikan pertimbangan etika                      | Menggunakan AI generatif tanpa persetujuan pengguna   | Hilangnya kepercayaan pelanggan<br>Masalah hukum |

### Arah Masa Depan untuk Analitik Prediktif Generatif

Bidang analitik prediktif generatif sangat siap untuk inovasi, dan kemajuan di masa depan kemungkinan akan didasarkan pada peningkatan ini:

- **Integrasi dengan teknologi AI lainnya.** Menggabungkan model generatif dengan sistem AI lain seperti pembelajaran penguatan dapat membuka jalan bagi model pemasaran yang lebih adaptif dan responsif yang dapat bereaksi terhadap umpan balik waktu nyata.
- **Model generatif yang lebih etis.** Dengan meningkatnya kekhawatiran tentang privasi dan persetujuan, kita mungkin akan melihat dorongan menuju model yang memprioritaskan pertimbangan etis, yang mengarah pada AI yang lebih transparan dan bertanggung jawab.

- **Personalisasi granular.** Seiring model generatif menjadi lebih canggih, tingkat personalisasi yang ditawarkannya akan menjadi lebih bernuansa, berpotensi memungkinkan bisnis untuk menyesuaikan konten dengan suasana hati, momen, atau bahkan kondisi fisiologis individu.
- **Pemahaman pelanggan holistik.** Masa depan mungkin akan melihat model generatif yang tidak hanya memprediksi perilaku pembelian tetapi juga menawarkan wawasan tentang seluruh gaya hidup pelanggan, mulai dari kebiasaan makan hingga preferensi hiburan, memberikan pemahaman holistik tentang konsumen modern.

Kesimpulannya, AI generatif dalam pemasaran adalah domain yang berkembang yang menjanjikan perpaduan kemungkinan dan tantangan inovatif. Seiring lanskap terus bergeser, kebutuhan akan pertimbangan etis, metodologi yang kuat, dan kolaborasi manusia-mesin yang seimbang akan menjadi semakin penting.

#### **Konsep Kunci**

- **Model generatif dalam analitik prediktif.**  
Model generatif, berbeda dengan model diskriminatif, tidak hanya mempelajari batasan antara kelas data yang berbeda tetapi juga distribusi data yang sebenarnya. Hal ini menjadikannya ideal untuk tugas-tugas seperti deteksi anomali, di mana tujuannya adalah untuk mengidentifikasi titik data yang menyimpang dari norma. Dalam analitik prediktif, ini bisa berarti mengidentifikasi perilaku pelanggan yang tidak biasa atau memperkirakan penjualan di masa mendatang berdasarkan tren historis.
- **Pemodelan perilaku pelanggan.** Model generatif dapat digunakan untuk mensimulasikan perilaku pelanggan. Model tersebut mempelajari perilaku pelanggan dari data historis dan dapat menghasilkan data baru yang mengikuti pola yang sama. Ini dapat sangat berguna dalam pengujian skenario dan pengambilan keputusan. Misalnya, sebuah perusahaan dapat menggunakan model generatif untuk memprediksi dampak strategi pemasaran baru terhadap perilaku pelanggan.
- **Imputasi data.** Data yang hilang adalah masalah umum dalam analitik perilaku pelanggan. Model generatif seperti VAE dapat digunakan untuk menghasilkan data yang hilang berdasarkan pola yang dipelajari dari data yang tersedia. Hal ini dapat menghasilkan model prediktif yang lebih kuat dan wawasan yang lebih baik tentang perilaku pelanggan.
- **Pemodelan temporal.** Banyak perilaku pelanggan bergantung pada waktu, seperti pola pembelian. Model generatif dapat digunakan untuk memodelkan ketergantungan waktu ini, membantu perusahaan memprediksi perilaku pelanggan di masa depan dan mengoptimalkan strategi pemasaran mereka sesuai dengan itu.

#### **Generative AI untuk Analisis Prediktif Ritel**

Untuk menghidupkan konsep AI generatif dalam analisis prediktif dan pemodelan perilaku pelanggan, mari kita pertimbangkan contoh nyata dari sebuah perusahaan ritel multinasional. Perusahaan ritel ini, sebut saja RetailX, mengoperasikan toko fisik dan platform online. RetailX memiliki sejumlah besar data perilaku pelanggan dari program loyalitasnya, yang mencakup informasi seperti frekuensi pembelian, jumlah yang dibelanjakan, dan produk

yang dibeli. Perusahaan ingin lebih memahami pelanggannya dan memprediksi perilaku mereka di masa depan untuk mengoptimalkan strategi pemasaran.

RetailX menggunakan AI generatif, khususnya jenis model yang dikenal sebagai GAN, untuk mempelajari distribusi perilaku pelanggan. Setelah pelatihan pada data pembelian historis, GAN dapat menghasilkan profil pelanggan sintetis yang sesuai dengan pola perilaku pelanggan nyata. Data sintetis ini berfungsi sebagai kumpulan data yang diperluas untuk pemodelan, memungkinkan RetailX untuk mensimulasikan basis pelanggan yang jauh lebih besar dan memprediksi berbagai skenario perilaku di masa mendatang.

Selain itu, RetailX menggunakan jenis model AI generatif yang disebut VAE untuk mengisi data yang hilang dalam dataset perilaku pelanggannya. VAE mempelajari distribusi data yang diamati dan dapat menghasilkan nilai yang mungkin untuk titik data yang hilang, sehingga data perilaku pelanggan menjadi lebih lengkap dan andal untuk analisis lebih lanjut dan pemodelan prediktif. Penggunaan model AI generatif dalam analitik prediktif dan pemodelan perilaku pelanggan memungkinkan RetailX untuk lebih memahami pelanggannya, mengantisipasi kebutuhan mereka, dan menyesuaikan strategi pemasaran sesuai dengan kebutuhan tersebut.

## **12.5 ETIKA & MASA DEPAN GENERATIVE AI DALAM PEMASARAN**

### **Gambaran Umum**

Sekuat apa pun AI generatif dalam meningkatkan praktik pemasaran, hal ini tidak lepas dari pertimbangan dan tantangan etika. Ini melampaui kekhawatiran umum tentang privasi dan keamanan data, meskipun hal-hal ini masih sangat relevan, hingga mencakup isu-isu baru yang unik bagi kekuatan dan kemampuan AI generatif.

Salah satu pertimbangan etika utama adalah potensi penyalahgunaan AI generatif dalam menciptakan konten yang menyesatkan atau bahkan berbahaya. Deepfake, contoh penting dari teknologi AI generatif, telah digunakan untuk membuat berita palsu dan menyebarkan disinformasi, yang menimbulkan ancaman serius bagi individu dan masyarakat. Demikian pula, dalam pemasaran, terdapat potensi untuk menghasilkan iklan yang menyesatkan atau menipu, yang menyebabkan ketidakpercayaan konsumen dan kerusakan reputasi merek.

Kekhawatiran etis lainnya terletak pada ketidaktransparanan model AI generatif. Model-model ini, yang sangat kompleks, seringkali kurang transparan dan mudah dijelaskan, sehingga sulit untuk menafsirkan outputnya atau proses yang mengarah padanya. Ketidaktransparanan ini dapat menyebabkan bias yang tidak disengaja, ketidakadilan, dan diskriminasi dalam praktik pemasaran.

Namun, prospek masa depan AI generatif dalam pemasaran sangat menjanjikan. Diharapkan dapat mendorong pengalaman pemasaran yang lebih personal dan menarik, menyederhanakan proses pembuatan konten, dan memberikan wawasan yang lebih dalam tentang perilaku dan preferensi pelanggan.

Selain itu, kemajuan dalam etika dan regulasi AI, bersama dengan pembahasan berkelanjutan untuk membuat AI lebih transparan dan mudah diinterpretasikan, diharapkan dapat mengatasi banyak kekhawatiran etis yang terkait dengan AI generatif.

#### **Tantangan Etika: Deepfake, Disinformasi, dan Lainnya**

Evolusi pesat AI generatif telah membawa serta berbagai tantangan etika. Deepfake, yang menggunakan GAN untuk menciptakan konten video yang sangat realistis tetapi sepenuhnya fiktif, mungkin merupakan ancaman yang paling terkenal. Teknologi semacam itu, di tangan mereka yang berniat jahat, dapat salah menggambarkan peristiwa, merusak reputasi, atau bahkan memanipulasi opini publik dan proses politik.

Selain itu, disinformasi tidak terbatas pada deepfake. AI generatif memiliki kemampuan untuk menciptakan artikel berita palsu, gambar palsu, atau suara palsu. Penyebaran konten palsu tersebut dapat mengikis kepercayaan pada sumber berita yang bereputasi, sehingga menggoyahkan struktur masyarakat.

#### **Penggunaan AI Generatif yang Bertanggung Jawab dalam Pemasaran**

Para profesional pemasaran harus menavigasi lanskap baru ini dengan kompas etika yang tajam. Kekuatan AI generatif harus digunakan secara bertanggung jawab (lihat Tabel 12.3):

- **Transparansi dan pengungkapan.** Perusahaan harus mengungkapkan kapan mereka menggunakan AI generatif, terutama saat membuat konten promosi. Bersikap transparan tentang penggunaan teknologi ini tidak hanya menumbuhkan kepercayaan tetapi juga menjaga integritas merek.
- **Pengumpulan dan penggunaan data yang etis.** Pemasar harus memastikan bahwa setiap data yang digunakan untuk melatih model generatif diperoleh secara etis, dengan menghormati hak privasi dan peraturan perlindungan data.
- **Menghindari representasi yang menyesatkan.** AI generatif tidak boleh digunakan untuk memalsukan testimoni, ulasan, atau dukungan. Keaslian tetap menjadi hal terpenting dalam membangun dan mempertahankan loyalitas merek.

**Tabel 12.3** Praktik Terbaik untuk Penggunaan Kecerdasan Buatan Generatif yang Bertanggung Jawab dalam Pemasaran.

| Praktik Terbaik                                            | Alasan                                                                   |
|------------------------------------------------------------|--------------------------------------------------------------------------|
| Selalu ungkapkan konten yang dihasilkan AI kepada pengguna | Membangun kepercayaan dan menjaga transparansi                           |
| Hindari menghasilkan konten yang menyesatkan               | Mencegah masalah hukum dan menjaga integritas                            |
| Lakukan audit rutin terhadap model AI                      | Memastikan keakuratan model dan penggunaan yang etis                     |
| Selalu perbarui pedoman etika                              | Sesuai dengan standar industri dan praktik terbaik                       |
| Pastikan transparansi dalam proses AI                      | Membantu dalam komunikasi pemangku kepentingan dan membangun kepercayaan |

### **Masa Depan Kecerdasan Buatan Generatif: Apa yang Ada di Cakrawala?**

Masa depan AI generatif menyimpan janji yang sangat besar. Kami mengantisipasi kemajuan dalam penyempurnaan alat deteksi deepfake, memanfaatkan "perlombaan senjata" antara generator dan detektor. Selain itu, seiring AI terus memahami perilaku dan preferensi manusia dengan lebih baik, kita mungkin akan melihat pengembangan kampanye pemasaran yang lebih personal dan efisien.

Inovasi juga akan segera hadir dalam hal pembuatan konten secara real-time. Bayangkan sebuah dunia di mana kampanye pemasaran dibuat secara dinamis secara real-time berdasarkan peristiwa terkini, sentimen konsumen, atau bahkan pola cuaca.

### **Mempersiapkan Diri untuk Lanskap Pemasaran yang Didominasi Kecerdasan Buatan Generatif**

Untuk berkembang di lanskap yang didominasi AI, para profesional pemasaran harus melakukan hal berikut:

- **Menerapkan pembelajaran berkelanjutan.** Bidang ini berkembang pesat, dan para profesional perlu terus memperbarui diri dengan alat, teknik, dan pedoman etika terbaru.
- **Berinvestasi dalam alat dan platform.** Perusahaan harus berinvestasi dalam platform yang memungkinkan mereka untuk memanfaatkan kekuatan AI generatif secara bertanggung jawab dan efektif.
- **Terlibat dalam diskusi etika.** Pemasar harus berpartisipasi dalam percakapan yang lebih luas di masyarakat dan industri tentang implikasi etika dan pedoman untuk penggunaan teknologi yang ampuh ini.

Kesimpulannya, meskipun pertumbuhan dan kemampuan AI generatif menghadirkan tantangan, mereka juga menawarkan peluang yang belum pernah terjadi sebelumnya untuk inovasi dan koneksi di dunia pemasaran.

### **Konsep Kunci**

- **Penyalahgunaan dan informasi yang salah.** Seperti yang disoroti oleh Chesney dan Citron, kemampuan AI generatif telah menyebabkan terciptanya deepfake, yang berpotensi menyesatkan konsumen dan publik. Pemasar harus menyadari implikasi dari pembuatan konten yang dapat disalahartikan sebagai konten yang dihasilkan manusia secara autentik, termasuk potensi untuk menyesatkan konsumen dan merusak kredibilitas merek.
- **Transparansi dan ketidaktransparanan.** Model AI generatif, mirip dengan banyak model pembelajaran mesin, dapat bersifat buram, sehingga sulit untuk memahami bagaimana mereka membuat keputusan atau menghasilkan output. Ketidaktransparanan ini menimbulkan masalah etika tentang keadilan, bias, dan akuntabilitas.
- **Bias dan keadilan.** Model AI generatif dapat secara tidak sengaja mempelajari dan menyebarkan bias yang ada dalam data pelatihannya. Bias ini dapat menyebabkan hasil yang tidak adil atau diskriminatif ketika digunakan dalam aplikasi pemasaran.

- **Regulasi dan tata kelola.** Seiring dengan semakin meluasnya penggunaan AI generatif dalam pemasaran, akan ada kebutuhan yang semakin besar akan regulasi dan mekanisme tata kelola untuk memastikan penggunaan yang etis dan adil. Ini termasuk mekanisme untuk privasi data, transparansi, akuntabilitas, dan persetujuan.
- **Masa depan AI generatif dalam pemasaran.** Terlepas dari pertimbangan etis ini, masa depan AI generatif dalam pemasaran sangat cerah. AI generatif berpotensi merevolusi pembuatan konten, personalisasi, dan keterlibatan pelanggan, serta memberikan wawasan yang lebih dalam tentang perilaku dan preferensi pelanggan. Namun, pertimbangan dan tantangan etis perlu diatasi untuk mewujudkan potensi ini sepenuhnya.

### **Contoh Praktis**

Mari kita pertimbangkan kasus pengecer fesyen online yang menggunakan AI generatif untuk membuat iklan yang dipersonalisasi. AI menganalisis data tentang pembelian pelanggan di masa lalu, riwayat penelusuran, dan preferensi yang dinyatakan, lalu menggunakan data ini untuk menghasilkan iklan yang sangat disesuaikan, termasuk membuat gambar digital pakaian yang diprediksi akan disukai pelanggan.

Penerapan AI generatif ini memiliki beberapa implikasi etis yang harus dipertimbangkan oleh pengecer:

- **Transparansi dan informasi yang salah.** Iklan yang dihasilkan bisa sangat realistis sehingga pelanggan mungkin salah mengira iklan tersebut sebagai produk sebenarnya yang tersedia untuk dibeli. Jika tidak diungkapkan dengan benar, ini dapat dianggap menyesatkan. Seperti yang disarankan oleh Chesney dan Citron, penting bagi perusahaan untuk secara jelas menunjukkan gambar yang dihasilkan AI untuk menjaga transparansi dan menghindari informasi yang salah.
- **Bias dan keadilan.** Misalkan AI generatif pengecer telah dilatih pada data historis yang sebagian besar mencakup model muda dan kurus. AI tersebut kemudian dapat menghasilkan gambar yang melanggengkan tipe tubuh ini, sehingga menghasilkan penggambaran standar kecantikan yang bias. Seperti yang dicatat oleh Mehrabi dan rekan-rekan, untuk mencegah hal ini, pengecer harus memastikan bahwa data pelatihan mewakili beragam tipe tubuh dan bahwa AI tidak secara tidak adil mengecualikan atau menstereotipkan kelompok tertentu.
- **Privasi data.** Pengecer harus memperhatikan peraturan privasi data saat mengumpulkan dan menggunakan data pelanggan untuk AI. Sesuai dengan Goodman dan Flaxman, perusahaan harus memastikan bahwa data pelanggan yang digunakan untuk mempersonalisasi iklan dikumpulkan dan disimpan sesuai dengan peraturan seperti GDPR dan CCPA.
- **Prospek masa depan.** Ke depannya, seiring dengan peningkatan kemampuan AI generatif oleh peritel dan penanganan pertimbangan etis ini, mereka dapat menciptakan pengalaman belanja yang lebih personal. Misalnya, AI dapat menghasilkan gambar yang menunjukkan bagaimana pakaian akan terlihat pada tubuh

pelanggan sendiri atau di rumah mereka sendiri. Namun, pertimbangan etis, terutama tentang privasi dan transparansi data, akan tetap menjadi prioritas utama.

## **12.6 KESIMPULAN**

AI Generatif berada di garis depan gelombang transformatif dalam pemasaran, siap untuk mendefinisikan kembali ranah kreativitas, efisiensi, dan hiper-personalisasi. AI ini menjanjikan untuk memperkaya interaksi konsumen, menciptakan konten yang beresonansi pada tingkat individual yang mendalam. Namun demikian, kekuatan tersebut membawa implikasi etis yang substansial. Potensi teknologi generatif untuk merusak kepercayaan publik melalui penyalahgunaan atau penggunaan berlebihan merupakan kekhawatiran yang mendesak, dan kesalahan langkah dapat memicu skeptisisme atau bahkan memicu tindakan regulasi.

Batasan yang semakin kabur antara konten yang dibuat oleh manusia dan yang dihasilkan oleh AI memperkuat pentingnya otentisitas dalam strategi pemasaran. Transparansi dalam penggunaan AI akan sangat penting bagi merek yang bertujuan untuk mempertahankan hubungan yang tulus dengan pelanggan mereka. Otentisitas dalam pemasaran melampaui sekadar kejujuran; ini tentang menjalin dan mengembangkan hubungan yang tulus.

Seiring AI generatif semakin mengakar, industri pemasaran akan melihat lonjakan permintaan akan para profesional yang dapat dengan mahir mengintegrasikan AI ke dalam strategi pemasaran sambil menavigasi dilema etis yang dihadapkannya. Keterampilan baru yang muncul akan menuntut perpaduan antara kecakapan teknologi dan kearifan etis.

Bisnis yang secara strategis memanfaatkan AI generatif, meskipun dengan penuh kesadaran menangani masalah etika dan menekankan hubungan konsumen yang autentik, akan membedakan diri mereka sendiri. Keberhasilan di era baru ini tidak hanya akan diukur dari metrik konversi tetapi juga dari kedalaman dan ketahanan hubungan pelanggan di pasar yang mengutamakan digital.



## **BAB 13**

### **ETIKA, PRIVASI, DAN MASA DEPAN ILMU DATA PEMASARAN**

#### **13.1 PENDAHULUAN**

Era digital telah menghadirkan ledakan data, yang membentuk kembali lanskap pemasaran. Saat bisnis semakin mendalami dunia ilmu data pemasaran untuk memanfaatkan kekuatan data ini, mereka menavigasi jaringan peluang dan tantangan yang rumit. Potensi besar dari wawasan berbasis data menjanjikan peningkatan pengalaman pelanggan, penargetan yang tepat, dan strategi pemasaran yang inovatif. Namun, alat dan teknik yang sama yang memberdayakan kemajuan ini juga menimbulkan masalah etika, privasi, dan transparansi yang kompleks.

Dalam perlombaan untuk mendapatkan keunggulan kompetitif, sangat penting bagi bisnis untuk tetap menyadari tanggung jawab mendalam yang menyertai penggunaan data pribadi dan sensitif. Di luar sekadar kepatuhan terhadap peraturan, ada keharusan etis untuk menangani data dengan hati-hati, memastikan penghormatan dan perlindungan hak privasi individu. Tindakan penyeimbangan yang rumit antara memanfaatkan data dan menjaga kepercayaan ini sangat penting karena kesalahan langkah dapat menyebabkan bukan hanya konsekuensi hukum tetapi juga terkikisnya kepercayaan pelanggan dan kerusakan merek.

Dalam bab ini, kita akan membahas secara mendalam ranah etika, privasi, dan prospek masa depan ilmu data pemasaran yang saling terkait. Kita akan mengeksplorasi pertimbangan etika kritis seputar penggunaan data, menganalisis peraturan privasi utama seperti Peraturan Perlindungan Data Umum (GDPR) dan Undang-Undang Privasi Konsumen California (CCPA), dan merenungkan tren yang muncul yang siap mendefinisikan batas berikutnya dari pemasaran berbasis data. Melalui perpaduan wawasan teoretis, contoh praktis, dan studi kasus, bab ini menerangi jalan bagi bisnis dan ilmuwan data, membimbing mereka melalui dilema dan peluang etika yang ada di depan dalam domain ilmu data pemasaran yang terus berkembang.

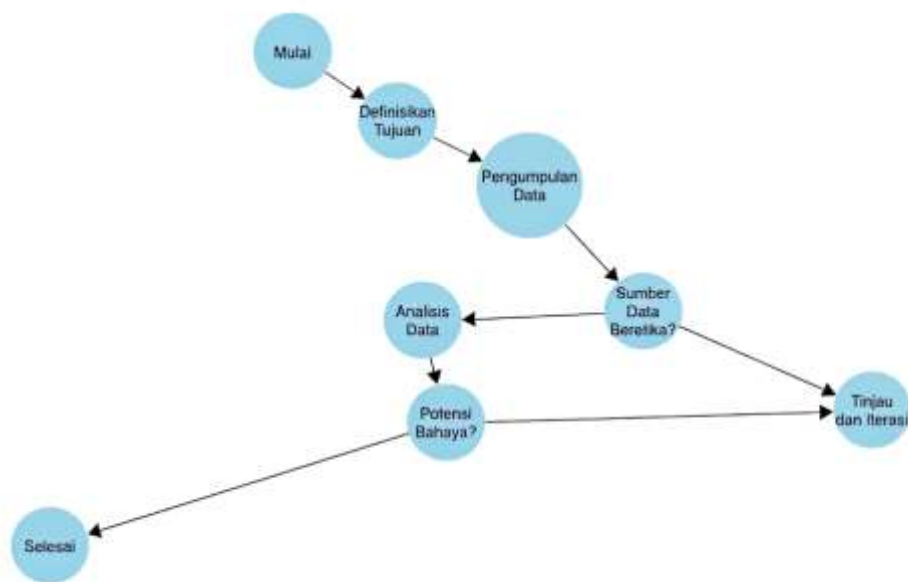
#### **13.2 PERTIMBANGAN ETIKA DALAM ILMU DATA PEMASARAN**

##### **Gambaran Umum Pertimbangan Etika**

Ilmu data pemasaran melibatkan pekerjaan dengan sejumlah besar data pribadi dan sensitif, yang menimbulkan banyak pertimbangan etika. Isu-isu etika ini terutama didasarkan pada penggunaan data yang bertanggung jawab dan perlindungan privasi serta hak konsumen.

Pertama, konsep persetujuan berdasarkan informasi merupakan perhatian etis utama. Persetujuan berdasarkan informasi memastikan bahwa pelanggan memahami bagaimana data mereka akan digunakan sebelum mereka memberikannya. Konsep ini mencakup penyediaan kebijakan privasi yang jelas dan transparan yang dapat dipahami oleh pelanggan rata-rata. Hal ini juga mengharuskan bisnis untuk menghormati pilihan pelanggan jika mereka memutuskan untuk tidak membagikan data mereka.

Kedua, prinsip minimalisasi data berperan, yang menyatakan bahwa hanya data yang diperlukan untuk tujuan yang dinyatakan yang harus dikumpulkan dan diproses. Prinsip ini membantu mengurangi risiko pelanggaran data dan penyalahgunaan data. Terakhir, masalah akurasi dan integritas data merupakan perhatian etis penting lainnya. Mempertahankan akurasi data memastikan bahwa keputusan yang dibuat berdasarkan data tersebut dapat diandalkan dan adil. Kesalahan penyajian atau ketidakakuratan dapat menyebabkan perlakuan tidak adil terhadap pelanggan dan berpotensi merusak reputasi perusahaan.



**Gambar 13.1** Proses Pengambilan Keputusan Etis dalam Ilmu Data.

Beberapa konsep kunci mendasari pertimbangan etis dalam ilmu data pemasaran (lihat Gambar 13.1):

- **Persetujuan berdasarkan informasi.** Ini adalah praktik mendapatkan izin eksplisit dari konsumen sebelum mengumpulkan, menggunakan, atau membagikan data mereka. Ini melibatkan penjelasan yang jelas tentang tujuan penggunaan data dan implikasi potensial dari berbagi data. Penghormatan terhadap otonomi ini membantu membangun kepercayaan antara bisnis dan pelanggan.
- **Minimalisasi data.** Prinsip ini menyarankan agar organisasi hanya mengumpulkan dan menyimpan data yang diperlukan untuk tujuan yang telah dinyatakan. Ini dipandang sebagai praktik kunci dalam menghormati privasi konsumen dan mengurangi risiko pelanggaran data.
- **Akurasi data.** Memastikan bahwa data yang digunakan dalam proses analitik akurat dan mutakhir sangat penting untuk pengambilan keputusan yang adil dan efektif. Data yang tidak akurat dapat menyebabkan hasil yang tidak adil atau diskriminatif.
- **Privasi sejak tahap desain.** Konsep ini mengusulkan bahwa pertimbangan privasi harus tertanam dalam desain sistem dan praktik sejak awal, daripada ditambahkan sebagai pertimbangan belakangan.

- **Transparansi.** Ini melibatkan komunikasi yang jelas dengan konsumen tentang praktik pengumpulan data, penggunaan data, dan langkah-langkah perlindungan data yang ada. Transparansi menumbuhkan kepercayaan dan memungkinkan konsumen untuk membuat keputusan yang tepat mengenai berbagi data.
- **Akuntabilitas.** Perusahaan harus bertanggung jawab atas praktik data mereka, termasuk kepatuhan terhadap hukum dan peraturan privasi, dan menerapkan langkah-langkah perlindungan data.

### **Pentingnya Etika dalam Penanganan Data**

Era digital membawa serta akses yang belum pernah terjadi sebelumnya ke data, yang sebagian besar bersifat pribadi. Meskipun akses ini memberikan banyak peluang bagi bisnis untuk meningkatkan strategi pemasaran dan pengalaman pelanggan mereka, hal ini juga menimbulkan tantangan etis yang signifikan. Etika dalam penanganan data berkaitan dengan prinsip dan standar moral yang memandu tindakan saat mengumpulkan, menyimpan, memproses, dan berbagi data. Konsekuensi dari tidak menjunjung tinggi standar ini dapat merugikan baik individu yang datanya disalahgunakan maupun bisnis yang bertanggung jawab atas penyalahgunaan tersebut.

Masalah seperti pelanggaran data dapat menyebabkan kerugian finansial, kerusakan reputasi, dan potensi konsekuensi hukum bagi bisnis. Bagi individu, penyalahgunaan data pribadi dapat mengakibatkan hilangnya privasi, dampak finansial, atau diskriminasi. Oleh karena itu, mematuhi pedoman etika bukan hanya persyaratan peraturan tetapi juga tanggung jawab sosial bagi bisnis.

### **Potensi Penyalahgunaan Data dan Analitik**

Jumlah data yang sangat besar yang tersedia bagi pemasar dan alat analitik canggih yang mereka miliki terkadang dapat menyebabkan konsekuensi yang tidak diinginkan. Potensi penyalahgunaan meliputi hal-hal berikut (lihat Tabel 13.1):

- **Diskriminasi data.** Algoritma mungkin secara tidak sengaja menguntungkan satu kelompok daripada kelompok lain. Misalnya, sistem penargetan iklan mungkin secara tidak proporsional menampilkan iklan pekerjaan bergaji tinggi kepada pria lebih banyak daripada wanita.
- **Pelanggaran privasi.** Pemasar mungkin mengumpulkan lebih banyak data daripada yang diperlukan atau menggunakan data dengan cara yang tidak disetujui pelanggan, yang menyebabkan potensi pelanggaran privasi.
- **Iklan yang menyesatkan.** Menggunakan analitik data untuk memanipulasi perilaku konsumen dengan mendorong produk atau layanan yang mungkin tidak mereka butuhkan, atau menggunakan trik psikologis, dapat dianggap tidak etis.
- **Penjualan kembali data.** Menjual data pengguna kepada pihak ketiga tanpa persetujuan eksplisit dapat menyebabkan pemasaran yang tidak diinginkan dan potensi pelanggaran kepercayaan pengguna.

**Tabel 13.1** Penyalahgunaan Data yang Umum dengan Konsekuensi dan Pertimbangan Etika yang Sesuai.

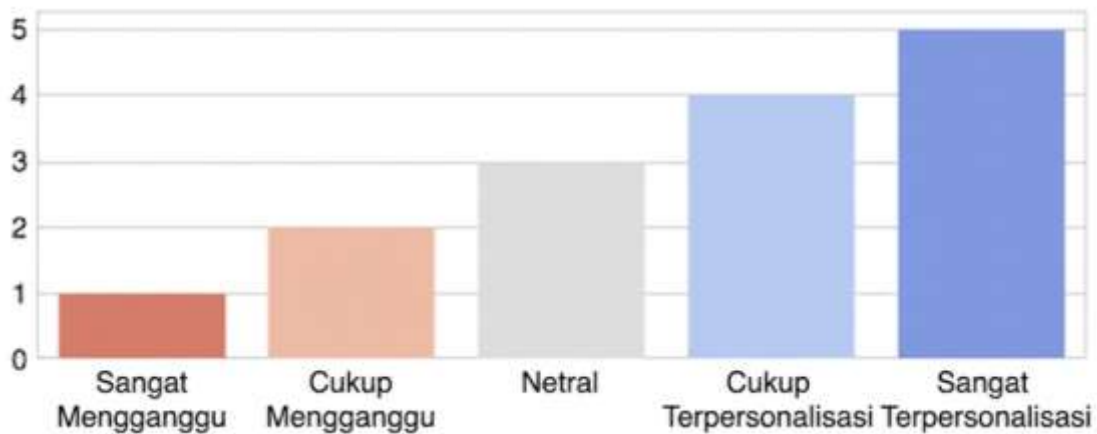
| Penyalahgunaan Data                                     | Konsekuensi                                                | Pertimbangan Etis                 |
|---------------------------------------------------------|------------------------------------------------------------|-----------------------------------|
| Pengintaian data tanpa persetujuan                      | Konsekuensi hukum dan hilangnya kepercayaan                | Menghormati privasi               |
| Penjualan data pribadi kepada pihak ketiga              | Hilangnya kepercayaan pelanggan dan potensi tindakan hukum | Menghormati privasi dan kejujuran |
| Penggunaan data untuk mendiskriminasi kelompok tertentu | Reaksi publik dan potensi konsekuensi hukum                | Keadilan dan tidak diskriminasi   |
| Pemalsuan hasil data                                    | Keputusan yang salah arah dan hilangnya kredibilitas       | Kejujuran dan ketelitian          |
| Mengabaikan hukum privasi data                          | Denda besar dan tindakan hukum                             | Menghormati hukum dan peraturan   |

### Menyeimbangkan Personalisasi dan Intrusi

Aplikasi utama ilmu data pemasaran adalah personalisasi, menyesuaikan konten dan pengalaman kepada pengguna individu berdasarkan preferensi dan perilaku mereka. Meskipun personalisasi dapat meningkatkan pengalaman pengguna, penting untuk mencapai keseimbangan agar tidak terkesan mengganggu atau menyeramkan. Sebuah studi oleh Marreiros dan rekan-rekan. menemukan bahwa pengguna dapat merasa dilanggar ketika personalisasi terlalu tepat, terutama jika mereka tidak menyadari data yang telah dikumpulkan tentang mereka.

Bagi bisnis, sangat penting untuk memastikan bahwa strategi personalisasi bersifat transparan, dan pengguna diberi tahu tentang bagaimana data mereka digunakan. Memberikan pengguna kendali atas data mereka dan penggunaannya juga dapat membantu mencapai keseimbangan yang tepat. Praktisi harus mempertimbangkan strategi berikut tentang bagaimana keseimbangan ini dapat dicapai (lihat Gambar 13.2):

- **Prioritaskan transparansi:**
  - **Kejelasan informasi.** Selalu pastikan bahwa setiap proses pengumpulan data dijelaskan dengan istilah sederhana, tanpa bersembunyi di balik jargon. Pengguna harus dapat dengan mudah memahami data apa yang dikumpulkan dan untuk tujuan apa.
  - **Mekanisme persetujuan.** Secara eksplisit mintalah persetujuan sebelum mengumpulkan data pribadi dan izinkan pengguna untuk menolak kapan saja. Pemberitahuan tepat waktu, yang memberikan informasi saat relevan, juga dapat bermanfaat.



**Gambar 13.2** Keseimbangan Antara Personalisasi dan Intrusi dalam Kampanye Pemasaran.

- **Berikan kendali kepada pengguna:**
  - **Akses data.** Izinkan pengguna untuk melihat data yang telah dikumpulkan tentang mereka. Ini tidak hanya membangun kepercayaan tetapi juga memberikan kesempatan untuk koreksi data, memastikan personalisasi yang lebih akurat.
  - **Opsi kustomisasi.** Biarkan pengguna memutuskan tingkat personalisasi yang mereka rasa nyaman. Beberapa mungkin lebih menyukai kategori yang lebih luas, dan yang lain mungkin menghargai personalisasi yang berorientasi pada detail.
- **Batasi penyimpanan data:**
  - **Kebijakan penyimpanan.** Terapkan kebijakan ketat tentang berapa lama data akan disimpan dan patuhi kebijakan tersebut. Ini mengurangi potensi penyalahgunaan dan seringkali dipandang positif oleh konsumen.
- **Edukasi pengguna:**
  - **Kampanye kesadaran.** Lakukan kampanye untuk mengedukasi pengguna tentang praktik pengumpulan data, manfaat personalisasi, dan potensi risiko. Pengguna yang terinformasi dengan baik lebih cenderung mempercayai dan terlibat dengan konten yang dipersonalisasi.
  - **Umpan balik.** Kumpulkan umpan balik secara teratur tentang strategi personalisasi untuk memahami sentimen pengguna dan sesuaikan strategi accordingly.
- **Anonimkan data:**
  - **Penyamaran data.** Jika memungkinkan, anonimkan data untuk memastikan bahwa meskipun diakses, hal itu tidak membahayakan privasi individu.
- **Pantau dan sesuaikan:**
  - **Algoritma dinamis.** Gunakan algoritma yang secara dinamis menyesuaikan tingkat personalisasi berdasarkan umpan balik dan keterlibatan pengguna. Jika pengguna secara konsisten mengabaikan atau tidak menyukai konten personalisasi tertentu, sistem harus belajar dan beradaptasi.
  - **Audit berkala.** Lakukan audit berkala untuk memastikan bahwa strategi personalisasi selaras dengan pedoman etika dan preferensi pengguna.

Meskipun kemampuan teknologi untuk personalisasi sudah ada, sentuhan manusia, pemahaman, dan etika lah yang menentukan efektivitas dan penerimaan strategi tersebut. Personalisasi, jika dilakukan dengan benar, berpotensi menjadi landasan keterlibatan pengguna di era digital. Namun, penghormatan terhadap otonomi dan privasi pengguna harus selalu menjadi prinsip panduan. Dengan menciptakan hubungan simbiosis di mana pengguna merasa dipahami daripada diawasi, bisnis dapat benar-benar memanfaatkan kekuatan ilmu data pemasaran.

#### **Dilema Etika dan Solusinya**

- **Model prediksi kehamilan Target.** Target, raksasa ritel AS, mengembangkan model untuk memprediksi status kehamilan pelanggannya berdasarkan pola belanja mereka. Namun, model tersebut secara tidak sengaja mengungkapkan kehamilan seorang remaja kepada ayahnya dengan mengirimkan kupon produk kehamilan kepadanya, yang menyebabkan bencana hubungan masyarakat. Kasus ini menggarisbawahi perlunya perusahaan untuk berhati-hati saat menggunakan analitik prediktif, terutama ketika berurusan dengan informasi sensitif.
- **Solusi.** Setelah insiden tersebut, Target menjadi lebih halus dalam pemasarannya. Alih-alih hanya mengirimkan kupon produk kehamilan, mereka mencampurnya dengan penawaran lain yang tidak terkait untuk membuatnya kurang mencolok.
- **Peta panas Strava.** Pada tahun 2018, Strava, aplikasi pelacak kebugaran, merilis peta panas global yang menampilkan rute lari populer. Namun, peta tersebut secara tidak sengaja mengungkapkan lokasi pangkalan militer dan rutinitas tentara, yang menimbulkan potensi ancaman keamanan.
- **Resolusi.** Strava menanggapi dengan menekankan pengaturan privasi pengguna dan berjanji untuk bekerja sama dengan pejabat militer dan pemerintah untuk mengatasi area sensitif di peta.

Studi kasus ini menekankan pentingnya pertimbangan etis dalam ilmu data pemasaran. Studi kasus ini menunjukkan bahwa bahkan dengan niat terbaik, data dapat disalahgunakan atau ditangani dengan tidak benar, yang menyebabkan konsekuensi yang tidak diinginkan. Seiring dengan terus berkembangnya bidang ilmu data pemasaran, perusahaan harus waspada dan memprioritaskan pertimbangan etis dalam praktik data mereka.

Berdasarkan diskusi tentang tantangan etis, Tabel 13.2 membahas lebih dalam dengan menyajikan studi kasus tambahan yang menyoroti dilema etis spesifik yang dihadapi dalam ilmu data pemasaran, bersama dengan solusi yang mungkin diterapkan dan hasil masing-masing, yang menggambarkan sifat kompleks pengambilan keputusan etis di bidang ini.

**Tabel 13.2** Studi Kasus Tambahan yang Menyoroti Dilema Etis yang Dihadapi, Solusi yang Mungkin, dan Hasil.

| Studi Kasus                                                | Dilema Etika                                         | Solusi yang Mungkin                               | Hasil                                 |
|------------------------------------------------------------|------------------------------------------------------|---------------------------------------------------|---------------------------------------|
| <b>Studi A: Iklan bertarget berdasarkan data kesehatan</b> | Kekhawatiran privasi terkait data kesehatan sensitif | Anonimkan data dan dapatkan persetujuan eksplisit | Meningkatkan kepercayaan pengguna dan |

|                                                                                      |                                                           |                                                                                        |                                                                                                |
|--------------------------------------------------------------------------------------|-----------------------------------------------------------|----------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|
|                                                                                      |                                                           |                                                                                        | keterlibatan iklan yang lebih baik                                                             |
| <b>Studi B: Pengalaman belanja yang dipersonalisasi menggunakan pengenalan wajah</b> | Pelanggaran privasi dan potensi penyalahgunaan data wajah | Beri tahu pelanggan dan berikan opsi untuk menolak                                     | Respon beragam; sebagian mengapresiasi, sebagian lainnya memilih untuk tidak menggunakan iklan |
| <b>Studi C: Menjual data pengguna kepada perusahaan asuransi</b>                     | Potensi penyalahgunaan yang menyebabkan premi tidak adil  | Pastikan data dianonimkan dan tidak digunakan untuk penetapan harga yang diskriminatif | Menghadapi reaksi negatif dan merevisi kebijakannya                                            |
| <b>Studi D: Kepolisian prediktif menggunakan data demografis</b>                     | Potensi bias dan diskriminasi rasial                      | Gunakan sumber data yang tidak diskriminatif dan audit berkala                         | Menurunkan angka kejahatan tetapi menghadapi kritik karena potensi bias                        |

### **Pertimbangan Etika dalam Pemasaran Media Sosial**

Pemasaran media sosial menyediakan sumber data yang kaya tentang preferensi, perilaku, dan sentimen pelanggan. Namun, penggunaan data ini harus diseimbangkan dengan pertimbangan etika. Misalnya, Cambridge Analytica, sebuah perusahaan konsultan politik Inggris, ditemukan telah mengumpulkan data pribadi jutaan profil Facebook orang tanpa persetujuan mereka dan menggunakannya untuk tujuan iklan politik. Kasus ini menimbulkan kekhawatiran etika yang serius tentang persetujuan yang diinformasikan dan privasi dalam pengumpulan dan penggunaan data. Hal ini juga memicu debat global tentang tanggung jawab platform media sosial untuk melindungi data pengguna.

Sebaliknya, pertimbangkan Patagonia, sebuah perusahaan pakaian luar ruangan yang dikenal dengan praktik bisnis etisnya. Dalam pemasaran media sosial mereka, mereka memprioritaskan transparansi dan persetujuan yang diinformasikan. Mereka secara jelas mengkomunikasikan bagaimana mereka mengumpulkan dan menggunakan data pelanggan, dan mereka menyediakan opsi bagi pelanggan untuk mengontrol data mereka. Dengan melakukan hal tersebut, mereka menumbuhkan kepercayaan dan loyalitas di antara pelanggan mereka sambil menghormati privasi mereka.

Contoh-contoh ini menyoroti pentingnya pertimbangan etis dalam pemasaran ilmu data. Perusahaan yang memprioritaskan praktik data etis tidak hanya mematuhi peraturan tetapi juga membangun kepercayaan dengan pelanggan mereka, yang dapat mengarah pada kesuksesan bisnis jangka panjang.

### **13.3 PERATURAN PRIVASI DATA**

#### **Gambaran Umum Peraturan Privasi Data**

Karena data telah menjadi aset penting dalam ekonomi digital, kekhawatiran tentang privasi data telah mendorong terciptanya berbagai peraturan di seluruh dunia untuk

melindungi informasi pribadi konsumen. Yang paling terkenal di antaranya adalah GDPR di Uni Eropa dan CCPA di Amerika Serikat (lihat Tabel 13.3).

**Tabel 13.3** Peraturan Utama Perlindungan Data Pribadi (GDPR, CCPA) yang Menunjukkan Tujuan Utama, Entitas yang Tercakup, dan Sanksi.

| Regulasi                                                | Tujuan Utama                                                                                                                                                           | Entitas yang Tercakup                                                                                                        | Sanksi                                                                                                                     |
|---------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| <b>GDPR (Peraturan Perlindungan Data Umum)</b>          | Melindungi hak individu terkait data pribadi dan membentuk kembali pendekatan organisasi terhadap privasi data                                                         | Semua perusahaan yang memproses data pribadi subjek yang berdomisili di Uni Eropa, terlepas dari lokasi perusahaan tersebut. | Hingga Rp220 miliar atau 4% dari omset global tahunan, mana pun yang lebih tinggi                                          |
| <b>CCPA (Undang-Undang Privasi Konsumen California)</b> | Memberikan hak kepada konsumen California untuk mengetahui data pribadi apa yang dikumpulkan, menolak penjualan data pribadi mereka, dan mengakses data pribadi mereka | Bisnis yang melayani penduduk California dan memiliki pendapatan tahunan setidaknya Rp25.000.000.                            | Hingga Rp75.000.000 untuk setiap pelanggaran yang disengaja dan Rp25.000.000 untuk setiap pelanggaran yang tidak disengaja |

GDPR, yang mulai berlaku pada Mei 2018, adalah undang-undang perlindungan data komprehensif yang mengatur pemrosesan data pribadi individu di Uni Eropa dan Wilayah Ekonomi Eropa. Undang-undang ini memberi individu lebih banyak kendali atas data pribadi mereka dan memberlakukan aturan ketat pada pihak yang menyimpan dan memproses data ini, di mana pun mereka berada.

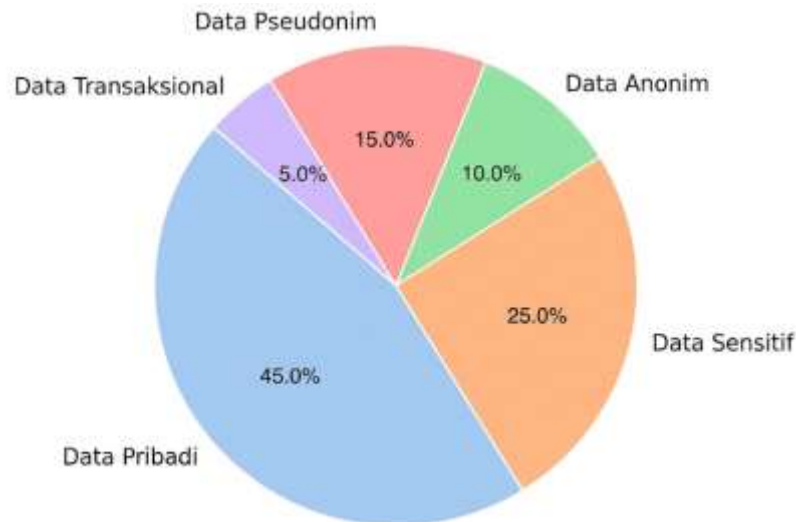
Di sisi lain Samudra Atlantik, CCPA, yang mulai berlaku pada Januari 2020, memberikan hak-hak khusus kepada penduduk California atas informasi pribadi mereka, termasuk hak untuk mengetahui informasi pribadi yang dikumpulkan oleh suatu bisnis tentang mereka dan hak untuk menghapus informasi pribadi yang dikumpulkan dari mereka (dengan beberapa pengecualian). CCPA juga memberikan hak untuk menolak penjualan informasi pribadi mereka.

Regulasi ini mencerminkan tren global menuju penguatan hak perlindungan data, dengan wilayah dan negara lain seperti Brasil, India, dan Tiongkok juga menerapkan atau merencanakan undang-undang serupa. Tren ini memiliki implikasi signifikan bagi ilmu data pemasaran karena memengaruhi bagaimana pemasar dapat mengumpulkan, menyimpan, memproses, dan menggunakan data konsumen.



### Implikasi untuk Pengumpulan dan Pemrosesan Data

Evolusi regulasi privasi data di seluruh dunia memiliki implikasi mendalam terhadap bagaimana bisnis mendekati pengumpulan dan pemrosesan data dalam inisiatif pemasaran mereka (lihat Gambar 13.3):



**Gambar 13.3** Persentase Jenis Data yang Terpengaruh oleh Regulasi.

- **Lingkup pengumpulan data.** Regulasi seperti GDPR mewajibkan data dikumpulkan untuk tujuan tertentu dan harus dibatasi pada apa yang diperlukan untuk tujuan tersebut. Ini berarti pemasar perlu tepat tentang mengapa mereka mengumpulkan data dan memastikan bahwa redundansi data diminimalkan.
- **Retensi data.** Data tidak lagi dapat disimpan tanpa batas waktu. Bisnis harus memiliki kebijakan retensi data yang jelas, di mana data dihapus atau dianonimkan setelah tujuan yang dimaksudkan telah tercapai.
- **Manajemen persetujuan.** Persetujuan implisit atau tersirat tidak lagi cukup. Regulasi menuntut persetujuan eksplisit, yang telah menyebabkan lonjakan formulir opt-in dan banner persetujuan cookie di situs web. Selain itu, bisnis harus menyediakan mekanisme bagi pengguna untuk menarik persetujuan mereka kapan saja dengan mudah.
- **Akurasi data.** Regulasi menggarisbawahi pentingnya akurasi data. Individu memiliki hak untuk memperbaiki ketidakakuratan dalam data pribadi mereka, sehingga organisasi berkewajiban untuk menerapkan proses yang memfasilitasi koreksi tersebut.
- **Keamanan data.** Dengan pelanggaran data sebagai perhatian utama, perusahaan sekarang diharuskan untuk memiliki langkah-langkah keamanan data yang kuat, dengan sanksi untuk kelalaian.

### Mencapai Kepatuhan: Praktik Terbaik dan Daftar Periksa

Memastikan kepatuhan terhadap peraturan privasi data yang terus berkembang membutuhkan pendekatan proaktif dan sistematis. Berikut beberapa praktik terbaik:

- **Kesadaran dan pelatihan.** Secara teratur didik dan latih anggota staf, terutama mereka yang menangani data, tentang pentingnya perlindungan data dan kepatuhan.

- **Penilaian dampak privasi (PIA).** Lakukan PIA sebelum memulai proyek baru atau mengadopsi teknologi baru untuk memahami potensi risiko privasi dan mengatasinya secara preventif.
- **Pemetaan data.** Pahami di mana data pribadi berada di sistem Anda, siapa yang memiliki akses, dan mengapa. Ini membantu dalam manajemen data dan kepatuhan yang efektif.
- **Audit berkala.** Tinjau dan audit aktivitas pemrosesan data secara berkala untuk mengidentifikasi potensi area ketidakpatuhan.
- **Libatkan petugas perlindungan data (DPO).** Bagi organisasi yang lebih besar atau yang terlibat dalam pemrosesan data yang ekstensif, akan bermanfaat untuk memiliki DPO (Data Protection Officer), peran yang diamanatkan oleh GDPR untuk bisnis tertentu.
- **Kebijakan yang jelas.** Susun kebijakan perlindungan data dan privasi yang jelas, sehingga dapat diakses oleh anggota staf dan pelanggan.
- **Manajemen insiden.** Miliki rencana yang jelas untuk menangani pelanggaran data, termasuk memberi tahu pihak berwenang yang diperlukan dan individu yang terkena dampak.

#### **Daftar periksa**

- Sesi pelatihan rutin yang dijadwalkan
- PIA (Protected Impact Assessment) dilakukan untuk proyek baru
- Peta data diperbarui
- Audit dilakukan dalam enam bulan terakhir
- Kebijakan perlindungan data telah diterapkan dan ditinjau
- Rencana manajemen insiden telah disiapkan

#### **Lanskap Global: Perbedaan dan Persamaan dalam Regulasi**

Regulasi perlindungan data bervariasi di seluruh dunia, mencerminkan perbedaan budaya, sosial, dan politik. Namun, mereka sering kali memiliki prinsip-prinsip umum:

- **Hak individu.** Sebagian besar regulasi memberikan hak kepada individu atas data mereka, termasuk akses, koreksi, penghapusan, dan terkadang portabilitas.
- **Akuntabilitas dan tata kelola.** Organisasi umumnya bertanggung jawab untuk melindungi data, sehingga memerlukan mekanisme tata kelola seperti penilaian dampak perlindungan data dan penunjukan petugas perlindungan data.
- **Arus data internasional.** Regulasi sering kali membahas transfer data pribadi lintas batas, memastikan bahwa data tetap terlindungi saat ditransfer secara internasional.
- **Persetujuan.** Kebutuhan akan persetujuan eksplisit sebelum pengumpulan dan pemrosesan data merupakan tema yang berulang di berbagai regulasi, meskipun sifat dan persyaratan yang tepat seputar persetujuan mungkin bervariasi.

Perbedaan muncul dalam nuansanya, dengan beberapa wilayah lebih menekankan prinsip-prinsip tertentu daripada yang lain (lihat Tabel 13.4). Misalnya, GDPR Uni Eropa sangat menekankan hak-hak individu, sedangkan Amerika Serikat memiliki pendekatan sektoral terhadap perlindungan data, dengan aturan yang berbeda untuk data kesehatan, data

keuangan, dan sebagainya. Negara-negara Asia-Pasifik, seperti Jepang dan Australia, memiliki perpaduan prinsip yang unik, yang mencerminkan pengaruh Barat dan regional.

**Tabel 13.4** Perbedaan dan Persamaan di Antara Regulasi Privasi Data Global.

| Aspek                                 | GDPR                                                                                           | CCPA                                                                                                         | Persamaan                                                                        |
|---------------------------------------|------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| <b>Lingkup</b>                        | Berlaku untuk semua organisasi yang memproses data penduduk Uni Eropa, tanpa memandang lokasi. | Berlaku untuk bisnis yang mengumpulkan data penduduk California dan memenuhi kriteria tertentu               | Keduanya bertujuan untuk melindungi hak privasi individu.                        |
| <b>Hak subjek data</b>                | Hak untuk dilupakan, portabilitas data, akses, perbaikan, dan keberatan                        | Hak untuk mengetahui, menolak, dan penghapusan data                                                          | Keduanya memberikan hak untuk mengakses dan menghapus data pribadi.              |
| <b>Pemberitahuan pelanggaran data</b> | 72 jam sejak diketahui adanya pelanggaran                                                      | Tidak didefinisikan secara spesifik                                                                          | Keduanya mewajibkan organisasi untuk memberi tahu jika terjadi pelanggaran data. |
| <b>Denda</b>                          | Hingga 4% dari omset global tahunan atau Rp220.000 juta, mana pun yang lebih tinggi.           | Hingga Rp75.000.000 untuk pelanggaran yang disengaja dan Rp25.000.000 untuk pelanggaran yang tidak disengaja | Keduanya mengenakan sanksi atas pelanggaran.                                     |

Kesimpulannya, meskipun terdapat nuansa dan persyaratan khusus di setiap yurisdiksi, prinsip-prinsip dasar perlindungan data tetap serupa, menekankan hak individu, akuntabilitas, dan penggunaan data secara etis.

#### **Konsep Kunci dalam Regulasi Privasi Data**

Memahami konsep-konsep kunci regulasi privasi data seperti GDPR dan CCPA sangat penting bagi organisasi untuk memastikan kepatuhan dan menghindari sanksi yang besar:

- **Data pribadi.** Berdasarkan GDPR, data pribadi mengacu pada informasi apa pun yang berkaitan dengan orang perseorangan yang teridentifikasi atau dapat diidentifikasi. Ini termasuk nama, nomor identifikasi, data lokasi, pengidentifikasi daring, atau satu atau lebih faktor spesifik yang berkaitan dengan identitas fisik, fisiologis, genetik, mental, ekonomi, budaya, atau sosial orang perseorangan tersebut. Demikian pula, CCPA mengacu pada informasi pribadi sebagai informasi yang mengidentifikasi, berkaitan dengan, menggambarkan, secara wajar mampu dikaitkan dengan, atau secara wajar

dapat dihubungkan, secara langsung atau tidak langsung, dengan konsumen atau rumah tangga tertentu.

- **Persetujuan.** GDPR dan CCPA mewajibkan bisnis untuk memperoleh persetujuan eksplisit dari individu sebelum mengumpulkan dan memproses data pribadi mereka. Persetujuan tersebut harus diberikan secara bebas, spesifik, terinformasi, dan tidak ambigu.
- **Hak untuk mengakses dan hak untuk penghapusan.** Kedua peraturan tersebut memberikan hak kepada individu untuk mengakses data pribadi mereka yang disimpan oleh suatu organisasi dan hak untuk meminta penghapusan data pribadi mereka dalam keadaan tertentu.
- **Petugas Perlindungan Data (DPO).** GDPR mewajibkan organisasi untuk menunjuk seorang DPO jika mereka melakukan pemantauan sistematis skala besar atau memproses banyak data pribadi yang sensitif. DPO mengawasi strategi dan implementasi perlindungan data untuk memastikan kepatuhan terhadap persyaratan GDPR.
- **Pemberitahuan pelanggaran data.** Berdasarkan GDPR, organisasi harus melaporkan jenis pelanggaran data pribadi tertentu kepada otoritas pengawas yang relevan dalam waktu 72 jam setelah mengetahui pelanggaran tersebut, jika memungkinkan.

Dengan memahami konsep-konsep kunci ini, organisasi dapat menavigasi lanskap kompleks peraturan privasi data dan membangun kepercayaan dengan pelanggan mereka.

#### **Kepatuhan GDPR dalam Pemasaran Email**

Kepatuhan GDPR merupakan aspek penting dari pemasaran email. Banyak perusahaan harus menyesuaikan praktik pemasaran email mereka untuk mematuhi peraturan baru. Mari kita pertimbangkan contoh bagaimana perusahaan ritel mematuhi GDPR dalam praktik pemasaran emailnya:

- **Persetujuan (Opt-in).** Langkah pertama dalam kepatuhan GDPR adalah memastikan bahwa semua penerima email telah memberikan persetujuan eksplisit untuk menerima email pemasaran. Ini biasanya dilakukan melalui proses persetujuan (opt-in). Misalnya, ketika pengguna membuat akun atau melakukan pembelian, perusahaan ritel dapat menyertakan kotak centang yang harus diklik pengguna untuk memilih menerima email pemasaran. Kotak ini tidak dapat dicentang sebelumnya; pengguna harus mengambil tindakan yang jelas dan tegas untuk memberikan persetujuan.
- **Persetujuan yang tidak ambigu.** Berdasarkan GDPR, persetujuan harus spesifik dan tidak ambigu. Ini berarti perusahaan ritel harus dengan jelas menyatakan apa yang disetujui pengguna. Misalnya, kotak persetujuan (opt-in) mungkin bertuliskan, “Ya, saya ingin menerima email promosi tentang produk dan layanan Anda.”
- **Pencabutan persetujuan.** GDPR juga mensyaratkan bahwa pencabutan persetujuan harus semudah memberikannya. Oleh karena itu, setiap email pemasaran yang dikirim oleh perusahaan ritel menyertakan tautan berhenti berlangganan di bagian bawah. Jika pengguna mengklik tautan ini, mereka harus dapat dengan mudah dan segera berhenti berlangganan dari email di masa mendatang.

- **Minimisasi data.** Perusahaan ritel mematuhi prinsip minimisasi data, yang berarti mereka hanya mengumpulkan data yang diperlukan untuk pemasaran email. Mereka tidak meminta atau menyimpan data pribadi tambahan apa pun dari pelanggan mereka.
- **Perlindungan data.** Perusahaan juga menerapkan langkah-langkah keamanan yang kuat untuk melindungi data pribadi pelanggannya, seperti enkripsi dan audit keamanan reguler, untuk mencegah pelanggaran data.

Contoh ini menggambarkan bagaimana GDPR memengaruhi praktik pemasaran email. Meskipun kepatuhan GDPR memerlukan beberapa perubahan dan ketelitian yang berkelanjutan, pada akhirnya hal itu membantu membangun kepercayaan dan hubungan yang lebih baik dengan pelanggan.

### **13.4 BIAS, KEADILAN, DAN TRANSPARANSI**

#### **Gambaran Umum Bias, Keadilan, dan Transparansi**

Bias, keadilan, dan transparansi merupakan pertimbangan penting dalam ilmu data pemasaran. Dengan meningkatnya penggunaan pembelajaran mesin (ML) dan kecerdasan buatan (AI) dalam pemasaran, aspek-aspek ini menjadi semakin signifikan.

- **Bias** dalam ilmu data mengacu pada kesalahan sistematis yang diperkenalkan oleh pengumpulan data, pemrosesan data, atau algoritma yang membuat hasil condong ke arah tertentu. Misalnya, jika algoritma rekomendasi dilatih pada data dari kelompok demografis tertentu, algoritma tersebut mungkin tidak berkinerja baik untuk kelompok lain, sehingga menciptakan bias. Hal ini dapat menyebabkan hasil yang tidak adil dan dapat merugikan kelompok pelanggan tertentu.
- **Keadilan** dalam ilmu data berarti bahwa hasil dari suatu algoritma tidak mendiskriminasi kelompok tertentu berdasarkan atribut sensitif seperti ras, jenis kelamin, usia, dan sebagainya. Penting untuk memastikan bahwa model yang digunakan dalam pemasaran tidak secara tidak sengaja menyebabkan perlakuan tidak adil terhadap kelompok pelanggan tertentu.
- **Transparansi** dalam ilmu data mengacu pada kemampuan untuk memahami dan menafsirkan cara kerja dan keputusan yang dibuat oleh algoritma. Hal ini sangat penting dalam pemasaran, di mana keputusan yang dibuat oleh algoritma dapat berdampak signifikan pada pelanggan. Transparansi juga membantu membangun kepercayaan dengan pelanggan karena mereka memahami bagaimana data mereka digunakan.

Dalam konteks ilmu data pemasaran, memastikan mitigasi bias, keadilan, dan transparansi bukan hanya tentang mematuhi peraturan atau menghindari bencana hubungan masyarakat. Ini tentang membangun kepercayaan dengan pelanggan, yang dapat mengarah pada hubungan pelanggan yang lebih baik dan pada akhirnya keunggulan kompetitif.

#### **Mengenali dan Mengatasi Bias dalam Data dan Model**

Bias adalah risiko bawaan ketika berurusan dengan pengumpulan data dan pembuatan model dalam ilmu data pemasaran. Penting untuk menyadari bahwa bias dapat muncul di

setiap tahap: dari data yang dikumpulkan hingga algoritma yang dikembangkan. Misalnya, jika suatu dataset sebagian besar terdiri dari satu kelompok demografis, wawasan yang diperoleh mungkin tidak berlaku untuk audiens yang lebih luas. Untuk mengatasi hal ini, bisnis perlu melakukan hal berikut.

#### **Audit Data**

Sebelum pengembangan model, evaluasi keterwakilan data. Pastikan data tersebut mencakup beragam perspektif dan tidak secara tidak sengaja mengecualikan segmen penting.

#### **Gunakan Data Pelatihan yang Beragam**

Dalam ML, data pelatihan menentukan pemahaman model. Menggabungkan data yang beragam membantu dalam mengembangkan model yang lebih umum dan kurang bias.

#### **Evaluasi Ulang Model Secara Teratur**

Penilaian berkelanjutan terhadap model untuk potensi bias dan penyempurnaan model tersebut memastikan bahwa model tetap relevan dan akurat dari waktu ke waktu.

**Tabel 13.5** Sumber Bias Umum dalam Data dan Metode yang Sesuai untuk Mengatasinya.

| <b>Sumber-Sumber Bias</b> | <b>Metode untuk Mengatasi Masalah</b>                                                         |
|---------------------------|-----------------------------------------------------------------------------------------------|
| <b>Pengambilan Sampel</b> | Gunakan pengambilan sampel bertingkat dan pastikan sampel representatif                       |
| <b>Pengukuran</b>         | Kalibrasi alat ukur secara berkala dan latih pengumpul data                                   |
| <b>Konfirmasi</b>         | Cari audit dan pemantauan eksternal Tantang dan validasi asumsi                               |
| <b>Algoritma</b>          | Gunakan model yang dapat diinterpretasikan dan validasi terhadap berbagai kelompok demografis |
| <b>Budaya</b>             | Komposisi tim yang beragam dan pelatihan berkala tentang kesadaran budaya                     |

#### **Prinsip-Prinsip Keadilan Algoritma**

Keadilan algoritma bertujuan untuk memastikan bahwa model tidak menghasilkan hasil yang diskriminatif atau tidak adil berdasarkan atribut tertentu seperti ras, jenis kelamin, atau usia. Prinsip-prinsip utamanya meliputi alasan-alasan berikut:

- **Kesetaraan demografis.** Keadilan memastikan bahwa keputusan yang dibuat oleh algoritma memberikan hasil positif yang sama di berbagai kelompok.
- **Peluang yang disamakan.** Ini berarti bahwa tingkat kesalahan algoritma harus konsisten di seluruh kelompok. Ini memastikan bahwa tidak ada kelompok yang secara tidak proporsional terpengaruh oleh kesalahan yang mungkin dilakukan algoritma.
- **Keadilan individual.** Setiap individu yang serupa dalam aspek-aspek yang relevan harus diperlakukan secara serupa oleh algoritma. Ini adalah pendekatan granular, yang berfokus pada tingkat individu daripada tingkat kelompok.

#### **Pentingnya Transparansi dan Penjelasan Model**

Dengan meningkatnya model yang kompleks, terutama dalam pembelajaran mendalam, memahami mengapa suatu model membuat keputusan tertentu bisa jadi sulit. Namun, transparansi dan penjelasan sangat penting karena alasan-alasan berikut:

- **Kepercayaan.** Konsumen dan pemangku kepentingan lebih cenderung mempercayai suatu model jika mereka memahami cara kerjanya dan alasan di balik keputusannya.
- **Kepatuhan terhadap peraturan.** Peraturan seperti GDPR memberi individu hak untuk mendapatkan penjelasan ketika keputusan algoritmik memengaruhi mereka, sehingga transparansi menjadi hal yang tidak dapat dinegosiasikan.
- **Pertimbangan etis.** Model yang transparan memungkinkan pengawasan etis, memastikan bahwa model tersebut beroperasi dalam norma-norma masyarakat yang diterima.

#### **Alat dan Teknik untuk Pemodelan yang Adil dan Transparan**

Beberapa alat dan teknik dapat digunakan untuk membangun model yang adil dan transparan:

- **Pustaka interpretasi.** Alat-alat seperti LIME (penjelasan model agnostik yang dapat diinterpretasikan secara lokal) atau SHAP (penjelasan aditif Shapley) dapat membantu menjelaskan mengapa suatu model membuat prediksi tertentu, sehingga model yang kompleks menjadi lebih mudah diinterpretasikan.
- **Alat deteksi bias.** Platform seperti AI Fairness 360 dari IBM atau What-If Tool dari Google dapat membantu mengidentifikasi dan mengurangi bias dalam model.
- **Teknik regularisasi.** Teknik ini dapat digunakan dalam ML untuk menghindari overfitting, memastikan bahwa model tidak hanya menghafal data pelatihan tetapi juga melakukan generalisasi dengan baik ke data baru yang belum pernah dilihat sebelumnya.
- **Privasi diferensial.** Ini adalah teknik yang menambahkan noise ke data, memastikan privasi sambil tetap memungkinkan analisis yang bermakna.

Dalam ranah ilmu data pemasaran, pengakuan bias, komitmen terhadap keadilan algoritmik, dan penekanan pada transparansi adalah prinsip-prinsip yang tidak dapat dinegosiasikan. Seiring bisnis semakin bergantung pada data, mereka harus memastikan bahwa model yang mereka gunakan kuat dan berprinsip.

#### **Konsep Kunci dalam Bias, Keadilan, dan Transparansi**

Dalam ranah ilmu data pemasaran, beberapa konsep kunci sangat penting untuk memahami dan mengatasi bias, memastikan keadilan, dan mempromosikan transparansi.

- **Bias algoritmik.** Bias algoritmik terjadi ketika suatu algoritma secara sistematis menghasilkan hasil yang condong ke arah tertentu, seringkali mengakibatkan perlakuan tidak adil terhadap kelompok tertentu. Misalnya, algoritma persetujuan pinjaman mungkin bias terhadap demografi tertentu jika data yang digunakan untuk melatihnya secara historis bias. Dalam pemasaran, ini dapat terwujud sebagai kelompok pelanggan tertentu yang secara tidak adil dikecualikan dari kampanye promosi.
- **Metrik keadilan.** Ada berbagai metrik dan teknik yang dapat digunakan untuk menilai keadilan algoritma, seperti paritas demografis, kesempatan yang sama, dan keadilan individu. Metrik ini memastikan bahwa hasil algoritma adil di berbagai kelompok dan individu. Dalam konteks pemasaran, ini dapat melibatkan penilaian apakah algoritma rekomendasi produk memberikan rekomendasi yang adil di berbagai demografi.

- **Kemampuan menjelaskan dan menafsirkan.** Istilah-istilah ini merujuk pada kemampuan untuk memahami cara kerja internal suatu algoritma dan menjelaskan keputusannya. Kemampuan menjelaskan sangat penting dalam ilmu data pemasaran karena memungkinkan pemasar untuk memahami mengapa rekomendasi atau keputusan tertentu dibuat, yang dapat membantu meningkatkan kampanye di masa mendatang dan membangun kepercayaan pelanggan.
- **Algoritma yang menjaga privasi.** Dengan meningkatnya kekhawatiran tentang privasi data, telah terjadi peningkatan dalam pengembangan algoritma yang dapat bekerja dengan data terenkripsi atau anonim, seperti algoritma privasi diferensial. Algoritma ini dapat membantu pemasar menggunakan data pelanggan untuk mendapatkan wawasan sambil tetap menghormati privasi mereka.

### **Memastikan Keadilan dan Transparansi dalam Algoritma Personalisasi**

Untuk mengilustrasikan penerapan keadilan dan transparansi dalam ilmu data pemasaran, mari kita pertimbangkan sebuah perusahaan ritel online yang menggunakan algoritma personalisasi untuk merekomendasikan produk kepada pelanggannya. Algoritma tersebut awalnya dikembangkan menggunakan data pembelian historis, dan seiring waktu, perusahaan menyadari bahwa algoritma tersebut lebih sering merekomendasikan jenis produk tertentu kepada pria daripada wanita, yang berpotensi menunjukkan bias gender.

Untuk mengatasi hal ini, perusahaan memutuskan untuk mengambil pendekatan dua langkah:

1. **Deteksi dan mitigasi bias.** Perusahaan menggunakan metrik keadilan seperti paritas demografis untuk menilai apakah rekomendasi tersebut memang bias. Setelah mengkonfirmasi bias tersebut, mereka menerapkan algoritma koreksi bias untuk menyesuaikan rekomendasi dan memastikan distribusi yang lebih merata di berbagai kelompok demografis.
2. **Transparansi dan penjelasan.** Untuk meningkatkan transparansi, perusahaan mengadopsi pendekatan ML yang dapat diinterpretasikan, di mana prediksi model dapat dengan mudah dipahami oleh manusia. Hal ini memungkinkan mereka untuk menjelaskan kepada pelanggan mengapa mereka menerima rekomendasi tertentu, sehingga meningkatkan kepercayaan dan kepuasan pelanggan.

Contoh ini menyoroti pentingnya terus memantau dan menyesuaikan algoritma untuk memastikan keadilan, dan memberikan transparansi dalam pengambilan keputusan.

## **13.5 TREN YANG MUNCUL DAN MASA DEPAN ILMU DATA PEMASARAN**

### **Gambaran Umum Tren yang Muncul dan Masa Depan**

Bidang ilmu data pemasaran berkembang pesat, didorong oleh berbagai tren dan teknologi yang muncul. Berikut adalah beberapa area utama yang perlu diperhatikan (lihat Tabel 13.6):

- **AI dan ML.** AI dan ML bukan lagi sekadar kata kunci; keduanya menjadi bagian integral dari strategi pemasaran. Teknologi ini dapat membantu bisnis mengotomatiskan



proses, mendapatkan wawasan, dan meningkatkan personalisasi, yang mengarah pada peningkatan pengalaman pelanggan.

**Tabel 13.6** Tren yang Muncul dalam Ilmu Data Pemasaran Beserta Potensi Dampaknya pada Bisnis.

| Tren yang Sedang Berkembang                         | Potensi Dampak pada Bisnis                                                             |
|-----------------------------------------------------|----------------------------------------------------------------------------------------|
| Komputasi kuantum dalam analitik pemasaran          | Pemrosesan data yang jauh lebih cepat dan kemampuan analitik yang ditingkatkan         |
| Penggunaan realitas tertambah untuk uji coba produk | Peningkatan keterlibatan pelanggan dan pengalaman belanja online yang lebih baik       |
| Pengalaman pemasaran realitas virtual               | Pengalaman merek yang mendalam yang mengarah pada loyalitas pelanggan yang lebih dalam |
| Konvergensi ranah pemasaran offline dan online      | Pelacakan perjalanan pelanggan yang terpadu dan strategi pemasaran holistik            |
| Segmentasi pelanggan berbasis AI tingkat lanjut     | Kampanye pemasaran yang sangat personal dengan ROI yang lebih baik                     |

- **Platform data pelanggan.** Platform ini adalah perangkat lunak yang mengkonsolidasikan data pelanggan dari berbagai sumber ke dalam basis data terpadu, sehingga memudahkan pemasar untuk melakukan segmentasi audiens dan memberikan pengalaman yang dipersonalisasi.
- **Analisis prediktif.** Penggunaan analisis prediktif dalam pemasaran diperkirakan akan meningkat, memungkinkan bisnis untuk mengantisipasi perilaku dan tren pelanggan serta mengoptimalkan upaya pemasaran mereka sesuai dengan itu.
- **Teknologi peningkatan privasi.** Dengan semakin meningkatnya penekanan pada privasi data, teknologi yang membantu bisnis melindungi informasi pelanggan sambil tetap memperoleh wawasan darinya akan menjadi semakin penting.

Ke depan, pemasar perlu mengikuti perkembangan ini dan bersiap untuk mengintegrasikannya ke dalam strategi berbasis data mereka.

#### **Evolusi Analisis Pemasaran: Masa Lalu, Masa Kini, Masa Depan**

Analisis pemasaran telah mengalami transformasi yang cukup besar selama bertahun-tahun, beradaptasi dengan lanskap teknologi dan perilaku konsumen yang terus berkembang (lihat Gambar 13.4):

- **Masa Lalu.** Awal mula analisis pemasaran adalah periklanan respons langsung di media cetak, di mana pemasar dapat melacak tingkat respons terhadap iklan tertentu. Seiring kemajuan teknologi, basis data menjadi populer, memungkinkan penyimpanan dan pengambilan informasi pelanggan, menandai awal mula pemasaran basis data.
- **Masa Kini.** Ledakan platform digital dan media sosial telah menghadirkan sejumlah besar data kepada pemasar. Penggunaan ML dan analitik prediktif telah menjadi hal biasa, memungkinkan bisnis untuk menargetkan dan mempersonalisasi pesan mereka dengan presisi yang belum pernah terjadi sebelumnya.

- **Masa Depan.** Seiring dengan terus berkembangnya Internet of Things, akan ada lebih banyak titik kontak untuk mengumpulkan data. Pemasar diharapkan untuk memanfaatkan hal ini, ditambah dengan kekuatan AI, untuk mengantisipasi kebutuhan pelanggan bahkan sebelum kebutuhan tersebut muncul, menciptakan pendekatan pemasaran yang lebih proaktif.



**Gambar 13.4** Evolusi Analisis Pemasaran Selama Bertahun-tahun dan Potensi Arah Masa Depan.

#### **Peran Komputasi Kuantum dan AI Tingkat Lanjut dalam Pemasaran**

Komputasi kuantum, dengan potensinya untuk memproses sejumlah besar data secara eksponensial lebih cepat daripada komputer klasik, siap merevolusi analitik pemasaran. Tugas-tugas seperti optimasi strategi pemasaran, yang saat ini membutuhkan waktu komputasi yang substansial, dapat dikurangi secara signifikan menggunakan algoritma kuantum. Sementara itu, AI tingkat lanjut dapat mensintesis dan menganalisis pola kompleks dalam perilaku konsumen, memungkinkan strategi segmentasi dan personalisasi yang lebih canggih.

#### **Mengintegrasikan Realitas Virtual dan Realitas Tertambah dalam Strategi Pemasaran**

Teknologi realitas virtual (VR) dan realitas tertambah (AR) menawarkan media imersif kepada pemasar untuk melibatkan konsumen. Merek telah mulai memanfaatkan VR untuk menyediakan ruang pameran virtual, menawarkan pengalaman belanja taktil dari kenyamanan rumah konsumen. Namun, aplikasi AR di ponsel cerdas dapat melapisi informasi produk dan uji coba virtual, memperkaya pengalaman belanja di toko. Pengalaman integratif seperti itu tidak hanya baru tetapi juga dapat secara signifikan meningkatkan keterlibatan dan niat pembelian.

#### **Perpaduan yang Berkelanjutan antara Ranah Pemasaran Offline dan Online**

Rakyat pemasaran offline (fisik) dan online (digital) yang dulunya terpisah dengan cepat menyatu. Prevalensi strategi omnichannel merupakan contoh nyata dari penggabungan ini. Konsumen saat ini dapat memulai perjalanan belanja mereka secara online melalui media sosial atau platform e-commerce, mengunjungi toko fisik untuk mencoba produk, dan kemudian menyelesaikan pembelian mereka di aplikasi seluler. Oleh karena itu, pemasar ditugaskan untuk memastikan branding yang konsisten dan transisi yang mulus antar saluran ini. Perpaduan ini juga menghadirkan peluang untuk memanfaatkan data dari satu ranah

(misalnya, perilaku penelusuran online) untuk meningkatkan pengalaman di ranah lain (misalnya, penawaran personalisasi di toko).

Kesimpulannya, lanskap pemasaran terus berkembang pesat, didorong oleh kemajuan teknologi dan perubahan perilaku konsumen. Seiring munculnya alat dan strategi baru, pemasar harus tetap gesit, selalu beradaptasi untuk melayani dan melibatkan audiens target mereka dengan sebaik-baiknya.

### **Konsep Kunci dalam Tren yang Muncul dan Masa Depan**

Masa depan ilmu data pemasaran akan dibentuk oleh beberapa konsep kunci yang muncul dari tren saat ini:

- **AI yang Dapat Dijelaskan.** Seiring dengan semakin meluasnya penggunaan AI dalam pemasaran, terdapat peningkatan permintaan akan transparansi dan interpretasi sistem AI. AI yang Dapat Dijelaskan bertujuan untuk membuat proses pengambilan keputusan sistem AI dapat dipahami oleh manusia.
- **Privasi sejak tahap desain.** Dengan munculnya peraturan privasi data yang ketat, konsep mengintegrasikan privasi ke dalam desain sistem dan praktik sejak awal menjadi semakin penting. Konsep ini memastikan kepatuhan privasi dan perlindungan data sejak awal.
- **Pemasaran Omnichannel.** Masa depan pemasaran terletak pada integrasi pengalaman pelanggan di semua saluran. Pemasar semakin berupaya menciptakan pengalaman pelanggan yang mulus, baik pelanggan berbelanja online dari perangkat seluler, laptop, atau di toko fisik.
- **Hiper-personalisasi.** Personalisasi dalam pemasaran bukanlah hal baru, tetapi dengan kemajuan AI dan ML, hiper-personalisasi, yang melibatkan personalisasi waktu nyata berdasarkan perilaku pelanggan, menjadi mungkin.
- **AI yang Etis.** Karena AI dan ML semakin banyak diterapkan dalam pemasaran, semakin banyak pengakuan akan perlunya pertimbangan etis dalam penggunaannya, seperti keadilan, akuntabilitas, dan transparansi.

### **AI dalam Pemasaran yang Dipersonalisasi**

AI semakin banyak digunakan dalam upaya pemasaran yang dipersonalisasi. Contoh penerapan ini dapat dilihat pada kasus Netflix, platform streaming terkemuka. Netflix menggunakan AI untuk mempersonalisasi pengalaman pengguna secara ekstensif. Berdasarkan riwayat tontonan pengguna individu, Netflix merekomendasikan acara dan film yang kemungkinan besar akan ditonton dan dinikmati pengguna. Mesin rekomendasi yang dipersonalisasi ini didorong oleh algoritma ML canggih yang memprediksi preferensi pengguna berdasarkan perilaku masa lalu.

Sistem AI di Netflix menggunakan teknik yang disebut collaborative filtering. Teknik ini mempertimbangkan tidak hanya perilaku pengguna individu tetapi juga perilaku pengguna lain dengan preferensi serupa. Misalnya, jika dua pengguna telah menonton dan menyukai serangkaian film yang serupa, dan salah satu dari mereka menonton dan menyukai film baru, sistem akan merekomendasikan film baru ini kepada pengguna lain.

Penggunaan AI oleh Netflix dalam pemasaran yang dipersonalisasi sangat sukses. Mesin rekomendasi mereka, yang merupakan komponen penting dari strategi pemasaran mereka, diperkirakan menghemat biaya sebesar \$1 miliar per tahun dengan mengurangi tingkat pembatalan langganan.

### **13.6 KESIMPULAN**

Di era digital saat ini, perpaduan antara pemasaran dan ilmu data memberikan peluang yang tak tertandingi bagi bisnis untuk memahami konsumen mereka dan mempersonalisasi pengalaman. Namun, dengan kekuatan besar datang tanggung jawab besar. Bab 13 telah menggarisbawahi pentingnya pertimbangan etika, masalah privasi, dan kepatuhan terhadap peraturan dalam ranah ilmu data pemasaran.

Pada intinya, praktik etika dalam ilmu data pemasaran bukan hanya tentang kepatuhan, tetapi tentang membangun kepercayaan. Konsumen, lebih dari sebelumnya, menyadari jejak digital mereka. Saat bisnis memanfaatkan data untuk mendapatkan wawasan, ada keharusan moral untuk memastikan bahwa data tersebut tidak disalahgunakan, disalahartikan, atau ditangani dengan tidak benar. Pelanggaran etika dapat merusak reputasi merek, kepercayaan konsumen, dan integritas ekosistem yang lebih luas secara permanen.

Privasi, yang terkait erat dengan etika, telah muncul sebagai landasan praktik pemasaran modern. Keseimbangan yang rumit antara personalisasi dan privasi adalah tali tipis yang harus dilalui pemasar dengan hati-hati. Memastikan anonimitas data, mematuhi prinsip minimalisasi data, dan menjaga transparansi dalam pengumpulan dan penggunaan data adalah hal yang tidak dapat ditawar dalam lanskap saat ini.

Regulasi, seperti GDPR dan CCPA, meskipun dipandang oleh sebagian orang sebagai ketat, merupakan lambang dorongan masyarakat untuk lingkungan data yang lebih terkontrol, transparan, dan berpusat pada konsumen. Regulasi ini menggarisbawahi hak individu atas data mereka, memaksa bisnis untuk mengadopsi pendekatan yang lebih hormat dan hati-hati terhadap pengumpulan dan penggunaan data.

Selain itu, seiring dengan kemajuan teknologi yang tak henti-hentinya—menghadirkan alat-alat seperti AI dan komputasi kuantum—pertimbangan etika, privasi, dan regulasi akan semakin penting. Tanggung jawab ada pada pemasar, ilmuwan data, dan bisnis saat ini dan di masa mendatang untuk bersikap proaktif, terus memperbarui pengetahuan mereka, meninjau kembali praktik mereka, dan terlibat dalam dialog terbuka tentang tantangan dan peluang yang terus berkembang.

Pada intinya, meskipun penggabungan pemasaran dan ilmu data menawarkan peluang yang menjanjikan bagi bisnis, sangat penting untuk menavigasi domain ini dengan kompas yang berlandaskan etika, menghormati privasi, dan pemahaman yang mendalam tentang peraturan. Hanya dengan melakukan hal tersebut, bisnis dapat benar-benar memanfaatkan kekuatan transformatif ilmu data pemasaran secara berkelanjutan dan berpusat pada konsumen.

## DAFTAR PUSTAKA

- Ali, M. (2022). Chapter-20 Digital Marketing Mastery: Igniting Growth for Your Firm. *Marketing Mastery: Strategies for Captivating Audiences*, 225.
- Allil, K. (2024). Integrating AI-driven marketing analytics techniques into the classroom: pedagogical strategies for enhancing student engagement and future business success. *Journal of Marketing Analytics*, 12(2), 142-168.
- Almgerbi, M., De Mauro, A., Kahlawi, A., & Poggioni, V. (2022). A systematic review of data analytics job requirements and online-courses. *Journal of Computer Information Systems*, 62(2), 422-434.
- Arora, M. (2023). Encapsulating role of persuasion and skill development in marketing communication for brand building: A perspective. In *International handbook of skill, education, learning, and research development in tourism and hospitality* (pp. 1-17). Singapore: Springer Nature Singapore.
- Brown, I. (2024). *Mastering marketing data science: a comprehensive guide for today's marketers*. John Wiley & Sons.
- Charan, A. (2023). *Marketing Analytics Practitioner's Guide, The-Volume 2: Product, Advertising, Packaging, Biometrics, Price And Promotion*. World Scientific.
- Charles, V., Garg, P., Gupta, N., & Agarwal, M. (2023). Data analytics and business intelligence. *Data Analytics and Business Intelligence*, 1-56.
- Confetto, M. G., Conte, F., Vollero, A., & Covucci, C. (2020). From dual marketing to marketing 4.0: the role played by digital technology and the internet.
- Devda, P., Shah, S., & Vasavada, M. (2023). Analytical Crm for Google Edge-Data Mining Framework With Reference To Pharmaceuticals Industry in India. *International Journal of Management, Public Policy and Research*, 2(1), 40-61.
- Dobrokhoto, V. (2023). *The art of after-sales marketing: Turning satisfied customers into loyal brand advocates*. Virtual Inspiration.
- Elhajjar, S. (2024). The current and future state of the marketing management profession. *Journal of Marketing Theory and Practice*, 32(2), 233-250.
- Ewel, J. (2020). *The six disciplines of agile marketing: Proven practices for more effective marketing and better business results*. John Wiley & Sons.

- Galety, M. G., Claver, J. H., Sriharsha, A. V., Vajjhala, N. R., & Natarajan, A. K. (Eds.). (2024). *Data Analytics and AI for Quantitative Risk Assessment and Financial Computation*. IGI Global.
- Glenister, G. (2024). *Influencer marketing strategy: How influencers can help grow your business*. Kogan Page Publishers.
- Gonzalez, L. R., & Stubberfield, A. (2024). *Cracking the Data Science Interview: Unlock insider tips from industry experts to master the data science field*. Packt Publishing Ltd.
- Grewal, D., Guha, A., Beccacece Satornino, C., & Becker, M. (2025). The future of marketing and marketing education. *Journal of Marketing Education*, 47(1), 61-77.
- Gupta, R., & Madan, S. (2024). *Digital Marketing: The science and magic of digital marketing can help you become a successful marketing professional*. BPB Publications.
- Holmlund, M., Van Vaerenbergh, Y., Ciuchita, R., Ravald, A., Sarantopoulos, P., Ordenes, F. V., & Zaki, M. (2020). Customer experience management in the age of big data analytics: A strategic framework. *Journal of business research*, 116, 356-365.
- Hossain, Q., Yasmin, F., Biswas, T. R., & Asha, N. B. (2024). Integration of Big Data Analytics in Management Information Systems for Business Intelligence. *Saudi J Bus Manag Stud*, 9(9), 192-203.
- Jarboe, G., Bailey, M., & Stebbins, M. (2023). *Digital Marketing Fundamentals: OMCP's Official Guide to OMCA Certification*. John Wiley & Sons.
- Johnson, D. S., Sihi, D., & Muzellec, L. (2021, September). Implementing big data analytics in marketing departments: Mixing organic and administered approaches to increase data-driven decision making. In *Informatics* (Vol. 8, No. 4, p. 66). MDPI.
- Kabir, M. H., Sultana, S., Hossain, M. M., & Islam, S. M. A. (2025). The impact of digital marketing strategies on consumer behavior: A comprehensive review. *Business and Social Sciences*, 3(1), 1-8.
- Kalla, D. (2024). *Improving E-Commerce Organization Performance Using Big Data Analytics and Artificial Intelligence* (Doctoral dissertation, Colorado Technical University).
- Kumar, V. M., Vijayaraghavan, P., Meshram, V. V., Sharma, M. K., MS, N., & Kumar, R. (2023, November). Transforming Data Analysis through AI-Powered Data Science. In *2023 2nd International Conference on Futuristic Technologies (INCOFT)* (pp. 1-5). IEEE.
- Laverie, D., Humphrey, W., Manis, K. T., & Freberg, K. (2020). The digital era has changed marketing: A guide to using industry certifications and exploration of student perceptions of effectiveness. *Marketing Education Review*, 30(1), 57-80.

- Lieberman, S. (2021). Commentary: managing human experience as a core marketing capability. *Journal of Marketing*, 85(1), 219-222.
- Maravilhas, S., & Ladeira, R. (Eds.). (2025). *Impact of Digital Transformation on Business Growth and Performance*. IGI Global.
- Moon, S., & Iacobucci, D. (2022). Social media analytics and its applications in marketing. *Foundations and Trends® in Marketing*, 15(4), 213-292.
- Muhammad, N. (2025). The impact of Industry 4.0 on digital marketing: Leveraging emerging technologies for business growth. *International journal of academic research in business & social sciences*.
- Palmatier, R. W., Petersen, J. A., & Germann, F. (2022). *Marketing Analytics: Based on First Principles*. Bloomsbury Publishing.
- Rachmad, Y. E. (2025). The Viral Shift: Harnessing the Power of Gimmicks in Changing Consumer Behavior. *United Nations Economic and Social Council*.
- Rajamannar, R. (2021). *Quantum marketing: mastering the new marketing mindset for tomorrow's consumers*. HarperCollins Leadership.
- Sachdeva, C., Grover, V., Kaur, A., & Gangwar, V. P. (2025). Mastering Data Transformation: Preparing Marketing Data for Actionable Insights.
- Salari, N. (2026). *Advanced marketing analytics: transform data into insights*. Kogan Page Publishers.
- Sedkaoui, S. (2024). *Automated data analytics: Combining human creativity and AI power using ChatGPT*. John Wiley & Sons.
- Sharma, S. (2023). *Managing Product Reviews: A Comprehensive Guide for Brands and Businesses*. CSMFL Publications.
- Shemshaki, M. (2024). *Data-Driven Digital Marketing The Art and Science of Intelligent Decision-Making*. Milad Shemshaki.
- Singh, B., & Kaunert, C. (2024). Future of digital marketing: Hyper-personalized customer dynamic experience with AI-based predictive models. In *Revolutionizing the AI-digital landscape* (pp. 189-203). Productivity Press.
- Smith, P. R., & Zook, Z. (2024). *Marketing communications: Integrating online and offline, customer engagement and digital technologies*. Kogan Page Publishers.

- Sotamaa, O., Tyni, H., & Myöhänen, T. (2024). 'Even if the algorithm is a terrible workmate, you just need to learn to live with it': Perceptions of data analytics among game industry professionals. *European Journal of Cultural Studies*, 27(4), 558-576.
- Sweeney, B. (2022). *Digital marketing QuickStart guide: The simplified beginner's guide to developing a scalable online strategy, finding your customers, and profitably growing your business*. ClydeBank Media LLC.
- Tanveer, M., Khan, N., & Ahmad, A. R. (2021, April). AI support marketing: Understanding the customer journey towards the business development. In *2021 1st international conference on artificial intelligence and data analytics (CAIDA)* (pp. 144-150). IEEE.
- Venkatesan, R., & Lecinski, J. (2021). *The AI marketing canvas: A five-stage road map to implementing artificial intelligence in marketing*. Stanford University Press.
- Verhoef, P. C., Kooge, E., Walk, N., & Wieringa, J. E. (2021). *Creating value with data analytics in marketing: mastering data science*. Routledge.
- Vesterinen, M., Mero, J., & Skippari, M. (2025). Big data analytics capability, marketing agility, and firm performance: a conceptual framework. *Journal of Marketing Theory and Practice*, 33(2), 310-330.
- Wongmonta, S. (2021). Marketing 5.0: The era of technology for humanity with a collaboration of humans and machines. *Nimitmai Review Journal*, 4(1), 83-97.
- Yu, B. (2023). Deep learning applications for interactive marketing in the contemporary digital age. *The Palgrave Handbook of Interactive Marketing*, 705-728.



## LAMPIRAN: KUNCI LATIHAN SOAL

### LATIHAN 2.1

Output kode Python untuk Latihan 2.1 adalah sebagai berikut.

DataFrame sekarang mencerminkan perubahan yang dilakukan selama proses pembersihan dan transformasi data:

```
Display the first few rows of the dataset data.head()
```

| ID Pelanggan | Usia  | Lama Berlangganan (hari) | Pengeluaran Bulanan (\$) | Tanggal Login Terakhir | Skor Umpan Balik | Pengeluaran Bulanan per Hari |
|--------------|-------|--------------------------|--------------------------|------------------------|------------------|------------------------------|
| 1            | 62.00 | 260                      | 112.89                   | 2022-01-01             | 5.0              | 0.434                        |
| 2            | 42.95 | 180                      | 288.36                   | 2022-01-02             | 5.0              | 1.602                        |
| 3            | 18.00 | 60                       | 72.97                    | 2022-01-03             | 1.0              | 1.216                        |
| 4            | 21.00 | 285                      | 199.07                   | 2022-01-04             | 3.0              | 0.698                        |
| 5            | 21.00 | 49                       | 194.15                   | 2022-01-05             | 5.0              | 3.962                        |

Berikut ringkasan singkat transformasi yang dilakukan:

- **Penanganan Nilai yang Hilang:** Nilai yang hilang pada kolom 'Usia', 'Pengeluaran Bulanan (\$)', dan 'Skor Umpan Balik' diisi dengan nilai rata-rata, median, dan modus dari kolom masing-masing.
- **Konversi Tanggal:** 'Tanggal Login Terakhir' dikonversi dari string menjadi objek DateTime untuk penanganan operasi terkait tanggal yang lebih baik.
- **Pembuatan Fitur Baru:** Kolom baru, 'Pengeluaran Bulanan per Hari', ditambahkan ke DataFrame, dihitung dengan membagi 'Pengeluaran Bulanan (\$)' dengan 'Lama Berlangganan (hari)'. Fitur baru ini memberikan wawasan tambahan tentang perilaku pengeluaran pelanggan.

### LATIHAN 2.2

Output kode Python untuk Latihan 2.2 adalah sebagai berikut:

1. **Hasil Agregasi Data:** Tabel ini menunjukkan data agregat berdasarkan wilayah, termasuk pengeluaran bulanan rata-rata dan total frekuensi pembelian per wilayah.

| Wilayah | Rata-rata Pengeluaran Bulanan | Total Frekuensi Pembelian |
|---------|-------------------------------|---------------------------|
| Timur   | 273.57                        | 2,423                     |
| Utara   | 264.17                        | 2,645                     |
| Selatan | 290.11                        | 2,712                     |
| Barat   | 265.98                        | 2,464                     |

2. Tabel tersebut memberikan ringkasan perilaku konsumen di berbagai wilayah, menyoroti perbedaan dalam pengeluaran dan frekuensi pembelian.
3. **Hasil Analisis Komponen Utama (PCA):** Tabel berikut menunjukkan lima baris pertama dari dataset setelah menerapkan PCA, mengurangi data menjadi dua komponen utama.

| Komponen Utama 1 | Komponen Utama 2 |
|------------------|------------------|
| -0.559           | -0.955           |
| 0.364            | 1.279            |
| -0.096           | -2.208           |
| -0.313           | -0.413           |
| -0.354           | 0.155            |

4. Kedua komponen utama ini merepresentasikan dataset yang telah ditransformasi dalam ruang dua dimensi. Transformasi ini membantu dalam memvisualisasikan dan menganalisis data dalam bentuk yang lebih ringkas, sehingga memudahkan identifikasi pola atau kluster.

### LATIHAN 3.1

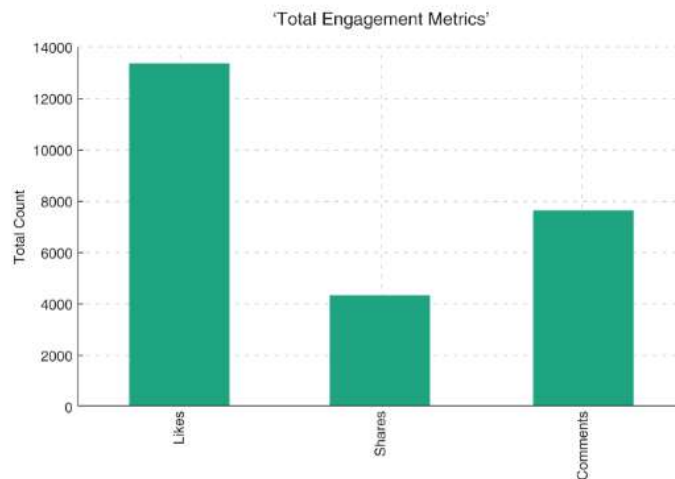
Statistik deskriptif untuk dataset adalah sebagai berikut:

- **'Pengeluaran Iklan':**
  - Rata-rata: Rp12.085.800
  - Deviasi Standar: Rp4.514.500
  - Minimum: Rp5.082.800
  - Maksimum: Rp19.803.300
- **'Tayangan':**
  - Rata-rata: 3096,93
  - Deviasi Standar: 1104,78
- **'Klik':**
  - Rata-rata: 123,28
  - Deviasi Standar: 48,49
- **'Konversi':**
  - Rata-rata: 27,97
  - Deviasi Standar: 13,48
- **Metrik Keterlibatan ('Suka', 'Bagikan', 'Komentar'):**
  - Suka: Rata-rata – 148,49, Deviasi Standar –83,67
  - Bagikan: Rata-rata – 48,63, Deviasi Standar –30,86
  - Komentar: Rata-rata – 85,52, Deviasi Standar –46,24

### Hasil Visualisasi

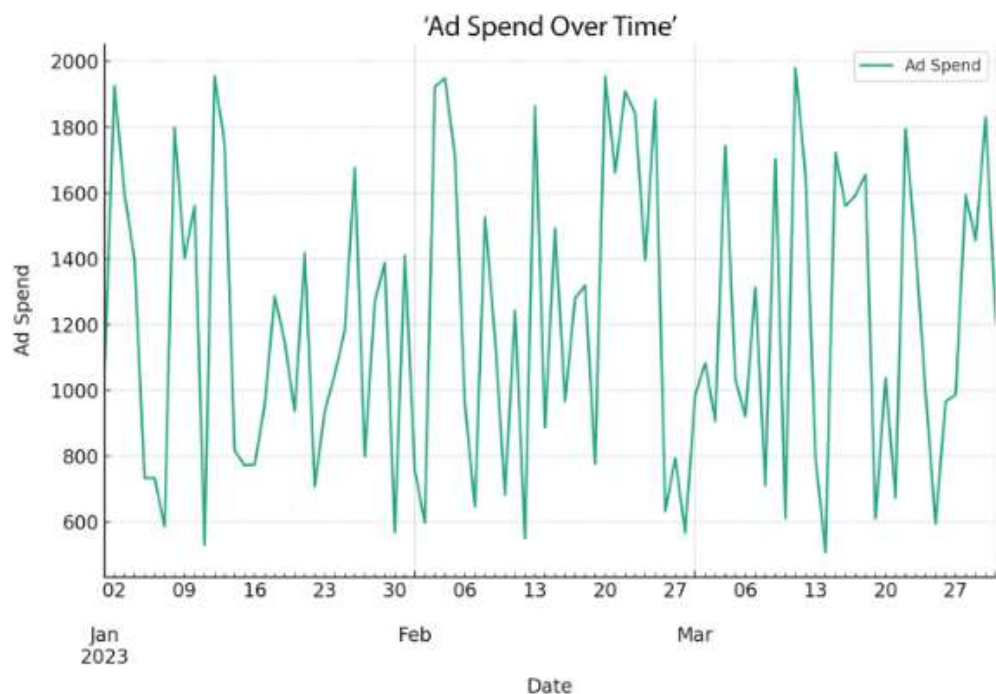
1. **Bagan Batang—'Total Metrik Keterlibatan':**

- Bagan ini menunjukkan jumlah total untuk 'Suka', 'Bagikan', dan 'Komentar'. 'Suka' adalah yang tertinggi, diikuti oleh 'Komentar' dan 'Bagikan'.



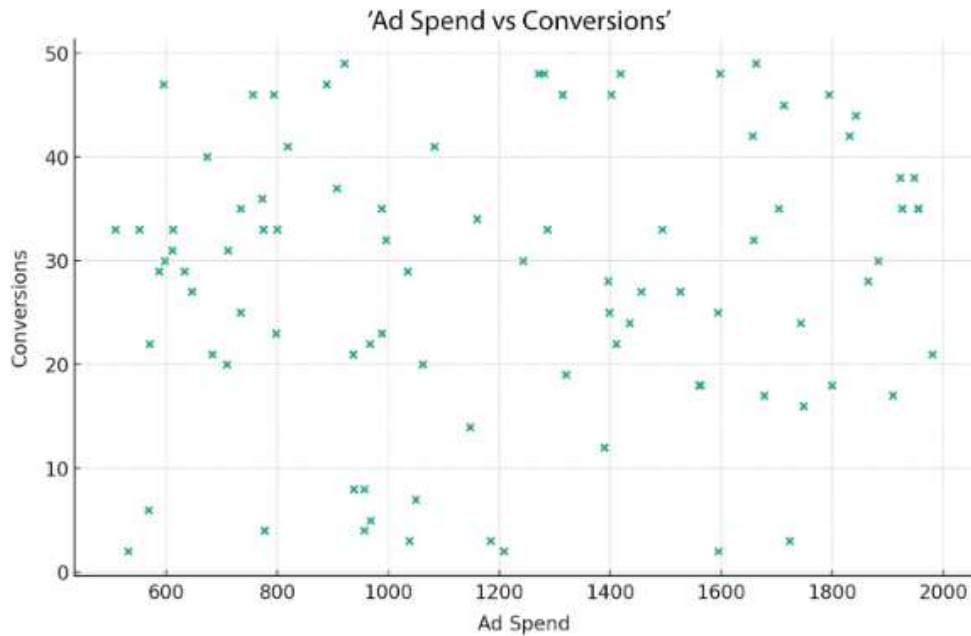
## 2. Grafik Garis—'Pengeluaran Iklan dari Waktu ke Waktu':

- Grafik garis menampilkan tren 'Pengeluaran Iklan' selama periode tiga bulan. Ini membantu memvisualisasikan fluktuasi pengeluaran dari waktu ke waktu.



## 3. Scatterplot—'Pengeluaran Iklan' vs 'Konversi':

- Scatterplot ini menggambarkan hubungan antara 'Pengeluaran Iklan' dan 'Konversi'. Setiap titik mewakili data satu hari, menunjukkan bagaimana konversi bervariasi dengan berbagai tingkat pengeluaran iklan.



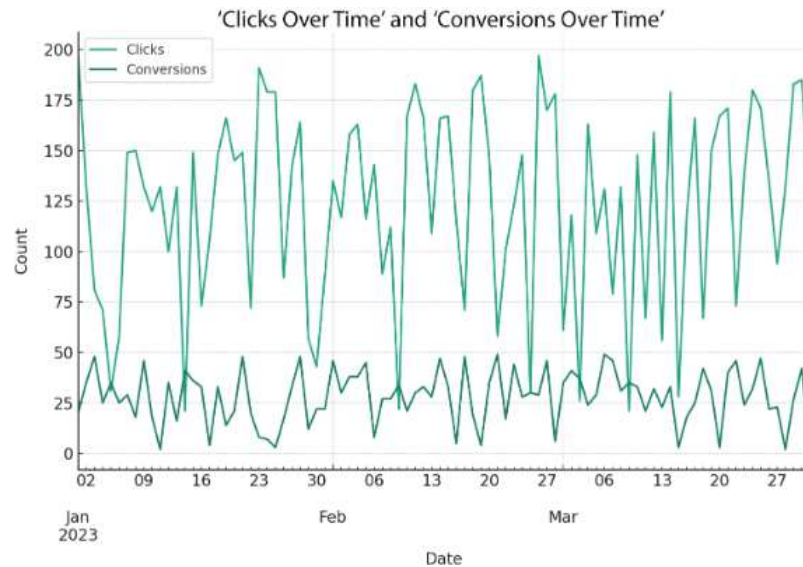
### Interpretasi

- Statistik deskriptif memberikan gambaran komprehensif tentang kecenderungan utama dan variabilitas dalam data pemasaran.
- Bagan batang menunjukkan bahwa 'Suka' adalah metrik keterlibatan yang paling signifikan untuk kampanye ini.
- Bagan garis untuk 'Pengeluaran Iklan' menunjukkan variabilitas dari waktu ke waktu, yang mungkin disebabkan oleh perubahan strategi pemasaran atau kondisi pasar.
- Diagram sebaran dapat diperiksa untuk pola korelasi antara 'Pengeluaran Iklan' dan 'Konversi'. Misalnya, tren positif akan menunjukkan bahwa pengeluaran iklan yang lebih tinggi berpotensi menghasilkan lebih banyak konversi, yang merupakan wawasan penting untuk alokasi anggaran dalam kampanye pemasaran.

### LATIHAN 3.2

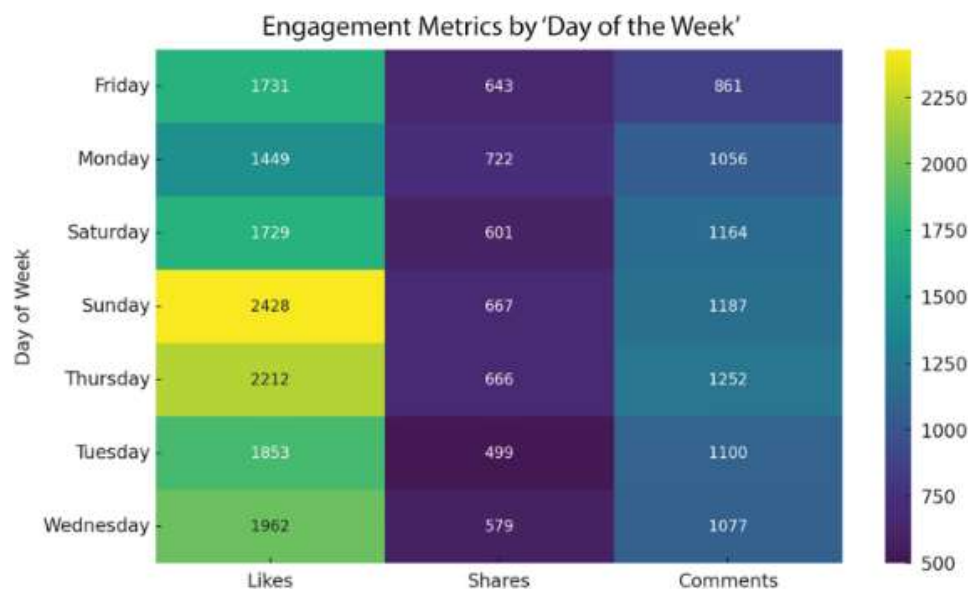
#### 1. Grafik Garis—'Klik dari Waktu ke Waktu' dan 'Konversi dari Waktu ke Waktu':

- Grafik ini menampilkan tren 'Klik' dan 'Konversi' selama periode tiga bulan, memungkinkan kita untuk mengamati bagaimana metrik ini telah berubah dari waktu ke waktu dan untuk mengidentifikasi pola atau anomali apa pun.



## 2. Peta Panas—Metrik Keterlibatan berdasarkan ‘Hari dalam Seminggu’:

- Peta panas memvisualisasikan jumlah total ‘Suka’, ‘Bagikan’, dan ‘Komentar’ untuk setiap hari dalam seminggu. Ini dapat membantu mengidentifikasi hari mana yang cenderung memiliki keterlibatan lebih tinggi, yang berpotensi memberikan informasi untuk penjadwalan konten dan strategi pemasaran.



## 3. Visualisasi Gaya Dasbor—Beberapa KPI:

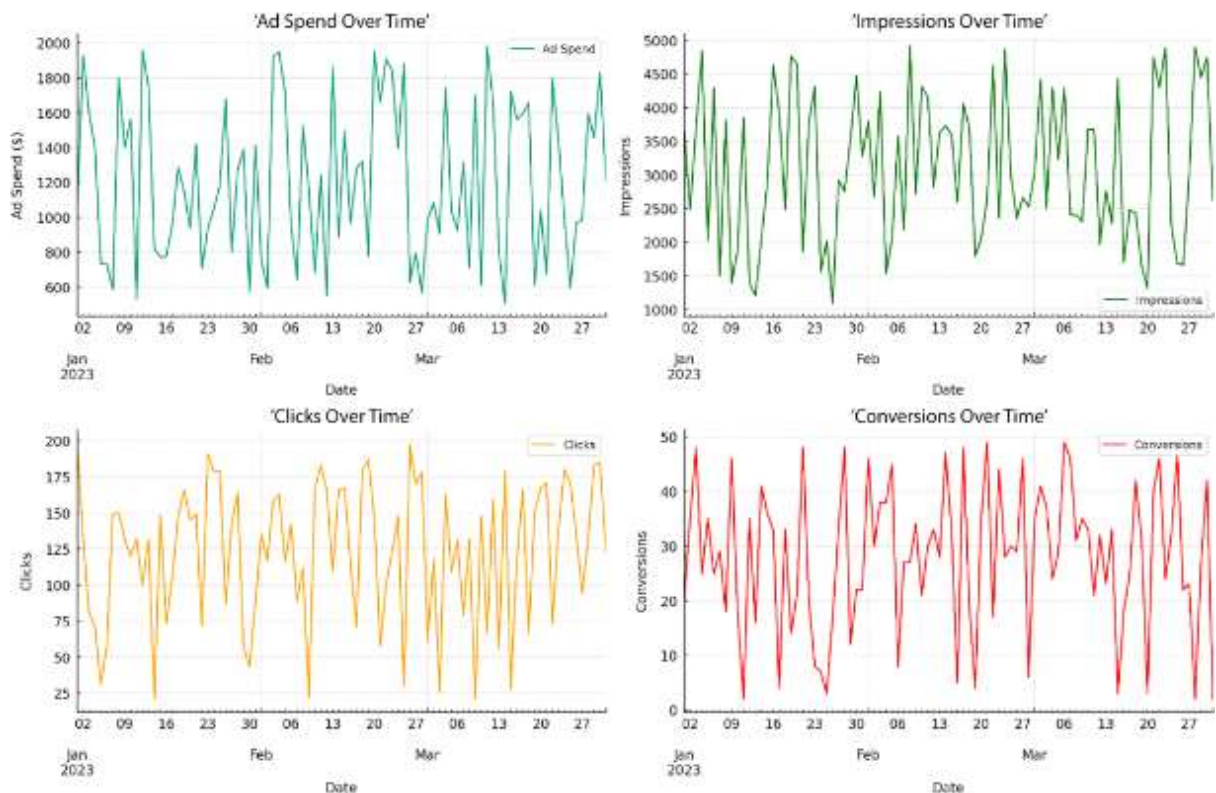
- Kisi-kisi plot 2x2 menampilkan ‘Pengeluaran Iklan’, ‘Tayangan dari Waktu ke Waktu’, ‘Klik dari Waktu ke Waktu’, dan ‘Konversi dari Waktu ke Waktu’. Tampilan komprehensif ini membantu untuk dengan cepat menilai kinerja berbagai aspek kampanye pemasaran, menyoroti tren dan area yang mungkin perlu diperhatikan.

### Interpretasi

- Dari analisis deret waktu, dimungkinkan untuk memahami korelasi antara berbagai aktivitas, seperti dampak pengeluaran iklan pada ‘Klik’ dan ‘Konversi’.

- Peta panas memberikan wawasan tentang efektivitas keterlibatan media sosial di berbagai hari, yang dapat sangat penting untuk perencanaan dan pengoptimalan strategi pemasaran media sosial.
- Visualisasi gaya dasbor menawarkan pandangan holistik tentang kinerja kampanye, memungkinkan penilaian cepat tentang bagaimana berbagai KPI telah berkembang dari waktu ke waktu.

Visualisasi ini, yang diambil dari kumpulan data, memberikan pemahaman yang lebih dalam tentang tren pasar dan kinerja kampanye, yang sangat penting untuk pengambilan keputusan yang tepat dalam pemasaran.



### LATIHAN 5.1

Model 'LogisticRegression' dilatih pada data pelanggan yang berhenti berlangganan dan dievaluasi pada set data uji. Laporan klasifikasi menyediakan beberapa metrik untuk menilai kinerja model:

| STDOUT/STDERR |         |           |         |          |
|---------------|---------|-----------|---------|----------|
|               | presisi | perolehan | skor-f1 | dukungan |
| 0             | 0.8     | 0.91      | 0.85    | 93       |
| 1             | 0.91    | 0.8       | 0.86    | 107      |
|               |         |           |         |          |

|                      |      |      |      |     |
|----------------------|------|------|------|-----|
| akurasi              |      |      | 0.85 | 200 |
| rata-rata makro      | 0.86 | 0.86 | 0.85 | 200 |
| rata-rata tertimbang | 0.86 | 0.85 | 0.86 | 200 |

- **Presisi:** Ini adalah rasio observasi positif yang diprediksi dengan benar terhadap total prediksi positif. Presisi tinggi berkaitan dengan tingkat positif palsu yang rendah. Kami memiliki nilai presisi 0,80 untuk kelas 0 (tidak akan berhenti berlangganan) dan 0,91 untuk kelas 1 (akan berhenti berlangganan), yang menunjukkan bahwa model lebih presisi dalam memprediksi pelanggan yang akan berhenti berlangganan daripada mereka yang tidak akan berhenti berlangganan.
- **Recall (Sensitivitas):** Ini adalah rasio observasi positif yang diprediksi dengan benar terhadap semua observasi dalam kelas sebenarnya. Recall adalah 0,91 untuk kelas 0 dan 0,80 untuk kelas 1, menunjukkan bahwa model lebih sensitif dalam memprediksi pelanggan yang tidak akan berhenti berlangganan dengan benar daripada pelanggan yang akan berhenti berlangganan.
- **Skor F1:** Ini adalah rata-rata harmonik tertimbang dari presisi dan recall. Skor F1 adalah 0,85 untuk kelas 0 dan 0,86 untuk kelas 1, menunjukkan bahwa model tersebut kuat dalam kinerja prediktifnya untuk kedua kelas.
- **Support:** Ini adalah jumlah kejadian sebenarnya dari kelas tersebut dalam dataset yang ditentukan. Untuk pelanggan yang tidak berhenti berlangganan (kelas 0), terdapat 93 instance, dan untuk pelanggan yang berhenti berlangganan (kelas 1), terdapat 107 instance.
- **Akurasi:** Model ini memiliki akurasi keseluruhan 0,85, yang berarti model ini memprediksi status berhenti berlangganan dengan benar 85% dari waktu pada set pengujian.
- **Rata-rata Makro:** Ini adalah rata-rata presisi, recall, dan skor F1 antar kelas. Rata-rata makro tidak memperhitungkan ketidakseimbangan kelas, yang sesuai di sini karena kelas-kelasnya seimbang.
- **Rata-rata Tertimbang:** Ini adalah rata-rata presisi, recall, dan skor F1 antar kelas yang diberi bobot berdasarkan jumlah instance di setiap kelas. Ini memberi kita ukuran yang lebih baik tentang kualitas sebenarnya dari pengklasifikasi, terutama ketika ada ketidakseimbangan kelas, yang bukan masalah signifikan dalam dataset ini.

Secara keseluruhan, dengan skor F1 sekitar 0,85 untuk kedua kelas, model tampaknya berkinerja baik pada dataset ini, yang menunjukkan bahwa model ini dapat menjadi titik awal yang baik untuk membuat prediksi dalam skenario dunia nyata.

## LATIHAN 5.2

Kesalahan kuadrat rata-rata yang dihitung adalah 0,01064, yang merupakan ukuran akurasi model. MSE yang lebih rendah menunjukkan kesesuaian model yang lebih baik dengan data. Dengan MSE yang rendah, kita dapat menyimpulkan bahwa model kita telah berkinerja baik pada dataset ini. Latihan ini mendemonstrasikan proses pembuatan dan evaluasi model

prediktif, yang merupakan aspek fundamental dari ilmu data dalam pemasaran dan banyak bidang lainnya. MSE yang kecil menunjukkan bahwa prediksi model sangat dekat dengan angka penjualan aktual, menjadikannya alat yang berpotensi berguna dalam konteks pemasaran dunia nyata.

### LATIHAN 6.1

Berikut adalah hasil setelah menjalankan pembahasan sentimen pada postingan baru ini:

| Postingan                                                                        | Pembahasan Sentimen |
|----------------------------------------------------------------------------------|---------------------|
| Saya suka smartphone baru ini. Kameranya luar biasa!                             | Positif             |
| Sangat tidak puas dengan layanan pelanggan. Pengalaman yang sangat mengecewakan. | Negatif             |
| Ini hanya produk biasa. Tidak ada yang istimewa.                                 | Positif             |
| Benar-benar fantastis! Tidak bisa meminta yang lebih baik.                       | Positif             |
| Pembelian terburuk yang pernah ada. Sangat menyesal membelinya.                  | Negatif             |

Seperti yang Anda lihat, setiap postingan telah dipembahasan oleh **TextBlob**, dan sentimen ('Positif', 'Negatif', atau 'Netral') telah ditetapkan berdasarkan isi postingan tersebut. Misalnya, postingan yang menyatakan kepuasan atau kebahagiaan diberi label 'Positif', sedangkan postingan yang menyatakan ketidakpuasan atau kekecewaan diberi label 'Negatif'. Ini menunjukkan bagaimana pembahasan sentimen dapat digunakan untuk mengkategorikan data teks berdasarkan sentimen yang diungkapkan di dalamnya.

### LATIHAN 6.2

Berikut adalah output dari evaluasi model:

| Kategori             | Presisi (Precision) | Recall | Skor F1 | Dukungan |
|----------------------|---------------------|--------|---------|----------|
| Pakaian              | 0,00                | 0,00   | 0,00    | 55       |
| Elektronik           | 0,00                | 0,00   | 0,00    | 63       |
| Makanan              | 0,41                | 1,00   | 0,58    | 82       |
| Akurasi              |                     |        | 0,41    | 200      |
| Rata-rata Makro      | 0,14                | 0,333  | 0,19    | 200      |
| Rata-rata Tertimbang | 0,17                | 0,310  | 0,24    | 200      |

#### Pembahasan Hasil:

- Kinerja model sangat bervariasi di berbagai kategori.
- Untuk kategori 'Pakaian' dan 'Elektronik', model menunjukkan presisi dan recall 0,00, yang mengindikasikan bahwa model gagal mengklasifikasikan ulasan dengan benar di kategori-kategori ini.



- Untuk kategori 'Makanan', model menunjukkan recall 1,00 tetapi presisi yang lebih rendah. Ini berarti model mengklasifikasikan sebagian besar ulasan sebagai 'Makanan', kemungkinan termasuk banyak ulasan yang termasuk dalam kategori lain.

Akurasi keseluruhan model adalah 0,41, yang menunjukkan bahwa hanya sekitar 41% klasifikasi yang benar.

#### **Interpretasi:**

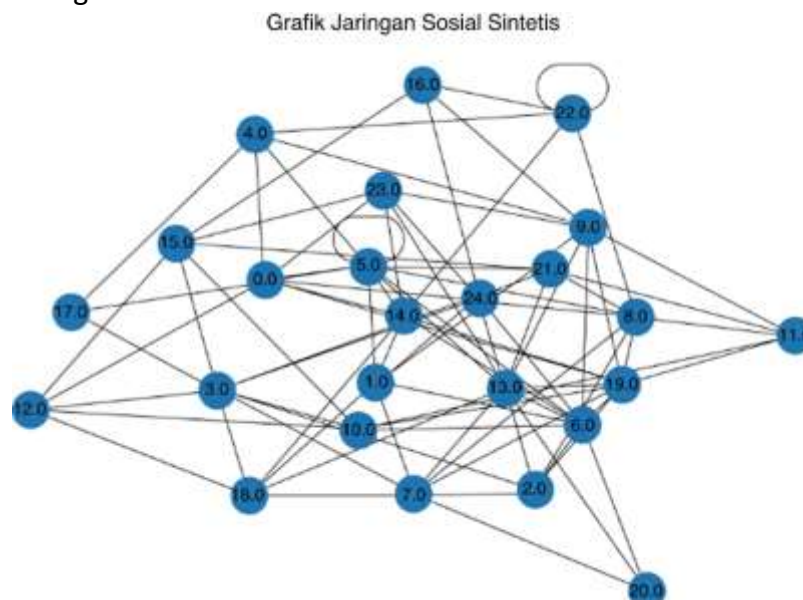
- Kecenderungan model untuk mengklasifikasikan sebagian besar ulasan sebagai 'Makanan' menunjukkan potensi bias. Hal ini mungkin disebabkan oleh data pelatihan yang tidak seimbang atau keterbatasan model Naive Bayes untuk tugas khusus ini.
- Peringatan tentang presisi dan skor F1 yang tidak terdefinisi dengan baik berasal dari kegagalan model untuk membuat prediksi yang benar untuk kategori 'Pakaian' dan 'Elektronik'.
- Contoh ini menyoroti aplikasi dasar klasifikasi teks. Perbaikan potensial dapat mencakup peningkatan pra-pemrosesan data, penggunaan model yang lebih canggih, atau penyempurnaan parameter model saat ini untuk hasil yang lebih seimbang dan akurat.

Latihan ini sengaja mendemonstrasikan aplikasi dasar klasifikasi teks. Perbaikan mungkin mencakup pra-pemrosesan data yang lebih baik, penggunaan model yang lebih canggih, atau penyempurnaan parameter model saat ini untuk mencapai hasil yang lebih seimbang dan akurat.

### **LATIHAN 7.1**

#### **1. Visualisasi Grafik Jaringan Sosial:**

- Grafik jaringan sosial berhasil dibuat dan divisualisasikan. Grafik ini mewakili jaringan sosial dengan node (pengguna) dan edge (hubungan). Catatan: Karena edge antar node dihasilkan secara acak, grafik ini akan berbeda dari pembuatan di masa mendatang.



#### **2. Lima Pengguna Teratas Berdasarkan Ukuran Sentralitas yang Berbeda:**

- **Sentralitas Derajat:**
  - Pengguna 5: 0,4583
  - Pengguna 6: 0,4167
  - Pengguna 13: 0,4167
  - Pengguna 19: 0,4167
  - Pengguna 3: 0,3333
- **Sentralitas Antara:**
  - Pengguna 13: 0,0713
  - Pengguna 6: 0,0685
  - Pengguna 9: 0,0629
  - Pengguna 5: 0,0590
  - Pengguna 19: 0,0588
- **Sentralitas Eigenvektor:**
  - Pengguna 6: 0,3079
  - Pengguna 5: 0,3020
  - Pengguna 19: 0,2995
  - Pengguna 13: 0,2982
  - Pengguna 14: 0,2483

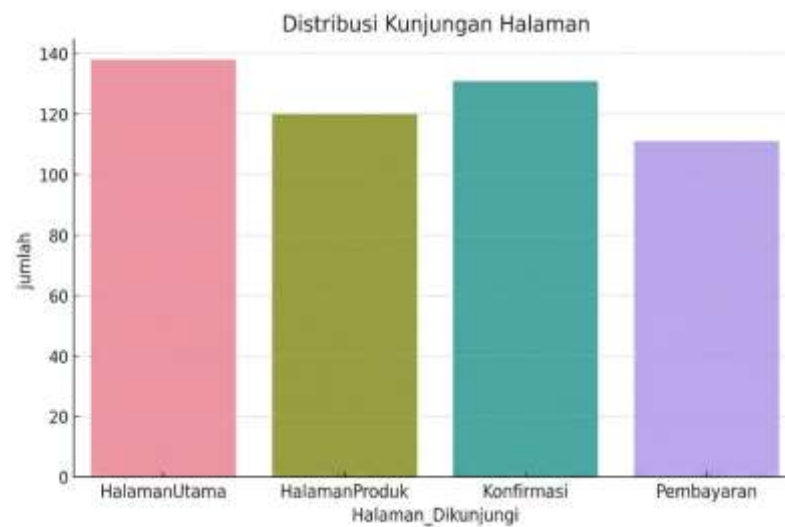
Hasil khusus ini menunjukkan pengguna yang paling berpengaruh dalam jaringan sosial berdasarkan ukuran sentralitas yang berbeda. Sentralitas derajat menyoroti pengguna dengan koneksi terbanyak, sentralitas antara mengidentifikasi pengguna yang bertindak sebagai jembatan antara bagian-bagian jaringan yang berbeda, dan sentralitas vektor eigen menunjukkan pengguna yang terhubung dengan pengguna berpengaruh lainnya.

Pembahasan ini memberikan wawasan berharga dalam mengidentifikasi influencer kunci dalam jaringan sosial, yang berguna untuk strategi pemasaran yang ditargetkan.

## **LATIHAN 7.2**

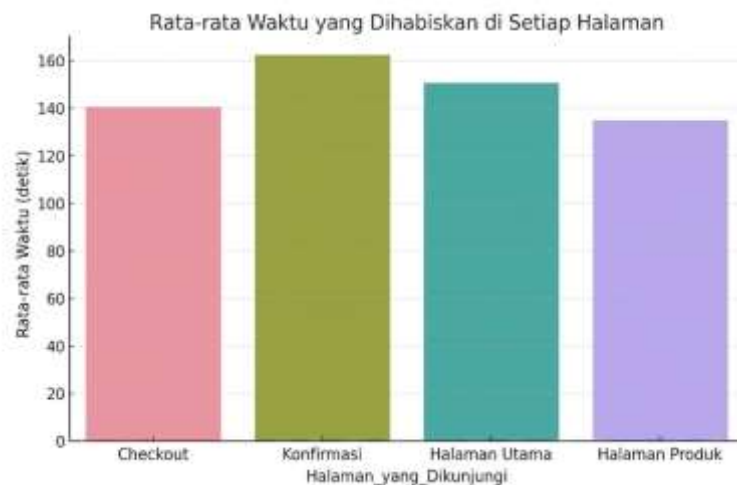
### **1. Visualisasi Distribusi Kunjungan Halaman:**

- Plot hitungan menunjukkan distribusi kunjungan di berbagai halaman. Halaman Beranda menerima kunjungan terbanyak, diikuti oleh halaman Konfirmasi, Halaman Produk, dan halaman Checkout.



## **2. Visualisasi Rata-Rata Waktu yang Dhabiskan di Setiap Halaman:**

- Grafik batang menunjukkan rata-rata waktu yang dihabiskan di setiap halaman. Halaman Konfirmasi memiliki rata-rata waktu yang dihabiskan tertinggi, diikuti oleh Halaman Beranda, Halaman Pembayaran, dan Halaman Produk.



## **3. Halaman yang Paling Sering Dikunjungi:**

- Halaman Utama: 138 kunjungan
- Konfirmasi: 131 kunjungan
- Halaman Produk: 120 kunjungan
- Checkout: 111 kunjungan

## **4. Rata-rata Waktu yang Dhabiskan di Halaman:**

- Checkout: 140,65 detik
- Konfirmasi: 162,15 detik
- Halaman Utama: 150,54 detik
- Halaman Produk: 134,92 detik

## **5. Rasio Bounce:**

- Rasio bounce sekitar 55,71%, menunjukkan bahwa lebih dari setengah sesi adalah sesi satu halaman.

#### 6. Rasio Konversi:

- Rasio konversi sekitar 29,6%, mewakili proporsi kunjungan yang menghasilkan konversi.

Hasil ini memberikan wawasan tentang perilaku pengguna di situs web, seperti halaman mana yang paling dan paling kurang menarik, dan efektivitas keseluruhan situs web dalam mempertahankan pengunjung dan mendorong konversi. Informasi ini dapat digunakan untuk mengoptimalkan situs web dan meningkatkan pengalaman pengguna.

### LATIHAN 8.1

#### 1. Pembahasan Data Eksplorasi (EDA):

- **Ringkasan Statistik:**
  - **Penjualan:** Berkisar antara sekitar 22.195 hingga 99.835 dengan rata-rata sekitar 62.417.
  - **Pengeluaran Iklan TV, Pengeluaran Iklan Online, Pengeluaran Iklan Radio:** Menunjukkan berbagai tingkat pengeluaran, dengan iklan TV umumnya memiliki pengeluaran tertinggi.
  - **Diskon Promosi:** Bervariasi antara 5% hingga sekitar 20%.
- **Matriks Korelasi:**
  - Korelasi antara pengeluaran pemasaran (**Pengeluaran Iklan TV, Pengeluaran Iklan Online, Pengeluaran Iklan Radio**) dan Penjualan relatif rendah. Ini menunjukkan bahwa variabel-variabel ini mungkin tidak memiliki hubungan linier yang kuat dengan penjualan dalam dataset ini.
  - **Pengeluaran\_Iklan\_Radio** memiliki korelasi negatif dengan Penjualan, yang merupakan aspek menarik yang perlu diselidiki lebih lanjut.

#### 2. Hasil Model Regresi:

- **Koefisien:**
  - **Pengeluaran\_Iklan\_TV, Pengeluaran\_Iklan\_Online, Pengeluaran\_Iklan\_Radio, dan Diskon\_Promosi** memiliki koefisien masing-masing sebesar 0,1358, 0,1138, -2,1305, dan -390,0831.
  - Patut dicatat bahwa **Pengeluaran\_Iklan\_Radio** memiliki koefisien negatif, yang menunjukkan potensi dampak negatif pada penjualan untuk setiap peningkatan satu unit dalam pengeluaran iklan radio.
- **Kesesuaian Model:**
  - Nilai R-kuadrat model adalah 0,018, yang cukup rendah. Ini menunjukkan bahwa model hanya menjelaskan sebagian kecil variabilitas dalam data penjualan.
  - Statistik F dan nilai p terkaitnya menunjukkan bahwa model tidak signifikan secara statistik pada tingkat signifikansi konvensional.
- **Interpretasi:**

- Daya penjelas model yang rendah (R-squared) dan kurangnya signifikansi statistik (nilai p dari statistik F) menunjukkan bahwa model tersebut mungkin bukan yang paling sesuai untuk data ini. Hal ini bisa disebabkan oleh sifat data sintetis atau kemungkinan bahwa hubungan antara variabel-variabel ini dan penjualan tidak linier.
- Koefisien negatif untuk **Radio\_Ad\_Spend** mungkin menyiratkan bahwa iklan radio tidak efektif untuk dataset khusus ini atau ada faktor pengganggu lain yang tidak diperhitungkan dalam model.

Dalam skenario dunia nyata, temuan seperti itu akan mengarah pada penyelidikan lebih lanjut, mungkin dengan mempertimbangkan variabel tambahan, mengeksplorasi model nonlinier, atau menyempurnakan metode pengumpulan data. Latihan ini berharga untuk memahami proses membangun dan menafsirkan model bauran pemasaran, meskipun sifat sintetis data dapat membatasi penerapan temuan spesifik ini di dunia nyata.

## **LATIHAN 8.2**

Koefisien yang diurutkan mewakili dampak setiap titik sentuh terhadap kemungkinan konversi. Koefisien positif menunjukkan pengaruh positif terhadap konversi, dan koefisien negatif menunjukkan pengaruh negatif.

### **1. Titik Sentuh dan Koefisiennya:**

- **Iklan Online:** Koefisien 0,210482. Ini menunjukkan pengaruh positif terkuat terhadap konversi di antara titik sentuh.
- **Media Sosial:** Koefisien 0,082429. Ini juga memengaruhi konversi secara positif tetapi dalam skala yang lebih kecil daripada iklan online.
- **Email:** Koefisien -0,079067. Titik sentuh ini tampaknya memiliki sedikit pengaruh negatif terhadap konversi.
- **Kunjungan Langsung:** Koefisien -0,290814. Ini menunjukkan pengaruh negatif terhadap konversi, lebih besar daripada email.
- **Iklan Pencarian:** Koefisien -0,381058. Hal ini memiliki dampak negatif paling signifikan terhadap konversi di antara titik kontak.

### **2. Interpretasi:**

- **Pengaruh Positif:** Koefisien positif untuk 'Iklan Online' dan 'Media Sosial' menunjukkan bahwa titik kontak ini efektif dalam mendorong konversi dalam dataset sintetis.
- **Pengaruh Negatif:** 'Email', 'Kunjungan Langsung', dan 'Iklan Pencarian' yang menunjukkan koefisien negatif mengindikasikan bahwa titik kontak ini mungkin kurang efektif atau bahkan kontraproduktif dalam menghasilkan konversi dalam dataset spesifik ini.

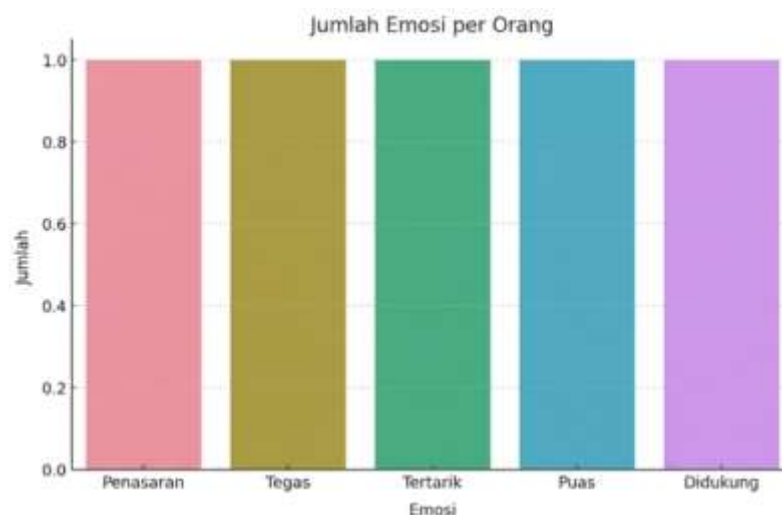
Pembahasan ini memberikan wawasan hipotetis tentang bagaimana berbagai titik kontak dapat memengaruhi konversi pelanggan. Dalam skenario dunia nyata, wawasan ini dapat berperan penting dalam memandu strategi pemasaran, meskipun hasilnya akan bergantung pada kualitas dan sifat data aktual.

Latihan ini, dengan fokusnya pada regresi logistik untuk atribusi, menyoroti potensi metode berbasis data dalam memahami perjalanan pelanggan dan mengoptimalkan titik kontak pemasaran untuk hasil konversi yang lebih baik.

### LATIHAN 9.1

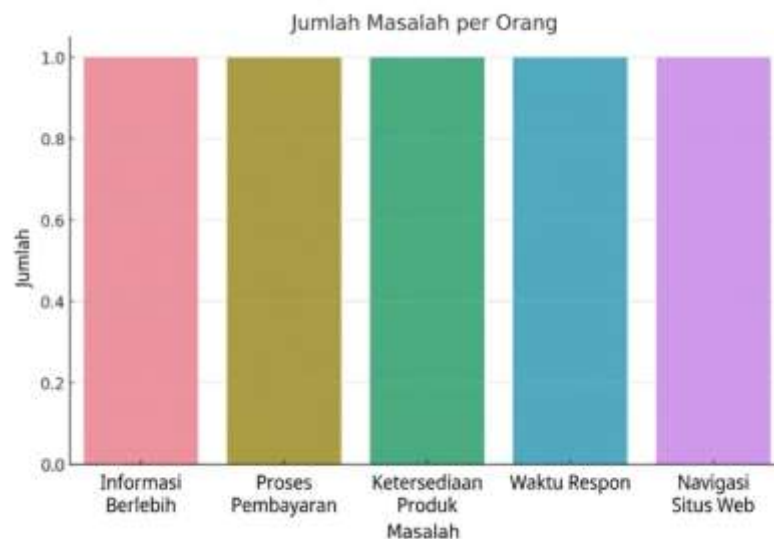
#### 1. Pembahasan Emosi dan Titik Masalah:

Jumlah emosi per persona menunjukkan distribusi yang seragam di berbagai persona untuk setiap emosi. Demikian pula, jumlah titik masalah juga menampilkan distribusi yang seragam di seluruh persona.



#### 2. Visualisasi:

- **Jumlah Emosi per Persona:** Grafik batang memvisualisasikan jumlah emosi yang berbeda untuk setiap persona. Setiap persona mengalami setiap emosi sekali, menunjukkan representasi yang seimbang dalam data.
- **Jumlah Masalah per Persona:** Grafik batang ini menunjukkan jumlah masalah yang berbeda untuk setiap persona. Mirip dengan emosi, setiap masalah juga terwakili secara seragam di seluruh persona.



Pembahasan dan visualisasi ini memberikan wawasan tentang bagaimana berbagai persona berinteraksi dengan berbagai titik kontak, beserta respons emosional dan masalah yang mereka hadapi. Informasi ini sangat penting untuk menyesuaikan strategi pengalaman pelanggan.

## LATIHAN 9.2

Berikut adalah hasil pembahasan dan visualisasi untuk Latihan 9.2:

### 1. Gambaran Umum Data:

Beberapa baris pertama dari dataset memberikan gambaran singkat tentang strukturnya, dengan setiap baris mewakili titik kontak dan metrik terkait seperti skor kepuasan pelanggan, tingkat konversi, dan kunjungan/pembelian berulang.

### 2. Ringkasan Dataset:

Ringkasan dataset memberi kita gambaran statistik. Ini menunjukkan bahwa skor kepuasan pelanggan rata-rata sekitar 82,6, dengan tingkat konversi rata-rata 6,22% dan rata-rata 33% untuk kunjungan/pembelian berulang.

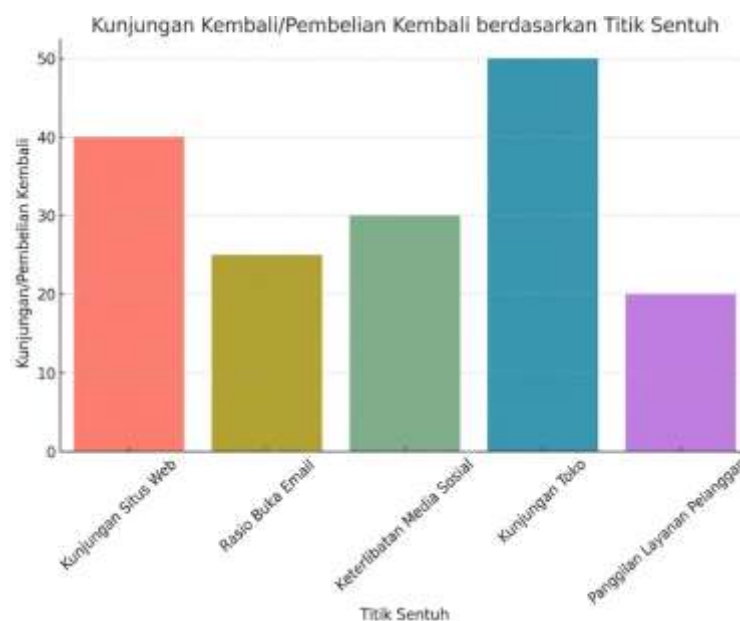
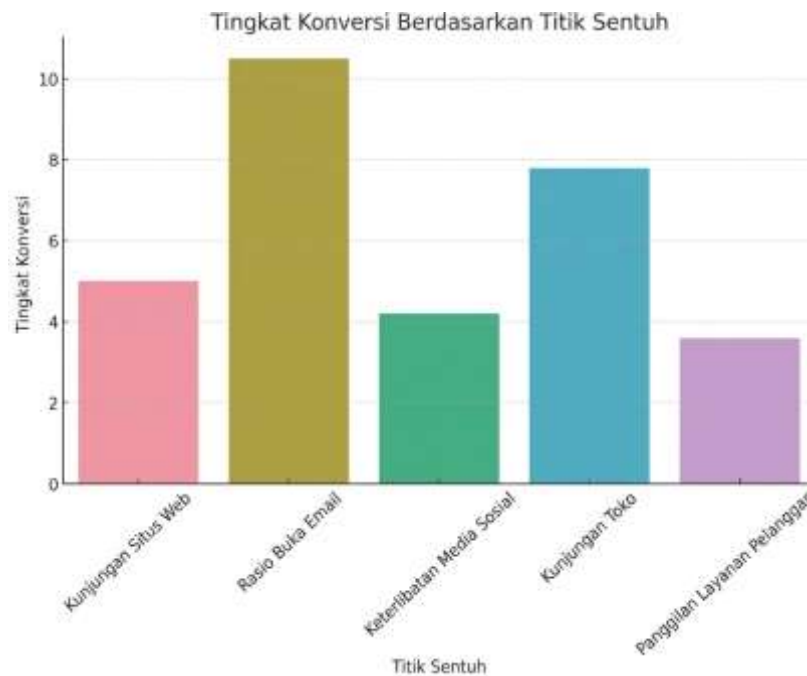
### 3. Visualisasi:

- **Skor Kepuasan Pelanggan berdasarkan Titik Kontak:** Grafik batang menunjukkan skor kepuasan pelanggan untuk setiap titik kontak. 'Kunjungan ke Toko' memiliki skor kepuasan tertinggi, yang menunjukkan kinerja yang kuat di area ini.



- **Tingkat Konversi berdasarkan Titik Kontak:** Grafik ini memvisualisasikan tingkat konversi yang terkait dengan setiap titik kontak. 'Tingkat Pembukaan Email' menonjol dengan tingkat konversi tertinggi, menunjukkan bahwa pemasaran email sangat efektif untuk dataset sintesis ini.

- **Kunjungan/Pembelian Berulang berdasarkan Titik Kontak:** Grafik terakhir menunjukkan persentase kunjungan atau pembelian berulang untuk setiap titik kontak. 'Kunjungan di Toko' sekali lagi menunjukkan kinerja yang kuat, menunjukkan bahwa pelanggan yang mengunjungi toko lebih cenderung untuk kembali atau melakukan pembelian berulang.



Visualisasi ini memberikan pemahaman yang jelas tentang kinerja setiap titik kontak dalam hal kepuasan pelanggan, konversi, dan retensi pelanggan. Wawasan seperti ini sangat penting bagi



bisnis untuk mengidentifikasi area mana yang berjalan dengan baik dan mana yang perlu ditingkatkan atau diinvestasikan lebih lanjut.

### **LATIHAN 10.1**

1. Hasil:

- **t\_stat**: -7.041427369013264
- **p\_value**: 3.059820094514218e-11

**t\_stat** adalah nilai statistik t yang dihitung, dan **p\_value** adalah probabilitas mengamati nilai yang ekstrem seperti statistik t di bawah hipotesis nol. Dalam hal ini, nilai p sangat rendah (jauh di bawah ambang batas tipikal 0,05), menunjukkan bahwa ada perbedaan yang signifikan secara statistik antara tingkat pembukaan email Grup A dan Grup B.

Kesimpulannya, berdasarkan analisis ini, kita dapat dengan yakin mengatakan bahwa tingkat pembukaan email dari kedua baris subjek tersebut berbeda secara signifikan. Jika rata-rata tingkat pembukaan email Grup B lebih tinggi, ini menunjukkan bahwa subjek email B lebih efektif dalam kampanye pemasaran email ini.

### **LATIHAN 10.2**

**Ringkasan Model:** Ringkasan model regresi memberikan berbagai statistik, termasuk koefisien untuk setiap variabel independen, kesalahan standar, nilai t, dan nilai p. Nilai-nilai ini membantu kita memahami dampak setiap faktor pada rasio klik-tayang.

- **Koefisien:** Menunjukkan perubahan pada variabel dependen untuk perubahan satu unit pada variabel independen.
- **Nilai p:** Membantu menentukan signifikansi statistik dari setiap koefisien. Nilai p yang rendah (biasanya < 0,05) menunjukkan bahwa faktor tersebut kemungkinan besar memiliki dampak signifikan pada rasio klik-tayang.
- **R-kuadrat dan R-kuadrat yang Disesuaikan:** Mengukur proporsi varians dalam variabel dependen yang dapat diprediksi dari variabel independen.

#### **Hasil Regresi OLS**

- **Variabel Dependen:** Rasio Klik Tayang
- **R-kuadrat:** 0,015, menunjukkan bahwa hanya sekitar 1,5% variabilitas dalam 'Rasio Klik Tayang' yang dijelaskan oleh model.
- **R-kuadrat yang disesuaikan:** -0,026, yang negatif, menunjukkan bahwa model mungkin tidak sesuai untuk data.
- **Statistik F:** 0,3724 dengan nilai p 0,828, menunjukkan bahwa model mungkin tidak signifikan secara statistik secara keseluruhan.
- **Koefisien:**
  - konstanta: 0,2808 (Intersep)
  - Warna\_Iklan\_Hijau: 0,0188
  - Warna\_Iklan\_Merah: 0,0134
  - Penempatan\_Atas: 0,0137
  - Ukuran\_Kecil: 0,0174

- **Nilai-p:** Nilai-p untuk koefisien individual tinggi, menunjukkan bahwa tidak satu pun faktor periklanan (Warna\_Iklan\_Hijau, Warna\_Iklan\_Merah, Penempatan\_Atas, Ukuran\_Kecil) memiliki dampak yang signifikan secara statistik pada 'Tingkat Klik Tayang' pada tingkat signifikansi 5%.

Dalam analisis khusus ini, nilai R-kuadrat cukup rendah, menunjukkan bahwa model tidak menjelaskan sebagian besar varians dalam tingkat klik tayang. Ini mungkin menunjukkan bahwa faktor lain yang tidak termasuk dalam model atau keacakan yang melekat memainkan peran penting dalam tingkat klik tayang. Faktor-faktor individual (warna iklan, penempatan, ukuran) tidak menunjukkan signifikansi statistik yang kuat dalam model ini, seperti yang ditunjukkan oleh nilai-p yang tinggi.

# ILMU DATA PEMASARAN

**Dr. Agus Wibowo, M.Kom, M.Si, MM.**

## BIO DATA PENULIS



Penulis memiliki berbagai disiplin ilmu yang diperoleh dari Universitas Diponegoro (UNDIP) Semarang. dan dari Universitas Kristen Satya Wacana (UKSW) Salatiga. Disiplin ilmu itu antara lain teknik elektro, komputer, manajemen dan ilmu sosiologi. Penulis memiliki pengalaman kerja pada industri elektronik dan sertifikasi keahlian dalam bidang Jaringan Internet, Telekomunikasi, Artificial Intelligence, Internet Of Things (IoT), Augmented Reality (AR), Technopreneurship, Internet Marketing dan bidang pengolahan dan analisa data (komputer statistik).

Penulis adalah pendiri dari Universitas Sains dan Teknologi Komputer (Universitas STEKOM ) dan juga seorang dosen yang memiliki Jabatan Fungsional Akademik Lektor Kepala (Associate Professor) yang telah menghasilkan puluhan Buku Ajar ber ISBN, HAKI dari beberapa karya cipta dan Hak Paten pada produk IPTEK. Sejak tahun 2023 penulis tercatat sebagai Dosen luar biasa di Fakultas Ekonomi & Bisnis (FEB) Universitas Diponegoro Semarang. Penulis juga terlibat dalam berbagai organisasi profesi dan industri yang terkait dengan dunia usaha dan industri, khususnya dalam pengembangan sumber daya manusia yang unggul untuk memenuhi kebutuhan dunia kerja secara nyata.



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

ISBN 978-634-7695-27-7 (PDF)



9

786347

695277