



Manfaat AI dalam Organisasi

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





Manfaat AI dalam Organisasi

Dr. Ir. 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 978-634-7227-57-7 (PDF)



9

786347

227577

Manfaat AI dalam Organisasi

Penulis :

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

ISBN : 978-634-7227-57-7 (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 syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa karena berkat rahmat dan karunia-Nya buku berjudul "*Manfaat AI dalam Organisasi*" ini dapat diselesaikan. Di era digital yang serba cepat seperti saat ini, kecerdasan buatan (AI) telah menjadi salah satu pilar utama transformasi organisasi dan bisnis. Teknologi ini bukan lagi sekadar alat bantu, melainkan sebuah kekuatan penggerak yang mampu menghadirkan perubahan revolusioner dalam berbagai sektor. Buku ini hadir sebagai panduan komprehensif yang dirancang untuk membantu pembaca memahami konsep dasar AI, aplikasi praktisnya, serta strategi implementasi yang efektif di lingkungan organisasi.

Bab 1 membuka perjalanan pembelajaran dengan pengantar menyeluruh mengenai kecerdasan buatan. Pembaca diajak menelusuri perkembangan AI dari masa ke masa, memahami revolusi yang dibawa oleh pembelajaran mesin, serta mengenal esensi sebenarnya dari AI. Bab ini turut membahas beragam metode pengajaran AI yang menjadi fondasi perkembangan teknologi ini.

Memasuki Bab 2, fokus beralih pada pemanfaatan AI dalam data bisnis inti. Bab ini memaparkan bagaimana organisasi dapat mengoptimalkan AI untuk mengelola dan menganalisis data penting mereka. Selain itu, terdapat studi kasus nyata yang menjelaskan penerapan AI, sekaligus pembahasan mengenai evaluasi kinerja dan risiko yang mungkin dihadapi.

Bab 3 menitikberatkan pada aplikasi AI dalam penjualan dan pemasaran. Di sini dijelaskan bagaimana AI dapat memprediksi perilaku pelanggan, meningkatkan rasio konversi dan upselling, serta melakukan segmentasi pelanggan secara otomatis. Pengukuran kinerja dan risiko bisnis berbasis pembelajaran mesin juga dibedah secara detail, lengkap dengan studi kasus yang memperkuat pemahaman.

Bab 4 menguraikan peranan AI dalam media, meliputi teknologi visi komputer untuk pengolahan gambar dan video, pembelajaran mendalam, pengenalan wajah, serta pembuatan konten kreatif. Bab ini juga menunjukkan bagaimana AI diaplikasikan pada audio dan bidang pertanian, termasuk studi kasus pendukung.

Di Bab 5, buku ini mengupas pemanfaatan AI dalam pemahaman bahasa alami atau *natural language processing* (NLP). Bab ini menguraikan kompleksitas NLP, manfaatnya bagi organisasi, serta contoh sukses penerapan NLP dalam dunia nyata.

Bab 6 membahas penggunaan AI untuk validasi konten dan pengembangan komunitas. Pembaca akan mengenal sistem rekomendasi yang meningkatkan keterlibatan pengguna, metode penyaringan kolaboratif, dan risiko yang patut diwaspadai. Studi kasus platform besar seperti Netflix memberikan gambaran nyata bagaimana AI bisa menghasilkan efisiensi besar.

Bab 7 menghadirkan panduan bagi pembaca untuk mengidentifikasi dan mengevaluasi peluang dalam pemanfaatan AI. Bab ini menekankan kewaspadaan terhadap hype yang

berlebihan, cara menetapkan prioritas proyek, analisis risiko, serta memecah produk AI menjadi komponen yang dapat dianalisis secara objektif.

Bab 8 menekankan pentingnya persiapan matang dalam membangun fondasi AI, meliputi strategi pengelolaan data, peningkatan kualitas data, dan pengembangan tim AI yang kompeten untuk menjawab kebutuhan organisasi.

Pada Bab 9, pembaca dibekali dengan strategi implementasi AI yang efektif. Isi bab ini mencakup keputusan antara membeli atau membangun solusi AI, penerapan strategi lean, pemahaman siklus hidup proyek, pengelolaan, serta tata cara menghadapi kegagalan yang mungkin terjadi.

Bab 10 menutup dengan pembahasan tentang masa depan AI, menyentuh potensi ancaman dan peluang yang muncul, serta refleksi mendalam mengenai perkembangan kecerdasan buatan umum (AGI). Bab ini mengajak pembaca untuk berpikir kritis sekaligus optimis atas peran AI ke depan dalam berbagai sektor masyarakat dan industri.

Melalui pembahasan lengkap dan terstruktur ini, buku "*Manfaat AI dalam Organisasi*" menawarkan wawasan yang menyeluruh sekaligus praktis. Kami berharap buku ini dapat menjadi panduan utama bagi para pemimpin, profesional, dan inovator yang ingin memanfaatkan kecerdasan buatan untuk mendorong kemajuan dan keberlanjutan organisasi mereka di masa depan.

Akhir kata, kami mengucapkan terima kasih kepada semua pihak yang telah mendukung proses penyusunan buku ini, mulai dari para ahli, praktisi, hingga pembaca yang memberikan inspirasi. Semoga buku ini dapat memberikan manfaat besar dan membuka wawasan baru bagi Anda.

Selamat dan Semangat Membaca...!!!

Semarang, Oktober 2025
Penulis

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

DAFTAR ISI

KATA PENGANTAR.....	ii
DAFTAR ISI.....	iv
BAB 1 PENGANTAR KECERDASAN BUATAN	1
1.1 Jalan Menuju AI Modern	1
1.2 Pembelajaran Mesin Sebagai Revolusi AI.....	4
1.3 Definisi Kecerdasan Buatan	6
1.4 Metode Pengajaran AI.....	8
BAB 2 KECERDASAN BUATAN UNTUK DATA BISNIS INTI	11
2.1 Memanfaatkan AI Pada Data Bisnis Inti	11
2.2 Menggunakan AI Dengan Data Bisnis Inti	13
2.3 Studi Kasus.....	23
2.4 Mengevaluasi Kinerja Dan Risiko.....	33
BAB 3 AI UNTUK PENJUALAN DAN PEMASARAN	36
3.1 Mengapa AI Untuk Penjualan Dan Pemasaran.....	36
3.2 Memprediksi Pelanggan Yang Churn	38
3.3 Optimalkan Konversi Dan Upselling Dengan AI.....	42
3.4 Melakukan Segmentasi Pelanggan Otomatis	45
3.5 Pengukuran Kinerja Prediksi Berbasis Machine Learning	53
3.6 Menghubungkan Metrik ML Dengan Hasil Dan Risiko Bisnis.....	57
3.7 Studi Kasus.....	60
BAB 4 ARTIFICIAL INTELLIGENCE UNTUK MEDIA.....	70
4.1 Meningkatkan Produk Dengan Visi Komputer	70
4.2 Pembelajaran Mendalam Untuk Klasifikasi Gambar	74
4.3 Pembelajaran Transfer Dengan Set Data Kecil	78
4.4 Melatih Komputer Untuk Pengenalan Wajah.....	81
4.5 Menggunakan Pembuatan Konten Dan Transfer Gaya.....	83
4.6 Yang Perlu Diwaspadai.....	85
4.7 AI Untuk Audio	87
4.8 Optimasi Pertanian Dengan Deep Learning: Studi Kasus	89
BAB 5 AI UNTUK BAHASA ALAMI	93
5.1 Daya Tarik Pemahaman Bahasa Alami	94
5.2 Menguraikan NLP: Mengukur Kompleksitas	95
5.3 Menambahkan Kapabilitas NLP Ke Organisasi Anda	99
5.4 Studi Kasus: Translated	116
BAB 6 AI UNTUK VALIDASI KONTEN DAN PENGEMBANGAN KOMUNITAS	123
6.1 Kutukan Pilihan.....	123

6.2	Meningkatkan Keterlibatan Dengan Sistem Rekomendasi.....	124
6.3	Kebijaksanaan Kerumunan: Penyaringan Kolaboratif	131
6.4	Rekomendasi Yang Salah	133
6.5	Studi Kasus: Netflix Menghemat Rp. 10 Triliun Per Tahun	135
BAB 7	SIAP MENEMUKAN PELUANG AI	142
7.1	Waspada Hype: Inovasi AI Untuk Bisnis	142
7.2	Penemuan: Mencari Peluang AI	147
7.3	Prioritas: Mengevaluasi Proyek AI.....	150
7.4	Validasi: Menganalisis Risiko	154
7.5	Mendekonstruksi Produk AI	156
7.6	Menerjemahkan Proyek AI Ke Dalam Istilah Yang Ramah ML	162
7.7	Latihan Studi Kasus.....	166
BAB 8	SIAPKAN DATA, TEKNOLOGI, DAN SDM	173
8.1	Strategi Data	173
8.2	Kualitas Data	183
8.3	Merekrut Tim AI	186
BAB 9	STRATEGI IMPLEMENTASI AI	191
9.1	Membeli Atau Membangun AI	191
9.2	Menggunakan Strategi Lean.....	198
9.3	Memahami Siklus Baik AI	204
9.4	Mengelola Proyek AI.....	210
9.5	Ketika AI Gagal.....	212
BAB 10	APA YANG AKAN DATANG	218
10.1	Bagaimana AI Mengancam Masyarakat	218
10.2	Peluang AI Dalam Masyarakat	228
10.3	Peluang AI Dalam Industri	230
10.4	AI Umum Adalah Topik Yang Penting	240
10.5	Pikiran Penutup	241
DAFTAR PUSTAKA		243

BAB 1

PENGANTAR KECERDASAN BUATAN

Bab ini membahas

- Memperoleh perspektif tentang sejarah kecerdasan buatan
- Memahami pembelajaran mesin dan hubungannya dengan AI
- Menjelajahi pendorong ledakan aplikasi AI

Kecerdasan buatan (AI) bukanlah teknologi baru. Selama beberapa dekade, ilmuwan komputer telah mencoba berbagai pendekatan untuk mencapai tujuan utama komputasi: mesin cerdas. Meskipun kita masih jauh dari mereplikasi keajaiban otak manusia, aplikasi AI telah mulai mengisi kehidupan kita sehari-hari dan memberi daya pada perangkat elektronik kita, mulai dari ponsel pintar hingga sistem alarm rumah.

Mengapa ledakan yang tampaknya tiba-tiba ini terjadi? Bab ini akan menjawab pertanyaan ini dengan mengajarkan Anda tentang AI modern termasuk prinsip-prinsip inti di baliknya, serta bagaimana dan mengapa kita sampai pada titik ini.

1.1 JALAN MENUJU AI MODERN

Sebagai manusia, kita selalu berusaha menemukan cara untuk memahami dunia di sekitar kita dan memanfaatkan alam untuk mencapai tujuan kita. Untuk melakukannya, kita selalu mengandalkan perangkat eksternal yang memperkuat kemampuan otak kita. Sempoa mungkin merupakan alat pertama semacam itu, ditemukan sekitar 5.000 hingga 6.000 tahun yang lalu untuk membantu orang berhitung.

Meskipun masih digunakan di sekolah untuk membantu anak-anak memvisualisasikan operasi matematika sederhana, itu tidak benar-benar menyelamatkan kita dari kerja keras untuk melakukannya. Kita harus menunggu hingga tahun 1960an untuk mesin pertama yang dapat menjumlahkan dan mengurangi angka secara otomatis. Komputer telah berkembang pesat sejak saat itu, tetapi jauh di lubuk hatinya, kemampuannya masih cukup sederhana: mengeksekusi perhitungan persis seperti yang diinstruksikan oleh manusia (ahli). Hanya ada sedikit "kecerdasan" di dalamnya.

Dua kata buatan dan kecerdasan pertama kali disatukan pada 31 Agustus 1955, ketika profesor John McCarthy dari Dartmouth College, bersama dengan M.L. Minsky dari Universitas Harvard, N. Rochester dari IBM, dan C. E. Shannon dari Bell Telephone Laboratories, meminta Rockefeller Foundation untuk mendanai penelitian musim panas tentang kecerdasan buatan. Proposal mereka menyatakan sebagai berikut:

Kami mengusulkan agar studi kecerdasan buatan selama 2 bulan yang melibatkan 10 orang dilakukan pada musim panas tahun 1956 di Dartmouth College di Hanover, New Hampshire. Akan dilakukan upaya untuk menemukan cara membuat

mesin menggunakan bahasa, membentuk abstraksi dan konsep, memecahkan berbagai masalah yang kini hanya diperuntukkan bagi manusia, dan meningkatkan kemampuan diri mereka sendiri. Kami berpikir bahwa kemajuan signifikan dapat dicapai dalam satu atau lebih masalah ini jika sekelompok ilmuwan yang dipilih secara cermat mengerjakannya bersama-sama selama musim panas.

Para peneliti menyadari bahwa menangani kecerdasan secara keseluruhan merupakan tantangan yang terlalu berat, baik karena keterbatasan teknis maupun kompleksitas tugas yang melekat. Alih-alih memecahkan konsep kecerdasan yang luas, mereka memutuskan untuk berfokus pada sub-masalah, seperti bahasa. Nantinya, aplikasi ini akan disebut AI sempit. Kecerdasan buatan yang mampu menyamai atau melampaui kemampuan manusia akan disebut AI umum. Dengan kata lain:

- *AI umum (atau AI kuat)*—Program kecerdasan buatan yang mampu menangani setiap jenis tugas yang disajikan. Ini mirip dengan manusia yang sangat banyak akal, dan Anda dapat menganggapnya sebagai robot dari *The Terminator* (atau, semoga saja, versi yang lebih damai).
- *AI Sempit* —Program kecerdasan buatan yang mampu menyelesaikan satu tugas yang terdefinisi dengan baik. AI ini bisa luas (mengenali objek dari gambar) atau sangat spesifik (memprediksi pelanggan mana yang membeli produk A yang lebih mungkin membeli produk B juga). Ini berarti satu tugas pada satu waktu, dan bukan yang lain: AI yang mengenali kucing dalam gambar tidak dapat menerjemahkan bahasa Inggris ke bahasa Italia, dan sebaliknya.

AI umum masih jauh: para peneliti masih belum tahu kapan kita akhirnya akan mendapatkannya. Beberapa berpendapat bahwa kita tidak akan pernah sampai di sana. Meskipun AI umum masih merupakan mimpi yang jauh dan samar, inilah yang ada dalam pikiran banyak orang ketika AI disebutkan dalam berita. Jika Anda salah satu dari orang-orang itu, dan sekarang kecewa karena AI umum belum ada, jangan putus asa. Aplikasi AI sempit masih mampu menciptakan nilai yang sangat besar. Misalnya, AI yang dapat mendeteksi kanker paru-paru merupakan aplikasi yang sempit tetapi tetap sangat berguna.

Hasil penelitian Dartmouth pada musim panas 1956 begitu menarik sehingga memicu gelombang kegembiraan dan harapan di antara para peserta. Antusiasme para ilmuwan menyebar ke pemerintah AS, yang mulai mendanai penelitian secara besar-besaran pada aplikasi spesifik: penerjemahan bahasa Inggris/Rusia. Menemukan penerjemah Rusia yang tepercaya pastilah tidak mudah di tengah Perang Dingin.

Setelah beberapa tahun pertama penelitian, sebuah komite pemerintah menghasilkan laporan Komite Penasihat Pemrosesan Bahasa Otomatis (ALPAC) tahun 1966 yang terkenal. Dokumen tersebut memuat pendapat banyak peneliti tentang perkembangan penelitian AI. Sebagian besar tidak terlalu positif:

Terjemahan mesin awal untuk teks sederhana atau terpilih... sama menggembirkannya dengan "terjemahan mesin" untuk teks ilmiah umum yang secara seragam mengecewakan. Tentu saja, tidak ada yang dapat menjamin bahwa kita tidak akan tiba-tiba atau setidaknya dengan cepat mencapai terjemahan mesin, tetapi kami merasa hal ini sangat tidak mungkin.
... tidak ada prospek terjemahan mesin yang bermanfaat secara langsung atau dapat diprediksi.

Laporan ALPAC menandai dimulainya periode yang disebut musim dingin AI pertama: pendanaan publik untuk penelitian AI terhenti, antusiasme mereda, dan para peneliti memfokuskan penelitian mereka pada bidang lain.

Ketertarikan pada AI memudar hingga tahun 1980-an, ketika perusahaan swasta seperti IBM dan Xerox mulai berinvestasi pada musim semi AI yang baru. Harapan baru didorong oleh teknologi yang disebut sistem pakar: program komputer yang mengodekan pengetahuan seorang pakar di bidang tertentu dalam bentuk aturan "jika-maka" yang presisi. Sebuah contoh akan membantu Anda memahami bagaimana sistem pakar dirancang untuk bekerja.

Misalkan Anda ingin membangun sistem AI yang dapat menggantikan seorang ahli gastroenterologi. Beginilah cara Anda melakukannya dengan sistem pakar: Anda meminta seorang dokter untuk menjelaskan dengan sangat presisi bagaimana mereka membuat keputusan tentang pasien. Anda kemudian meminta seorang programmer untuk dengan cermat mengubah pengetahuan dan alur diagnosis dokter menjadi aturan "jika-maka" yang dapat dipahami dan dijalankan oleh komputer. Versi yang sangat disederhanakan akan terlihat seperti ini:

Jika pasien sakit perut dan suhu tubuhnya tinggi, maka pasien tersebut terkena flu.
Jika pasien sakit perut dan mengonsumsi makanan kedaluwarsa, maka pasien tersebut mengalami keracunan makanan.

Dan seterusnya. Setelah pengetahuan dokter dikodekan ke dalam perangkat lunak dan pasien datang, perangkat lunak mengikuti alur keputusan yang sama dengan dokter dan (semoga) menghasilkan diagnosis yang sama. Pendekatan ini memiliki beberapa masalah:

- *Adaptabilitas yang buruk* — Satu-satunya cara untuk meningkatkan perangkat lunak adalah dengan kembali ke papan gambar bersama ilmuwan komputer dan pakar (dalam hal ini, dokter).
- *Kerapuhan yang ekstrem* — Sistem akan gagal dalam situasi yang bukan bagian dari desain awal. Bagaimana jika pasien sakit perut tetapi suhu tubuhnya normal, dan belum mengonsumsi makanan basi?
- *Sulit dirawat* — Kompleksitas sistem seperti itu sangat besar. Ketika ribuan aturan disusun, memperbaiki atau mengubahnya menjadi sangat rumit, lambat, dan mahal. Pernahkah Anda bekerja dengan lembar Microsoft Excel yang sangat besar dan

kesulitan menemukan akar penyebab kesalahan? Bayangkan lembar Excel yang 100 kali lebih besar. Sistem pakar mengalami kegagalan komersial. Pada akhir 1980an, banyak perusahaan yang mengembangkannya gulung tikar, menandai dimulainya musim dingin AI kedua. Baru pada awal 2000an, generasi kesuksesan AI berikutnya muncul, didorong oleh ide lama yang menjadi baru lagi: **pembelajaran mesin**.

1.2 PEMBELAJARAN MESIN SEBAGAI REVOLUSI AI

Definisi pertama pembelajaran mesin berasal dari tahun 1959, dari pelopor AI Amerika Arthur Samuel:

Pembelajaran mesin adalah bidang studi yang memberi komputer kemampuan untuk belajar tanpa diprogram secara eksplisit.

Elemen kuncinya di sini adalah belajar dan tanpa diprogram secara eksplisit. Mari kita fokus pada yang terakhir terlebih dahulu. Memprogram komputer secara eksplisit berarti mendefinisikan aturan dan instruksi yang harus diikuti untuk melakukan tugas tertentu. Inilah yang dilakukan oleh para insinyur perangkat lunak ketika mereka menulis perangkat lunak yang menangani tugas sehari-hari Anda seperti menghitung pajak atau mengisi lembar kerja.

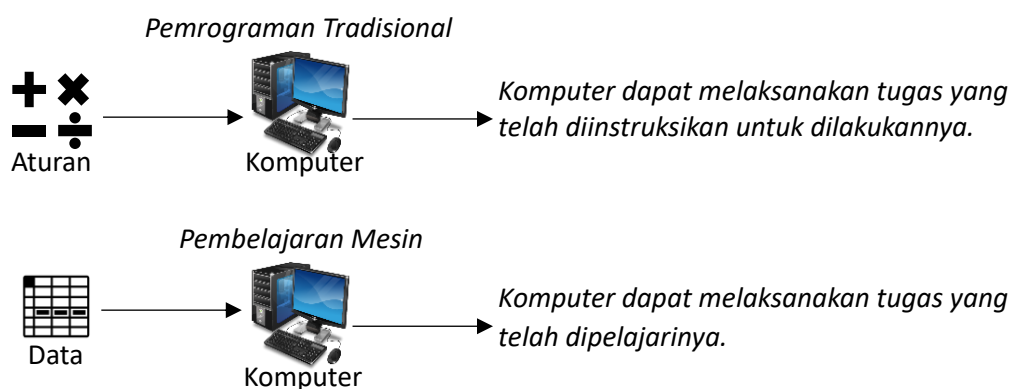
Orang-orang tanpa pengalaman pemrograman sering kali merasa bahwa insinyur perangkat lunak adalah makhluk hebat yang dapat mengendalikan mesin sesuai keinginan mereka. Sayangnya, segala sesuatunya tidak selalu semudah itu. Coba pikirkan berbagai keputusan yang Anda buat saat melakukan beberapa tindakan sepele: bisakah Anda menjelaskan proses yang Anda ikuti untuk mengenali teman-teman Anda saat bertemu mereka? Semua keputusan yang Anda buat dalam sekejap saat mengemudi? Bisakah Anda menyebutkan semua aturan tata bahasa Inggris yang Anda terapkan saat berbicara? Jika Anda tidak bisa menjelaskan secara tepat bagaimana Anda melakukan sesuatu, mustahil Anda bisa memerintahkan komputer untuk melakukannya.

Samuel mengusulkan untuk mengganti "menginstruksikan komputer" dengan "memberi mereka kemampuan untuk belajar." Jika dipikir-pikir, belajar, alih-alih mengikuti instruksi, adalah (kebetulan?) yang dilakukan manusia sepanjang waktu. Ibu dan ayah kita tidak mengajarkan kita bahasa ibu mereka dengan memberi kita buku tata bahasa di usia satu tahun. Mereka hanya berbicara kepada kita secara alami, dan kita belajar dari contoh mereka, menerapkan ribuan aturan tata bahasa tanpa menyadarinya. Faktanya, otak kita mampu secara otomatis mengekstrak aturan jauh sebelum ia mampu memahami tata bahasa secara rasional di sekolah! Bahkan bagi kita manusia, mempelajari aturan dari contoh tampaknya lebih mudah daripada hanya diberi tahu.

Layaknya kita belajar dari pengalaman, teknik pembelajaran mesin (ML) memungkinkan komputer belajar dari data. Mari kita buat lebih konkret dengan contoh mainan klasik: mengajarkan komputer membedakan anjing dari kucing melalui gambar. Jika Anda harus mengajarkan seorang anak untuk melakukan tugas ini, Anda tidak akan mengambil buku kedokteran hewan dan mulai membaca tentang perbedaan bentuk telinga atau warna bulu.

Sebaliknya, Anda mungkin hanya akan mengarahkan mereka ke beberapa gambar dan membiarkan otak mereka bekerja.

Solusi ML untuk masalah "anjing atau kucing" serupa dengan pengalaman belajar kita di masa kecil. Kita memberi komputer ribuan gambar kucing dan memberi tahunya "ini kucing," lalu ribuan gambar anjing dan memberi tahunya "ini anjing." Akhirnya, kita membiarkannya memahami perbedaan antara kedua hewan peliharaan tersebut secara otomatis. Kita tidak perlu menjelaskan elemen-elemen kunci yang membedakan anjing dari kucing. Aplikasi ML yang baik belajar memahami hal itu dari contoh-contoh yang diterimanya. Gambar 1.1 menunjukkan perbedaan antara pemrograman tradisional dan pembelajaran mesin.



Gambar 1.1 Perbedaan antara pendekatan pemrograman tradisional dan pembelajaran mesin: yang pertama bergantung pada aturan dan instruksi yang tepat, sedangkan yang terakhir bergantung pada data dan pembelajaran.

Anda mungkin mulai merasakan mengapa ML tidak mungkin berkembang sebelum tahun 2000an. Bahan utama dari serangkaian teknik ini adalah data, dan internet telah membuat pengumpulan data jauh lebih mudah. Bahan penting lainnya untuk ML adalah daya komputasi: belajar dari data tidak terjadi secara gratis, dan komputer membutuhkan prosesor yang cepat untuk melakukan tugas ini. Berkat komputasi awan dan peningkatan daya pemrosesan, akses ke komputer canggih tidak pernah semudah dan semurah ini.

Untuk memberi Anda gambaran tentang seberapa banyak hal telah berubah hanya dalam beberapa tahun, kami bertanya kepada Alex Waibel, salah satu pelopor AI dalam pengenalan suara dan salah satu karyawan pertama tim AI Facebook, betapa berbedanya bekerja di ML 20 tahun yang lalu. Komputer paling canggih yang dapat ia gunakan di awal tahun 2000an sebesar apartemen, harganya beberapa juta, dan ia perlu menyewanya untuk melatih modelnya. Saat ini, ia memiliki daya komputasi yang jauh lebih besar di mejanya dengan harga beberapa ribu dolar. Ponsel Anda mungkin lebih canggih daripada yang dimiliki para peneliti terkemuka hanya 20 tahun yang lalu.

Ketersediaan data dan daya komputasi yang murah menciptakan lingkungan yang sempurna bagi perkembangan pembelajaran mesin. Memang, banyak (sebagian besar) aplikasi terkeren yang dihadapi konsumen dari apa yang kita sebut AI saat ini sangat

bergantung pada ML: asisten virtual suara Siri, Google Translate, mobil self-driving, dan masih banyak lagi.

Menilik kembali sejarah AI, tampaknya ML adalah mesin yang mendorong ledakan AI saat ini, yang akhirnya membawa secercah harapan setelah musim dingin AI terakhir di tahun 1980an. Faktanya, keberhasilan AI modern sangat bergantung pada teknik ML sehingga orang sering bingung membedakan keduanya. Lalu, apa itu kecerdasan buatan? Mari kita cari tahu.

1.3 DEFINISI KECERDASAN BUATAN

Dalam pengalaman kami sebagai teknolog, konsultan, dan pembicara publik, kami terus-menerus bertemu orang-orang dengan pendapat berbeda tentang definisi AI, ilmu data, dan ML. Meskipun banyak yang cukup berpendirian, hanya sedikit yang dapat mempertahankan posisi mereka. Memang, menemukan definisi universal AI tidaklah semudah kelihatannya.

Berdasarkan namanya, kita mungkin mencoba mendefinisikan kecerdasan buatan dengan menemukan ciri-ciri manusia yang kita kaitkan dengan kecerdasan. Setelah kita sepakat tentang apa yang membuat manusia cerdas, kita dapat mengatakan bahwa komputer apa pun yang melakukan hal yang sama adalah AI. Masuk akal, bukan? Meskipun ini merupakan pendekatan yang umum, pendekatan ini tidak dapat diterapkan bahkan dalam skenario sederhana. Misalnya, manusia yang dapat membagi 13,856 dengan 13 hingga kesepuluh desimal dalam sepersekian detik pasti akan disebut cerdas, tetapi padanan buatanya adalah kalkulator saku seharga Rp. 20.000 yang tak seorang pun berani menyebutnya AI.

Pada saat yang sama, kita tidak akan pernah menyebut seseorang cerdas hanya karena mereka mampu mengemudi di lalu lintas yang padat, tetapi mobil tanpa pengemudi umumnya dianggap sebagai salah satu bentuk AI tersulit yang sedang dikembangkan oleh industri teknologi saat ini. Kita seharusnya tidak terkejut betapa sulitnya mendefinisikan kecerdasan; lagipula, para filsuf dan ilmuwan telah memperdebatkannya selama berabad-abad. Kita tidak hanya memiliki bobot yang berbeda untuk mengukur kecerdasan manusia dan mesin, tetapi kita juga tampaknya berubah pikiran dengan sangat cepat tentang apa itu AI dan apa yang bukan.

Mari kita ambil contoh dari Paul Graham, pendiri Y Combinator, akselerator startup Silicon Valley paling sukses, dan bisa dibilang salah satu orang yang paling berwawasan ke depan di bidang teknologi. Pada tahun 2002, Graham menulis esai yang mengusulkan solusi baru untuk mendeteksi email spam. Saat itu, email baru saja diluncurkan, dan spam (email yang tidak diinginkan) adalah salah satu ancaman paling serius bagi penggunaan internet secara luas oleh nonteknisi. Tampaknya sulit untuk membayangkannya sekarang, tetapi para ilmuwan komputer terbaik sibuk mencoba menulis aturan rumit yang memungkinkan komputer secara otomatis memilah iklan Viagra.

Dalam esainya, Graham memikirkan pendekatan berbasis ML baru yang akan belajar mengklasifikasikan email dengan memproses ribuan email "baik" dan spam. Perangkat lunak sederhana Paul belajar mengenali spam lebih baik daripada aturan rumit yang dibuat oleh para

insinyur. Maju cepat 20 tahun, dan detektor spam otomatis menjadi teknologi yang begitu membosankan sehingga kita akan ditertawakan habis-habisan jika berani menyebutnya AI.

Faktanya, AI seolah-olah menguasai tugas-tugas yang menurut imajinasi kita seharusnya tidak dapat dilakukan oleh komputer. Setelah kita terbiasa dengan suatu teknologi dalam kehidupan sehari-hari, kita melepas lencana kehormatan AI dan mulai menyebutnya perangkat lunak komputer saja. Ini adalah fenomena yang telah dipelajari dengan baik yang disebut efek AI.

Karena efek AI, tujuan untuk apa yang kita sebut AI terus berubah secepat perkembangan teknologi. Definisi AI yang kita ambil dari pertimbangan ini adalah "label sementara untuk suatu perangkat lunak yang melakukan sesuatu yang keren dan mengejutkan, hingga kita terbiasa." Kami tidak tahu tentang Anda, tetapi itu rasanya bukan definisi yang memuaskan.

Kami harap kami telah meyakinkan Anda bahwa sangat sulit menemukan definisi yang memuaskan semua orang dan valid seiring perkembangan teknologi. Dengan mempertimbangkan efek AI, kami memutuskan untuk menghindari definisi AI yang sempit yang hanya memberi penghargaan kepada aplikasi "mencolok" hanya untuk membuangnya setelah hype-nya hilang. Kami merangkul definisi yang lebih luas yang mencakup aplikasi yang kurang mencolok. Inilah definisi AI kami:

Perangkat lunak yang memecahkan masalah tanpa instruksi manusia yang eksplisit.

Seperti yang Anda lihat, definisi kami berfokus pada hasil teknologi, alih-alih teknik spesifik yang digunakan untuk membangunnya. Beberapa orang mungkin tidak setuju, karena hampir setara dengan apa yang kami katakan tentang pembelajaran mesin sebelumnya di bab ini. Sebenarnya, pembelajaran adalah sifat cerdas, dan meskipun ML hanyalah sebuah alat, ia adalah alat di balik 99% aplikasi sukses yang kita sebut AI saat ini. Hal ini mungkin berubah di masa mendatang, tetapi kami tidak melihat adanya pendekatan baru di masa mendatang yang memiliki potensi yang sama seperti ML. Inilah sebabnya setiap aplikasi AI yang akan kita bahas dalam buku ini didasarkan pada ML: ini hanyalah gambaran paling akurat tentang lanskap AI saat ini dan di masa mendatang. Kita sekarang memiliki pandangan yang jelas tentang apa itu ML, definisi kerja AI modern, dan beberapa perspektif tentang bagaimana istilah-istilah ini berevolusi. Kita hanya melewati kata kunci ketiga yang mungkin pernah Anda dengar: ilmu data.

Ilmu data (DS) adalah bidang multidisiplin yang luas yang menggunakan metode dan proses ilmiah untuk menganalisis data dan mengekstrak wawasan. Teknik ML adalah beberapa alat dalam kotak peralatan DS. Dalam praktiknya, ketika orang merujuk pada proyek ilmu data, mereka sering kali memaksudkan sesuatu yang statis: mengekstrak wawasan dari data dan menyajikannya sebagai presentasi atau laporan. Di sisi lain, AI lebih umum digunakan dalam konteks perangkat lunak langsung.

Misalnya, menganalisis data lalu lintas untuk merancang rencana tata kota baru bagi sebuah kota guna meminimalkan kemacetan kemungkinan besar termasuk dalam ranah ilmu

data. Namun, jika Anda menggunakan data yang sama untuk mengendalikan lalu lintas secara real-time dan mengarahkan mobil melalui rute yang tidak terlalu padat, kebanyakan orang akan mengatakan bahwa proyek tersebut adalah tentang AI. Dalam kasus pertama, keluaran proyek Anda berupa laporan, dan dalam kasus kedua, berupa perangkat lunak "langsung" yang berjalan 24/7. Perlu diingat bahwa pembagian ini sebagian besar konvensional: tidak ada aturan baku tentang apa itu AI dan apa itu ilmu data. Tabel 1.1 merangkum perbedaannya sebagaimana yang kita lihat.

Tabel 1.1 Perbedaan utama antara AI dan ilmu data

Kecerdasan Buatan	Kecerdasan data
Mengotomatiskan tugas atau memprediksi kejadian masa depan berdasarkan data.	Menghasilkan wawasan berdasarkan data.
Umumnya digunakan "langsung": terus-menerus menguraikan data baru dan menghasilkan jawaban.	Umumnya "satu kali": menghasilkan beberapa wawasan yang menginformasikan keputusan.
Umumnya berbentuk perangkat lunak.	Umumnya berbentuk presentasi atau laporan.

Semoga bagian-bagian ini membantu mengungkap beberapa istilah yang sering disalahpahami dan menciptakan konteks untuk teknologi-teknologi ini. Sekarang Anda dapat mulai mempelajari prinsip-prinsip inti AI, apa yang berpotensi Anda lakukan dengannya, dan bagaimana membawa teknologi transformatif ini ke dalam organisasi Anda. Di bagian selanjutnya, kami akan menjelaskan langkah-langkah perjalanan ini dan bagaimana buku ini memandu Anda melaluinya.

1.4 METODE PENGAJARAN AI

Jika Anda ingin menggunakan AI secara produktif dalam kehidupan kerja Anda, sangat penting bagi Anda untuk memahami seluk-beluknya terlebih dahulu. Kami memperhatikan bahwa orang-orang non-teknis yang mendekati AI tanpa pemahaman yang mendalam tentang prinsip-prinsipnya sering kali berakhir dengan memimpikan proyek-proyek yang mustahil untuk dibangun, atau melewatkan hasil yang mudah dicapai yang sebenarnya dapat dengan mudah ditangani. Setelah bagian pertama buku ini, Anda akan mengetahui semua prinsip AI yang Anda butuhkan untuk menghindari jalan buntu ini dan mendapatkan hasil terbaik dari teknologi ini.

Bahkan setelah bab pertama ini saja, Anda sudah memahami bahwa hampir semua aplikasi AI modern bergantung pada pembelajaran mesin, dan pembelajaran mesin adalah tentang belajar dari data. Inilah mengapa kami menggunakan data sebagai panduan Anda untuk memahami AI. Setiap bab di bagian pertama buku ini berfokus pada satu jenis data spesifik, menunjukkan cara menemukannya di organisasi Anda, apa yang dapat Anda lakukan dengannya, dan bagaimana data tersebut sesuai dengan dunia AI.

Setiap bab menggunakan contoh mainan untuk memperkenalkan konsep ML yang Anda butuhkan. Kami menemukan ini sebagai cara paling efisien untuk mengajarkan konsep

ML yang jika tidak akan terlalu kering dan abstrak. Kami tidak mendalami aspek teknologi karena dua alasan sederhana:

- Teknologi berubah begitu cepat sehingga detail implementasinya akan segera menjadi usang.
- Sederhananya, Anda tidak membutuhkannya. Kecuali Anda ingin mengalihkan karier Anda ke penulisan kode, kami percaya ada nilai lebih dalam menambahkan AI ke dalam kekayaan pengetahuan Anda dan membiarkan orang lain secara praktis mengimplementasikan visi Anda dalam istilah komputer.

Ini tidak berarti kami akan sepenuhnya membebaskan Anda dari hal-hal teknis. Dari pengalaman kami sebagai insinyur, kami tahu bahwa mungkin sulit bagi tim teknis Anda untuk berkomunikasi dengan orang lain tanpa sedikit pun pemahaman teknis. Kami tidak ingin mereka kesulitan berbicara dengan Anda, jadi kami memastikan Anda mempelajari aspek teknis terpenting dari perangkat AI. Meskipun Anda menyerahkannya kepada tim Anda, memahaminya akan membantu Anda merencanakan dan mengelola upaya tersebut.

Setiap bab mencakup satu atau lebih kasus bisnis nyata tentang perusahaan yang mencapai hasil luar biasa. Sehubungan dengan penyebutan perusahaan, produk, atau layanan tertentu, perlu diingat bahwa kami melakukannya karena kami ingin Anda meningkatkan kesadaran, tetapi Anda tidak boleh merasa dibatasi oleh mereka. Kami tidak berafiliasi atau berkepentingan dengan perusahaan mana pun dalam studi kasus ini; kebetulan saja mereka sedang membangun produk-produk hebat yang dapat kita pelajari bersama.

Saat mempresentasikan kasus, kami mengikuti metodologi yang terinspirasi oleh metode kasus Harvard Business School: pertama-tama kami akan mempresentasikan kasus dengan cara yang paling netral, dan mengajukan pertanyaan terbuka di akhir. Setelah itu, kami menyertakan pemikiran kami tentang pertanyaan-pertanyaan ini dan pertanyaan-pertanyaan yang perlu didiskusikan lebih lanjut. Kami sarankan Anda untuk tidak langsung membaca jawaban-jawaban ini, tetapi cobalah pikirkan bagaimana Anda akan menjawab berdasarkan pengetahuan Anda dan apa yang telah Anda baca dalam kasus tersebut, baru kemudian baca pendapat kami. Perlu diketahui bahwa tidak ada solusi tunggal untuk pertanyaan-pertanyaan yang kami ajukan: jika Anda menemukan pandangan menarik tentang kasus-kasus yang tidak kami sertakan dalam jawaban, bagus! Ini berarti Anda telah mempelajari apa yang Anda butuhkan dan mampu mengekstrak wawasan sendiri (jadi jika itu terjadi, kami telah mencapai tujuan kami dengan buku ini juga).

- AI memiliki sejarah panjang keberhasilan dan kegagalan yang berawal dari tahun 1950an.
- AI umum adalah impian untuk memiliki mesin yang mahatahu. Semua aplikasi AI yang kita miliki saat ini justru sempit; mereka berfokus pada tugas-tugas spesifik.
- Pembelajaran mesin adalah cara yang umum untuk mengimplementasikan AI saat ini, dan didasarkan pada membiarkan mesin belajar secara mandiri dari data.
- Ilmu data terkait dengan AI dan ML, tetapi lebih berfokus pada penggalian wawasan daripada kecerdasan persisten.



Revolusi AI saat ini didasarkan pada pelatihan komputer untuk belajar dari data, dan itulah sebabnya kami memutuskan untuk menyusun buku ini berdasarkan berbagai bentuk dan wujud data. Setiap bab berfokus pada jenis data tertentu dan menggunakan contoh sederhana untuk membantu Anda mempelajari konsep-konsep kunci tentang AI. Di akhir setiap bab, Anda akan menemukan studi kasus dari perusahaan nyata yang telah menggunakan teknologi dan data yang telah kita bahas untuk mencapai hasil yang menakjubkan.

BAB 2

KECERDASAN BUATAN UNTUK DATA BISNIS INTI

Bab ini membahas

- Melihat data bisnis dalam hal kepadatan dolar dan dampak bisnis
- Menggunakan pembelajaran terawasi untuk memprediksi informasi penting
- Membandingkan pembelajaran mesin dengan rekayasa perangkat lunak konvensional
- Studi kasus: Menggunakan AI untuk menciptakan lini bisnis baru dan mengoptimalkan proses

Bab ini membahas data bisnis inti, jenis data yang paling dekat dengan proposisi nilai bisnis Anda. Data bisnis inti adalah kunci keberhasilan organisasi Anda. Data tersebut adalah data pesanan yang dikumpulkan oleh operasi komersial, interaksi manusia yang dilacak oleh jejaring sosial, dan pengukuran tekanan darah yang dilacak oleh perusahaan rintisan layanan kesehatan. Karena data bisnis inti merupakan inti dari organisasi, data tersebut juga merupakan kandidat utama untuk aplikasi AI pertama Anda. Menerapkan AI pada data bisnis inti seperti curang: data tersebut sangat dekat dengan mesin penghasil nilai organisasi sehingga proyek yang sukses hampir pasti akan memberikan dampak yang dramatis. Untuk menggugah selera Anda, mari kita bahas secara singkat dua studi kasus yang akan Anda temukan di akhir bab ini.

Studi kasus pertama membahas inisiatif Jim Gao, mantan karyawan Google yang bertanggung jawab atas operasional pusat datanya. Ia mengamati data yang dikumpulkan dari sistem pendingin udara besar yang digunakan untuk mendinginkan komputer raksasa Google dan berpikir untuk menggunakan pembelajaran mesin guna mengoptimalkan konsumsi dayanya. Hasilnya adalah pemotongan tagihan sebesar 40% bagi raksasa teknologi tersebut. Studi kasus kedua membahas Square, sebuah perusahaan jasa pembayaran yang berbasis di San Francisco. Square telah memproses pembayaran kartu kredit untuk usaha kecil dan menyadari bahwa semua data yang dikumpulkannya dapat digunakan untuk menawarkan pinjaman yang disesuaikan dan berisiko rendah kepada usaha kecil.

Dengan menggunakan model pembelajaran mesin, Square menciptakan lini bisnis yang benar-benar baru dan menggarap samudra biru pinjaman usaha kecil. Kualitas layanannya tak tertandingi: dengan mengotomatiskan proses verifikasi pemberi pinjaman, Square dapat menyetorkan pinjaman ke rekening bank pelanggan hanya dalam satu hari kerja setelah pengajuan. Hasilnya, Square telah menyalurkan pinjaman lebih dari Rp. 30 Triliun dalam empat tahun, dengan tingkat tunggakan yang sangat rendah, yaitu 4%. Di akhir bab ini, Anda akan mengetahui bagaimana perusahaan-perusahaan ini mewujudkan perubahan tersebut.

2.1 MEMANFAATKAN AI PADA DATA BISNIS INTI

Kami mendefinisikan data bisnis inti sebagai "data yang berdampak langsung pada lini atas atau bawah organisasi." Data inti terlihat sangat berbeda tergantung pada apa yang

dilakukan organisasi Anda: riwayat keranjang belanja untuk operasi e-commerce, pengukuran fisik untuk organisasi teknik, dan perilaku pasien untuk perusahaan layanan kesehatan. Apa pun bentuknya, data inti berharga karena menggambarkan peristiwa dan pola yang berdampak langsung pada kinerja organisasi, dan mudah untuk dikaitkan dengan nilai moneter.

Kami telah menyebutkan dua studi kasus yang menanti Anda di akhir bab ini. Mari kita lihat mengapa data yang digunakan oleh kedua perusahaan tersebut merupakan data bisnis inti:

- Perusahaan seperti Google bergantung pada pusat data besar untuk menawarkan layanannya (memproses pencarian web, menyimpan foto, merutekan email, dan sebagainya). Mungkin satu-satunya biaya variabel bagi Google adalah energi yang dihabiskan untuk menjaga komputer pusat data tetap dingin. Oleh karena itu, data tentang instalasi pendinginnya merupakan inti dari bisnis Google, karena berkorelasi langsung dengan salah satu biaya utamanya.
- Produk inti Square adalah solusi *point-of-sale* (POS). Melalui produk tersebut, Square memproses semua pembayaran yang masuk ke pelanggannya. Karena misi Square adalah memberdayakan usaha kecil, data transaksinya terkait erat dengan visi tersebut dan berharga bagi pelanggannya.

Cara yang baik untuk melihat nilai data adalah dengan memikirkan metrik yang kita sebut kepadatan dolar data: seberapa besar data memengaruhi laba atas atau laba bawah organisasi. Data bisnis inti memiliki kepadatan rupiah yang tinggi: setiap pesanan e-commerce, prospek pekerjaan, atau transaksi keuangan memiliki dampak langsung pada laba atas atau laba bawah Anda. Saat Anda menjauh dari proposisi nilai inti organisasi Anda, kepadatan dolar data yang Anda kumpulkan akan menurun. Mencatat kunjungan ke situs web Anda memang berharga, tetapi tidak seberharga melacak pesanan yang masuk. Bagi rumah sakit, data pusat panggilan mungkin bermanfaat, tetapi tidak seberharga catatan pasien. Inilah sebabnya kami memutuskan untuk memulai buku ini dengan data dengan kepadatan dolar tertinggi: data bisnis inti Anda.

Sering kali kita melihat bahwa data bisnis inti memiliki bentuk terstruktur, seperti baris dan kolom rapi pada lembar kerja Microsoft Excel. Dalam istilah teknik, kami menyebut jenis data ini terstruktur. Contoh lain dari data terstruktur adalah laporan cuaca, pengukuran proses fisik, transaksi keuangan, sebagian besar pasar, serta metrik rantai pasokan dan pergudangan. Sebagai aturan praktis, apa pun yang dapat Anda muat di Excel kemungkinan besar termasuk dalam payung data terstruktur. Jenis data lain lebih sulit untuk dimasukkan ke dalam kolom Excel yang rapi: bayangkan gambar, rekaman suara, atau teks dalam buku.

Penting untuk dipahami bahwa informasi yang sama dapat ada dalam bentuk terstruktur dan tidak terstruktur. Misalnya, pertimbangkan cara berikut untuk mencatat diagnosis medis:

Pasien Gianluca Mauro mengalami peradangan parah pada sendi bahu; terapinya adalah mengonsumsi dua pil Kortison per hari selama lima hari.

Informasi yang sama dapat dicatat secara terstruktur, seperti pada Tabel 2.1.

Tabel 2.1 Diagnosis medis dalam bentuk terstruktur

Paten	Diagnosa	Daerah	Pengobatan	Frekuensi (kali/hari)	Durasi terapi (hari)
Gianluca Mauro	Peradangan parah	Sendi bahu	Kortison	2	5

Informasinya sama tetapi dalam kasus pertama, direpresentasikan secara tidak terstruktur (teks), sedangkan dalam kasus kedua, terstruktur. Data terstruktur jauh lebih mudah diproses oleh komputer daripada data tidak terstruktur. Bab-bab selanjutnya membahas teknik AI untuk data tidak terstruktur, termasuk gambar dan bahasa tulis. Untuk saat ini, mari kita tetap menggunakan data terstruktur, karena sebagian besar data bisnis inti termasuk dalam kategori ini.

2.2 MENGGUNAKAN AI DENGAN DATA BISNIS INTI

Setelah Anda mengetahui cara menemukan dan mengenali data bisnis inti, mari kita lihat apa yang dapat dilakukan AI dengannya. Karena kami percaya pada kekuatan pengajaran cerita, kami mengembangkan contoh sederhana yang akan menemani Anda di sepanjang bagian pertama buku ini. FutureHouse adalah bisnis fiktif yang mengoperasikan pasar real estat daring tempat pemilik rumah dapat mengiklankan rumah mereka untuk dijual dan diharapkan menarik minat pembeli. FutureHouse selalu memprioritaskan layanan pelanggan dan mempekerjakan agen yang dapat menawarkan dukungan mereka untuk membantu penjual menilai harga rumah mereka. Pembeli menggunakan situs ini untuk mencari rumah impian mereka.

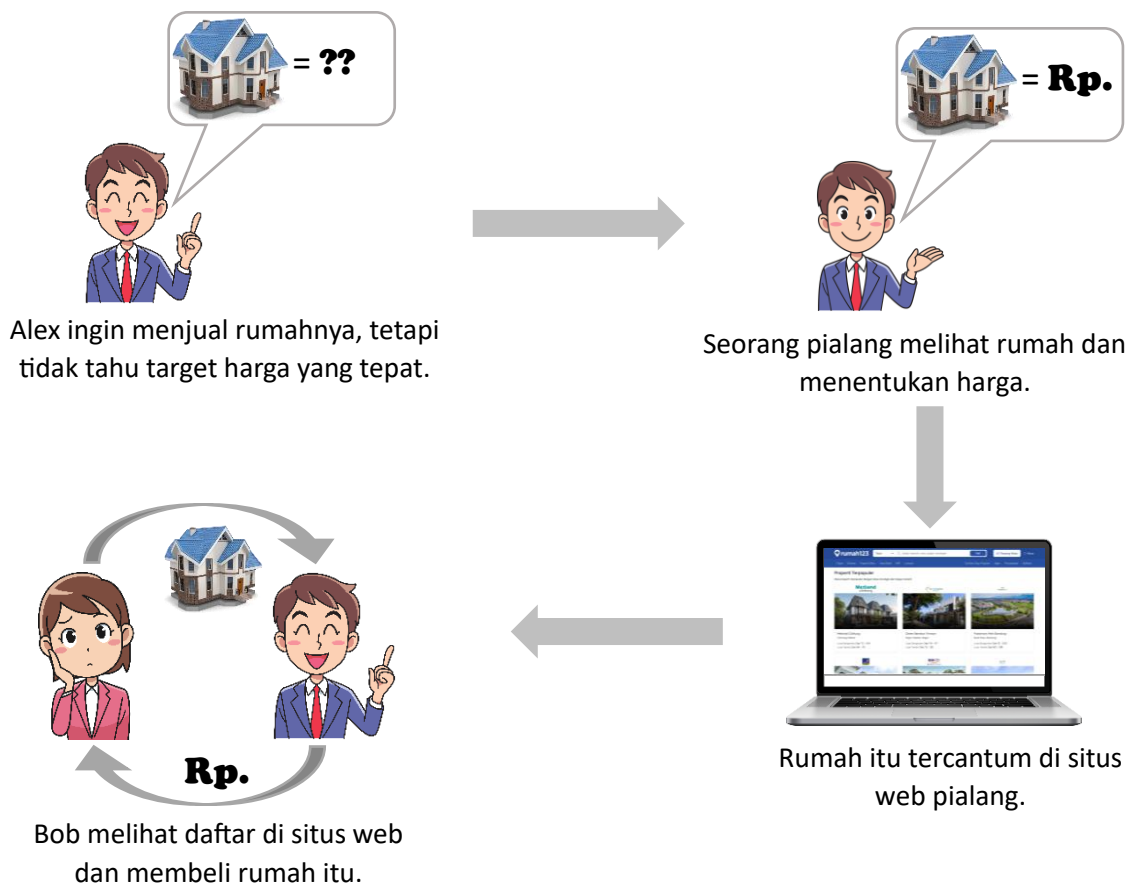
Kami memilih contoh ini karena akan familier bagi banyak pembaca, sekaligus memberi kami kesempatan untuk mengeksplorasi berbagai aspek lanskap AI. Bonusnya adalah pasar perumahan merupakan contoh umum dalam literatur ML. Jika Anda memutuskan untuk mendalami teknis ML lebih dalam dan mungkin menulis kode Anda sendiri, Anda akan menemukan banyak referensi di web.

Kami akan mulai dengan menjelaskan sedikit tentang cara kerja pasar perumahan, lalu kami akan memperkenalkan penerapan ML dalam bisnis ini. Kami akan menggunakan deskripsi ini sebagai alasan untuk memperkenalkan terminologi dan konsep kunci, dan menyimpulkan dengan memperluas konsep-konsep ini ke aplikasi lain yang lebih umum.

Contoh Pasar Real Estat

Agan real estat berada di pusat pasar perumahan. Mereka menghubungkan penjual dan pembeli serta membantu mereka menentukan harga yang tepat untuk properti mereka. Gambar 2.1 menunjukkan bagaimana transaksi penjualan rumah pada umumnya berlangsung di bawah pengawasan ketat seorang agen properti:

1. Seorang pelanggan datang untuk menjual rumah mereka. Sebelum memasarkannya, mereka menginginkan pendapat profesional tentang nilainya.
2. Agen properti memeriksa rumah dan menetapkan harga berdasarkan luas bangunan, usia, peralatan yang disertakan, penawaran lain di lingkungan tersebut, dan sebagainya.
3. Rumah tersebut terdaftar.
4. Calon pembeli menemukan listing tersebut dan datang untuk acara open house. Satu pembeli yang beruntung akhirnya berhasil membeli properti tersebut.



Gambar 2.1 Proses mendaftarkan dan menjual rumah di pasaran

FutureHouse mengumpulkan data di setiap langkah proses ini. Berikut aset data utamanya:

- Foto semua rumah yang dijual
- Deskripsi rumah (luas, lokasi, dan sebagainya)
- Riwayat kunjungan dan klik web untuk semua pengguna situs web
- Catatan historis properti yang terjual di wilayah tersebut
- Ulasan rumah yang ditinggalkan oleh orang-orang yang pernah mengunjunginya
- Transkrip negosiasi antara agen dan pembeli
- Artikel surat kabar tentang pembangunan yang akan datang di wilayah tersebut

Semua sumber data ini tidak memiliki nilai yang sama; kepadatan dolarnya berbeda. Misalnya, artikel surat kabar merupakan kumpulan data yang sangat umum yang dapat diakses oleh

setiap perusahaan lain. Oleh karena itu, meskipun mungkin berharga, hal itu bukanlah inti dari FutureHouse. Kunjungan situs web, gambar, dan ulasan mungkin penting karena bersifat spesifik dan unik untuk bisnis tersebut, sehingga kepadatan dolarnya lebih tinggi daripada artikel surat kabar. Namun, data tersebut tidak berdampak langsung pada kesuksesan bisnis. Catatan historis rumah yang terjual jauh lebih terkait dengan kinerja bisnis dibandingkan semua sumber data lainnya, karena model bisnis tradisional agen properti adalah biaya keberhasilan per rumah yang terjual: ini adalah data inti bisnis Anda.

Misalnya, kita ingin memanfaatkan data ini untuk membangun alat evaluasi properti bertenaga AI yang secara otomatis memprediksi harga terbaik untuk rumah yang terdaftar di platform. Estimasi harga biasanya dilakukan oleh agen yang berpengalaman di area tersebut dan rumah-rumah serupa. Mereka telah mengawasi penjualan banyak properti sebelumnya, mengembangkan intuisi yang mendalam tentang harga wajar properti tersebut. Inilah tugas yang kita inginkan untuk diambil alih oleh AI.

Mari kita coba menyusun daftar faktor yang lebih lengkap yang mungkin dipertimbangkan agen sebelum menyarankan harga jual:

- Luas bangunan
- Jumlah lantai
- Jumlah kamar tidur
- Jumlah kamar mandi
- Tahun pembangunan
- Apakah ada kolam renang?
- Apakah ada garasi?
- Efisiensi energi
- Lokasi
- Kualitas lingkungan
- Transportasi umum

Jika Anda bertanya bagaimana agen menentukan harga, mereka mungkin akan menjawab seperti "berdasarkan pengalaman." Mereka mungkin merujuk pada pedoman standar seperti harga rata-rata per kaki persegi di lingkungan tersebut, tetapi sulit bagi mereka untuk menjelaskan bagaimana faktor-faktor lain memengaruhi keputusan mereka. Jika Anda tidak dapat menjelaskan proses berpikir Anda, Anda bahkan tidak dapat mengkodekannya ke dalam program komputer.

Seorang agen belajar cara menilai rumah dengan melihat harga jual banyak properti dari waktu ke waktu. Gagasan utama di balik ML adalah bahwa komputer dapat melakukan hal yang sama: ia dapat mempelajari cara menilai rumah dari data historis rumah yang terjual sebelumnya, memasukkannya dengan karakteristik rumah (ukuran, lingkungan, dan sebagainya) dan harga jual akhir.

AI ini memungkinkan Anda memiliki keahlian yang sama dengan broker berpengalaman, dengan kecepatan, skala, ketersediaan, dan biaya perangkat lunak yang sama, sehingga setiap penjual rumah dapat memperkirakan nilai properti mereka dalam sekejap. Perpaduan antara ketersediaan data dan nilai bisnis ini merupakan tujuan utama inovasi

berbasis AI. Kami telah mengidentifikasi proposisi nilai yang jelas, didukung oleh jenis data yang tepat untuk membangunnya.

Menambahkan Kemampuan AI ke FutureHouse

Salah satu tujuan buku ini adalah agar Anda mempelajari cara berkomunikasi dengan ilmuwan dan insinyur data, jadi di paragraf inilah Anda akan terbiasa dengan istilah dan detail teknisnya. Pada tahap proses ideasi ini, kami telah memutuskan untuk membangun mesin prediksi harga rumah otomatis. Kami yakin ini akan membantu mengoptimalkan proses internal kami dan menawarkan nilai kepada pelanggan kami, karena mereka dapat memperoleh penawaran harga rumah mereka secara langsung secara daring. Kami juga memiliki catatan transaksi di wilayah tersebut, termasuk beberapa fitur penting rumah (misalnya, luas persegi) dan harga jual. Jika Anda menyampaikan semua ini kepada seorang insinyur, beginilah kira-kira percakapan hipotetisnya:

- Anda : Kami ingin menambahkan prediktor harga otomatis ke situs web listing properti kami. Bisakah Anda membantu kami?
- Insinyur : Bagus, ini adalah masalah standar dalam pembelajaran mesin. Ada beberapa algoritma pembelajaran mesin terkenal yang saya cukup yakin dapat menghasilkan model yang baik. Apa saja masukan yang kami miliki?
- Anda : Kami telah berbicara dengan para agen, dan mereka biasanya memperhitungkan luas persegi, jumlah kamar mandi, jarak ke transportasi umum, dan hal-hal seperti itu. Berikut daftar lengkapnya.
- Insinyur : Tentu, daftar fitur ini sepertinya awal yang baik. Dan target apa yang Anda butuhkan dari model ini? Hanya harga properti yang diharapkan?
- Anda : Benar, hanya itu yang kami butuhkan. Harganya akan membantu kami menawarkan penawaran harga real-time di situs web kami secara gratis.
- Insinyur : Keren. Data apa yang Anda miliki untuk ini?
- Anda : Kami memiliki catatan untuk semua transaksi rumah di wilayah ini selama dekade terakhir.
- Insinyur : Hebat! Apakah catatan tersebut mencakup semua fitur yang Anda sebutkan sebelumnya?
- Anda : Ya, sebagian besar. Kami tentu memiliki luas persegi dan hal-hal seperti itu, tetapi Anda mungkin harus memikirkan sesuatu untuk transportasi umum.
- Insinyur : Oke, saya bisa mengerjakannya. Apakah Anda juga memiliki label untuk semuanya?
- Anda : Ya, catatan tersebut mencakup harga jual akhir setiap properti.
- Insinyur : Luar biasa. Ada berapa banyak contoh yang kita bicarakan? Anda: Kami memiliki sekitar 200.000 transaksi selama lima tahun terakhir.
- Insinyur : Fantastis! Itu seharusnya cukup untuk melatih model kita.

Mari kita mundur sedikit dan merumuskan kembali apa yang baru saja Anda lalui. Pertama-tama, insinyur tersebut menyebut proyek Anda sebagai masalah pembelajaran mesin. Kata masalah dalam bidang seperti ML tidak memiliki konotasi negatif; Istilah ini

digunakan untuk merujuk pada tugas-tugas yang terdefinisi dengan baik yang memiliki tujuan spesifik dan kondisi yang diketahui. Tujuan proyek ini adalah membangun sebuah model: sebuah program komputer otomatis yang dapat memperkirakan nilai properti untuk area spesifik kita. Meskipun cara kerjanya jauh lebih kompleks, sebuah model seperti kalkulator saku: ia memiliki serangkaian masukan, memprosesnya melalui algoritma pembelajaran mesin, dan kemudian menghasilkan target yang kita inginkan. Banyak orang juga suka mengatakan bahwa model memprediksi target, atau membuat prediksi (misalnya, Anda akan sering membaca tentang kualitas prediksi sebuah model). Dalam kasus kita, masukannya adalah serangkaian fitur yang menggambarkan setiap rumah: luas persegi, jumlah kamar, dan sebagainya.

Insinyur juga menanyakan tentang jumlah contoh, yang berarti jumlah penjualan rumah sebelumnya yang kita miliki dalam catatan kita. Ketika kita berbicara tentang nilai yang diharapkan dari keluaran untuk contoh tertentu (rumah), kita biasanya menyebutnya label, seolah-olah kita memberi label setiap contoh dengan nilai keluaran yang benar. Mudah untuk memvisualisasikannya dalam lembar Excel seperti pada gambar 2.2.

The diagram illustrates an Excel spreadsheet with five rows of data. The first row contains headers: 'A', 'B', 'C', and 'D'. The second row contains labels: 'Kamar', 'Meter Persegi', 'Jarak dari Pusat Kota', and 'Harga'. The following three rows contain numerical and text data. To the left of the table, the word 'Contoh' has arrows pointing to each of the five rows. Above the table, the word 'Fitur' has arrows pointing to columns A, B, and C, while 'Target/Label' has an arrow pointing to column D.

	A	B	C	D
1	Kamar	Meter Persegi	Jarak dari Pusat Kota	Harga
2	2	80	5.4 Km	100.000
3	1	42	7 Km	80.000
4	3	120	23 Km	160.000
5	2	65	2 Km	70.000

Gambar 2.2 Lembar Excel berisi fitur dan label untuk beberapa contoh

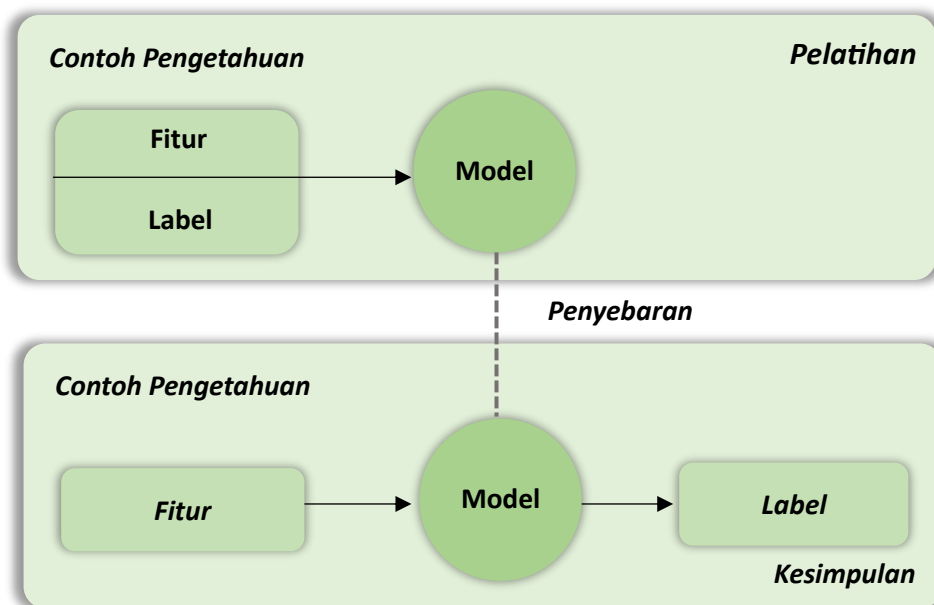
Insinyur pembelajaran mesin kami meminta data historis karena mereka harus menginstruksikan komputer untuk menjalani proses yang disebut pelatihan. Sebagaimana agen properti berpengalaman dapat menebak harga rumah baru berdasarkan pengalaman mereka menjual rumah di masa lalu, algoritma pembelajaran mesin dapat mempelajari cara memprediksi harga rumah berdasarkan contoh-contoh sebelumnya dalam fase yang disebut pelatihan. Pada fase awal ini, para insinyur menulis kode untuk memasukkan ribuan contoh yang kami miliki ke dalam model. Setiap contoh berisi fitur (karakteristik rumah yang akan dipelajari) dan labelnya (harga jual), sehingga algoritma dapat belajar dari pengalaman masa lalu ini. Setelah pelatihan selesai, kami memiliki program komputer mandiri (model) yang mengemas semua kekuatan prediktif prediksi harga rumah. Karena model sekarang menanamkan pengetahuan yang terdapat dalam set data pelatihan, kami dapat mulai menggunakannya secara independen dari data pelatihan.

Kapan pun kami ingin mengevaluasi rumah baru, kami dapat memasukkan fitur-fiturnya ke dalam model dan mendapatkan jawaban tentang kemungkinan harganya di

pasaran. Langkah kedua ini, yang disebut inferensi, menjadikan model bermanfaat bagi organisasi karena menghasilkan estimasi rumah baru yang terdaftar di platform kami. Setelah model dilatih, kami dapat meminta para insinyur untuk menghubungkannya ke situs web FutureHouse dan mulai menjawab permintaan pelanggan yang masuk.

Wajar jika batasan antara pelatihan dan inferensi masih agak kabur bagi Anda. Lagipula, manusia menjalani hidup dengan berlatih pengetahuan baru dan menyimpulkan keyakinan masa lalu secara bersamaan: teman agen real estat kita telah mengevaluasi properti dan belajar dari setiap properti yang terjual secara bersamaan sejak awal karier mereka. Namun, penting bagi Anda untuk menyadari bahwa keduanya merupakan langkah yang sangat terpisah dalam dunia pembelajaran mesin. Setelah dataset siap, pelatihan tidak lebih dari sekadar penyambungan mekanis antara data dan berbagai model yang mungkin diputuskan untuk diuji oleh seorang insinyur.

Nilai sebenarnya dihasilkan selama inferensi, ketika kita dapat menggunakan semua pengalaman masa lalu tersebut untuk membuat prediksi baru. Perbedaan antara pelatihan dan inferensi ini diungkapkan secara grafis pada Gambar 2.3.



Gambar 2.3 Dua fase pembelajaran mesin: pelatihan dan inferensi

Sejauh ini, kami telah berhasil merumuskan kembali masalah prediksi harga properti kami secara umum, yang akan berguna ketika kami menangani masalah lain. Kami meminta para insinyur untuk membangun model yang dapat memetakan input (alias fitur) ke nilai target (harga rumah). Kami juga akan memberikan beberapa data, dalam bentuk daftar panjang contoh (penjualan sebelumnya), masing-masing dengan beberapa fitur (luas persegi dan angka lainnya) dan label (harga jual).

Kami ingin model tersebut dapat memprediksi harga rumah baru yang akan dipasarkan, dan melakukannya dengan cara yang konsisten dengan contoh yang kami berikan. Dengan kata lain, kami berharap model tersebut akan menetapkan harga flat dua kamar tidur

baru di lingkungan Tribeca, New York, serupa dengan harga flat dua kamar tidur lain yang telah terjual di dekatnya tahun lalu. Kita telah memperkenalkan beberapa abstraksi sejauh ini, jadi mari kita berhenti sejenak dan merangkum:

- *Algoritma pembelajaran mesin*—Teknologi yang memungkinkan komputer belajar dari data
- *Fitur*—Seperangkat karakteristik suatu objek yang dapat dipelajari oleh algoritma
- *Label*—Keluaran atau target yang ingin kita prediksi oleh algoritma
- *Pelatihan*—Fase di mana algoritma pembelajaran mesin diberi contoh-contoh sebelumnya dan belajar darinya
- *Model*—Keluaran dari fase pelatihan—Program komputer mandiri yang dapat memprediksi label, dengan serangkaian fitur
- *Inferensi*—Fase di mana model digunakan dengan contoh-contoh baru

Hal yang hebat dari semua abstraksi ini adalah kita sekarang dapat menerapkannya pada masalah lain. Misalnya, katakanlah kita ingin memprediksi harga mobil bekas, alih-alih rumah.

Bisakah Anda memikirkan fitur apa yang akan Anda gunakan? Model dan jarak tempuh mobil mungkin merupakan tebakan yang bagus, begitu pula riwayat perawatannya. Namun, Anda dapat memiliki imajinasi yang lebih luas: pikirkan tentang subsidi pemerintah untuk kendaraan rendah polusi, jumlah pencarian Google untuk model-model tersebut, dan banyak lagi. Secara lebih umum, semua masalah yang ingin kita prediksi nilainya (yang belum kita ketahui) berdasarkan serangkaian informasi yang kita miliki adalah contoh pembelajaran terawasi.

Pembelajaran terawasi adalah area pembelajaran mesin yang paling banyak diterapkan dalam industri dan penelitian saat ini. Contoh kita sejauh ini termasuk dalam subset ini: ia menggunakan serangkaian teknik yang memungkinkan mesin mempelajari pemetaan antara serangkaian informasi yang disebut fitur dan nilai target yang disebut label. Keindahan pembelajaran terawasi, dan pembelajaran mesin secara umum, adalah komputer akan secara otomatis menemukan rumus ini, betapa pun rumitnya. Model yang berhasil akan membuat prediksi yang konsisten dengan label, sehingga mentransfer pengalaman yang tertanam dalam kumpulan data ke kasus-kasus baru. Gambar 2.4 menunjukkan bagaimana fitur dan label berinteraksi dalam aplikasi pembelajaran terawasi.

Sekarang, Anda telah mempelajari konsep dasar di balik pembelajaran terawasi. Seperti yang Anda lihat, ini bukan ilmu roket. Sebenarnya mudah dipahami. Kompleksitasnya terletak pada cara kerja internalnya: algoritma pembelajaran mesin, mesin yang memungkinkan komputer melakukan pembelajaran. Dalam buku ini, kami tidak akan membahas algoritma ini secara mendalam kecuali kami menganggapnya menarik dan bermanfaat bagi Anda. Dengan perangkat dan layanan masa kini, kami yakin Anda dapat menghasilkan nilai bagi organisasi Anda hanya dengan pengetahuan yang cukup untuk menghasilkan produk atau layanan AI dan menyerahkannya kepada insinyur ML ketika saatnya tiba. Sekarang Anda mulai tahu cara melakukan keduanya, jadi selamat!



Gambar 2.4 Konsep inti pembelajaran terawasi: menemukan pemetaan antara sekumpulan fitur dan label

Meskipun konsep sebelumnya tampak sederhana, jangan remehkan pengetahuan baru Anda: miliaran rupiah telah dihasilkan oleh perusahaan yang menggunakan teknik sederhana yang baru saja Anda pelajari. Lalu, mengapa pembelajaran mesin begitu hebat? Inilah topik bagian selanjutnya.

Keunggulan Pembelajaran Mesin

Di bagian sebelumnya, kita berbincang menyenangkan dengan seorang insinyur imajiner, dan kami berpisah karena tahu bahwa mereka akan membangun sebuah model untuk memprediksi harga rumah berdasarkan beberapa fiturnya, seperti luas bangunan. Abaikan dulu masalah prediksi harga, dan bayangkan kita menginginkan sebuah kalkulator untuk pajak properti yang harus dibayar pembeli. Jika demikian, percakapan kita dengan insinyur tersebut akan jauh lebih singkat:

Anda : Kami ingin menambahkan kalkulator pajak properti ke situs web kami. Bisakah Anda membantu?

Insinyur : Keren. Apa bagian yang relevan dalam undang-undang perpajakan?

Anda : Anda dapat merujuk ke 4123 dan 564b.

Insinyur : Bagus, hanya itu yang saya butuhkan.

Perbedaan utama antara membangun prediktor harga dan kalkulator pajak adalah bahwa kalkulator pajak diatur oleh aturan-aturan presisi yang sudah dinyatakan dalam bentuk matematika. Bahkan akuntan paling junior pun dapat menerapkan rumus matematika yang menentukan pajak properti dan menerapkannya pada properti baru, dengan sedikit kebutuhan akan pengalaman.

Hal yang sama berlaku untuk fisika: sejak Isaac Newton menemukan hukum dinamika pada tahun 1600an, kita telah mampu memprediksi gerakan benda dan membangun roket untuk mengirim manusia ke bulan. Kapan pun kita menghadapi situasi seperti ini, matematika dan ilmu komputer konvensional akan berfungsi dengan baik: kita tinggal menerjemahkan aturan dan rumus ke dalam kode komputer, dan hasilnya akan 100% akurat.

Hal ini tidak berlaku untuk contoh prediksi harga: hubungan antara fitur-fitur rumah dan harga jualnya tidak jelas, kabur, dan jelas tidak tertulis dalam buku aturan. Bahkan agen yang paling berpengalaman pun tidak mungkin dapat menuliskan aturan yang mudah

dipahami komputer untuk menilai properti, dan seorang programmer pun tidak mungkin dapat menerjemahkannya ke dalam perangkat lunak.

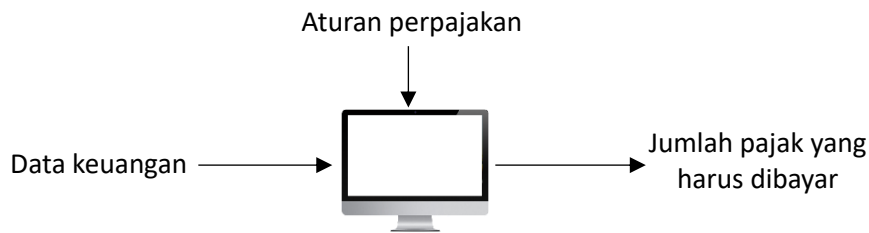
Pada abad yang lalu, umat manusia menggunakan komputer untuk mewujudkan tujuan-tujuan yang luar biasa, bahkan sampai meluncurkan roket ke luar angkasa dan mengatur miliaran transaksi keuangan setiap hari. Namun, robot kita yang paling canggih pun masih belum dapat menyeberang jalan seaman anak berusia enam tahun. Faktanya, ilmu komputer telah lama terjebak pada banyak masalah penting karena terlalu rumit untuk dipahami secara analitis. Menerjemahkan puisi ke dalam bahasa asing, mengendarai mobil, atau menjawab telepon adalah tugas-tugas yang dipelajari manusia melalui pengalaman, alih-alih dengan mengikuti instruksi. Pembelajaran mesin akhirnya memungkinkan komputer untuk melakukan hal yang sama, mengabaikan aturan matematika sepenuhnya dan dengan demikian memungkinkan domain aplikasi yang benar-benar baru. Wawasan mendasarnya adalah kita dapat menggunakan pengalaman untuk mengatasi kekurangan aturan dan hubungan matematis. Dengan kata lain, model dikembangkan dengan mempelajari informasi relevan dari contoh-contoh sebelumnya. Hal ini berbeda dengan pemrograman tradisional, di mana model dikembangkan oleh seseorang dan dengan susah payah dikodekan ke dalam komputer (yaitu, diprogram).

Gambar 2.5 merepresentasikan perbedaan mendasar ini. Ide inti dari pembelajaran mesin adalah kita dapat menggunakan fitur dan label dalam data agar komputer secara otomatis mempelajari hubungan di antara keduanya. Setelah menyimpan informasi ini, tujuannya adalah agar model mereplikasi hubungan yang sama pada contoh yang belum pernah dilihat sebelumnya (rumah baru), memprediksi label (harga rumah), dengan serangkaian fitur (karakteristik rumah). Konsep umumnya tidak jauh berbeda dengan apa yang biasanya dilakukan manusia: anak-anak perlu diajari perbedaan antara kuda dan kuda poni hanya beberapa kali sebelum mereka "memahaminya" dan menginternalisasi perbedaannya.

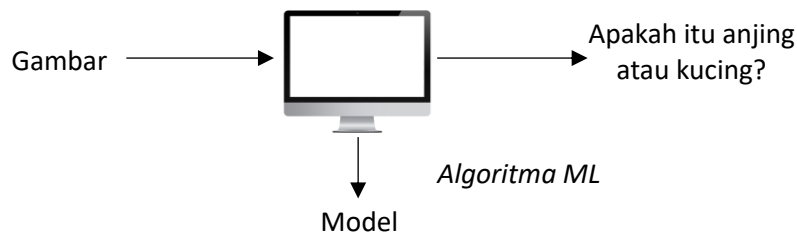
Dibandingkan dengan rekayasa perangkat lunak konvensional, pembelajaran mesin memiliki dua keunggulan. Kita sudah membahas yang pertama: ada beberapa masalah yang tidak bisa kita selesaikan dengan komputer dengan cara lain. Alasan kedua untuk menggunakan pembelajaran mesin adalah karena perangkat lunak bersifat statis: meskipun para insinyur dapat membuat aturan logika yang kompleks, aturan tersebut tidak dapat berubah setelah produk dikirimkan. Dengan pembelajaran mesin, menambahkan data pelatihan baru dapat meningkatkan kinerja dan memecahkan situasi yang benar-benar baru, tanpa memerlukan upaya rekayasa tambahan. Misalnya, ketika gelombang kondominium baru terjual, kita cukup mengulangi pelatihan dengan data baru, dan prediksinya akan disesuaikan.

Kedua keunggulan ini telah menjadi kunci untuk menghasilkan nilai dengan pembelajaran mesin dalam dekade terakhir. Setelah Anda memahaminya, mari kita lihat kasus-kasus lain yang dapat menerapkannya dan contoh nyata perusahaan-perusahaan yang meraih keuntungan berkat pembelajaran mesin.

Masalah pemrograman tradisional dengan aturan dan hubungan yang diketahui



Pembelajaran mesin-masalah yang tidak dapat dijelaskan secara analitis tetapi dapat dipelajari dari pengalaman.



Gambar 2.5 Perbedaan antara metode pemrograman tradisional dan pembelajaran mesin

Menerapkan AI pada Data Bisnis Inti Umum

Anda mungkin bisa menebak bahwa AI dapat diterapkan ke lebih banyak industri daripada real estat. AI dapat diterapkan ke banyak domain. Bagian ini akan membantu Anda memperluas wawasan dengan menggeneralisasi alat-alat dasar yang diperkenalkan dalam bab ini. Algoritma prediksi harga kami termasuk dalam cabang pembelajaran mesin yang disebut pembelajaran terawasi, seperti yang telah kami sebutkan sebelumnya. Algoritma pembelajaran terawasi jelas tidak terbatas pada menemukan hubungan antara fitur rumah dan harga jualnya. Keunggulan algoritma ini adalah bahwa mereka merupakan mesin pembelajaran umum yang dapat diadaptasi untuk hampir semua fenomena di mana fitur memengaruhi nilai suatu label.

Mengganti label adalah cara paling cepat untuk memahami fleksibilitas algoritma pembelajaran terawasi. Katakanlah kita mengganti label kita saat ini (harga jual) dengan waktu yang dibutuhkan untuk menjual rumah tersebut. Metrik ini juga mudah diperoleh dari catatan historis kami, dan akan berguna bagi pelanggan dan agen sehingga mereka dapat mengukur permintaan rumah dan bertindak sesuai dengan itu. Tampaknya masuk akal jika rumah mewah yang besar akan terjual lebih lambat daripada apartemen dua kamar tidur. Bahkan dengan label yang berbeda, kami kemungkinan dapat menggunakan kembali serangkaian fitur yang sama yang kami identifikasi untuk prediksi harga untuk menghasilkan algoritma baru yang, alih-alih memprediksi harga terbaik untuk sebuah rumah, akan memprediksi waktu yang dibutuhkan untuk menemukan pembeli.

Jika setiap kali Anda membuka lembar kerja Excel, Anda melihat kolom-kolom dalam hal fitur dan label potensial untuk masalah pembelajaran mesin, dunia peluang akan terbuka

untuk Anda. Misalnya, jika Anda bekerja di bank dan terbiasa melihat lembar Excel berisi data pelanggan, Anda sekarang dapat melihat bagaimana Anda dapat menggunakan data tersebut sebagai fitur dan label. Bank terkenal menggunakan pembelajaran terawasi untuk memprediksi risiko kredit pelanggan mereka sebelum memberikan pinjaman, menggunakan fitur-fitur seperti pendapatan, usia, dan status perkawinan. Jika Anda bekerja di perusahaan pemeliharaan dan memiliki beberapa data tentang kegagalan peralatan, Anda tidak perlu heran bahwa beberapa perusahaan dapat memprediksi kemungkinan kerusakan peralatan dalam bulan berikutnya dengan menggunakan fitur-fitur seperti usia dan lamanya waktu sejak perbaikan terakhir. Anda sekarang memiliki semua pengetahuan yang dibutuhkan untuk memahami dan merumuskan masalah seperti ini, dan Anda juga akan mendapatkan lebih banyak latihan dengan studi kasus di akhir bab ini.

Dalam bab-bab berikutnya, kami akan membahas teknologi ini dalam dua cara utama: melihat masukan atau keluaran model. Mengenai masukan, kami akan mengajarkan Anda tentang teknik AI yang dapat menggunakan gambar atau teks sebagai masukan, yang sangat memperluas jenis masalah yang dapat kami tangani. Mengenai keluaran, kami akan membahas keluarga model yang tidak hanya memprediksi label yang cocok dengan data pelatihan, tetapi juga melakukan hal-hal bermanfaat lainnya seperti mengelompokkan item serupa, atau membantu pelanggan menemukan konten yang menarik.

Sekarang saatnya untuk mengakhiri bagian teknis dari bab ini, jadi mari kita rangkum kembali konsep-konsep terpenting. Anda telah mempelajari bahwa keajaiban pembelajaran mesin terletak pada penggantian pemahaman analitis tentang proses kompleks (seperti pasar real estat) dengan model yang telah belajar secara otomatis dari titik data. Selama contoh-contoh tersebut merupakan representasi yang baik dari perilaku sistem, kita dapat menggunakan pengetahuan yang tertanam dalam model untuk memprediksi peristiwa yang tak terlihat (seperti rumah baru yang akan dipasarkan). Anda telah melihat bahwa ada beberapa pekerjaan yang harus kita lakukan sebelumnya, dalam mendefinisikan visi bisnis dan mengidentifikasi data. Kemudian, kita siap meminta para insinyur untuk melatih model tersebut. Setelah pelatihan selesai, kita mendapatkan model yang telah selesai yang dapat kita gunakan untuk inferensi, yang akhirnya menghasilkan nilai bisnis.

2.3 STUDI KASUS

Bagian ini menyajikan dua studi kasus dari dua perusahaan yang mengembangkan dua aplikasi AI yang sangat berbeda, tetapi memiliki banyak kesamaan: keduanya menggunakan data bisnis inti dan pembelajaran terawasi untuk meraih keunggulan kompetitif yang luar biasa.

Studi kasus pertama berasal dari Google, yang menggunakan data bisnis inti dari konsumsi energi pusat datanya untuk memangkas tagihan energi hingga 40%. Tokoh utama dalam studi kasus kedua adalah Square, yang menggunakan data bisnis inti dari transaksi POS pelanggannya untuk memenangkan samudra biru pinjaman usaha kecil.

Bagaimana Google Menggunakan AI untuk Memangkas Tagihan Energinya

Pada tahun 2014, Jim Gao adalah seorang insinyur di Google. Ia bertanggung jawab untuk memastikan sistem pendingin udara (AC) masif di pusat data (DC) raksasa teknologi tersebut berjalan semulus dan seefisien mungkin. Dengan latar belakang teknik mesin, Gao mengikuti praktik terbaik yang telah mapan di industri energi dan memperoleh hasil yang luar biasa. Setelah langkah-langkah penghematan energi yang paling umum diadopsi diterapkan, kinerja pusat data Google mulai stagnan, mengungkap keterbatasan pendekatan tradisional terhadap penghematan energi. Jelas bahwa pendekatan baru diperlukan.

Gao memutuskan untuk menempuh jalur yang belum pernah ditempuh sebelumnya, memanfaatkan kebijakan 20% Google, sebuah inisiatif yang memungkinkan karyawan untuk menghabiskan 20% waktu mereka mengerjakan apa yang menurut mereka paling menguntungkan Google. Sebagai seorang insinyur pusat data, ia sangat menyadari sensor yang digunakan di pusat data dan sejumlah besar data yang dikumpulkan darinya untuk keperluan operasional. Gao memutuskan untuk mempelajari pembelajaran mesin dan mencoba membangun model untuk memprediksi dan meningkatkan kinerja pusat data.

Masalah Konsumsi Energi Pusat Data

Pusat data adalah bangunan yang menampung server jaringan. Dalam kasus Google, mesin-mesin ini melayani kueri Penelusuran dan Maps, menyimpan foto dan dokumen, serta melakukan semua tugas lain yang dibutuhkan Google untuk menawarkan layanannya kepada pengguna. Konsumsi energi merupakan pendorong utama biaya pusat data (DC) karena banyaknya komputer yang boros daya. Biaya tinggi bukan satu-satunya faktor yang perlu dipertimbangkan ketika mempertimbangkan konsumsi energi DC; dampak lingkungan juga penting. DC saat ini mengonsumsi 2% listrik dunia, angka yang pasti akan meningkat seiring meningkatnya kebutuhan akan layanan jaringan. Jumlah energi yang digunakan untuk menyalakan komputer tidak dapat dengan mudah dioptimalkan oleh tim operasi karena bergantung pada beban kerja komputasi dan efisiensi chip. Karena alasan ini, para insinyur pusat data berupaya mengurangi semua konsumsi tambahan.

Efisiensi pusat data biasanya diukur dengan melacak metrik yang disebut efektivitas penggunaan daya (PUE). Metrik ini mencerminkan berapa banyak energi yang digunakan untuk hal lain selain komputer sebenarnya yang membentuk pusat data:

$$PUE = \frac{\text{Total Facility Energy}}{\text{IT Equipment Energy}}$$

Pusat data yang sempurna memiliki PUE 1: semua energi dihabiskan untuk menyalakan komputer. Semakin tinggi PUE, semakin banyak energi yang dihabiskan untuk sistem lain, di antaranya pendinginan adalah yang paling penting. Misalnya, PUE sebesar 1,5 berarti bahwa untuk setiap kilowatt-jam (kWh) energi yang dikonsumsi untuk menyalakan komputer, dibutuhkan tambahan 0,5 kWh energi untuk pendinginan dan kebutuhan kecil lainnya.

Google selalu menjadi pemimpin dalam efisiensi PUE. Menurut survei tahun 2018 terhadap 900 operator DC oleh Uptime Institute, rata-rata PUE di industri ini adalah 1,58.

Google terus meningkatkan PUE-nya hingga mencapai 1,12 pada tahun 2013. Sayangnya, nilai ini baru membaik pada tahun 2017.

Pendekatan ML untuk Efisiensi Pusat Data

Gao menyadari bahwa salah satu kendala untuk menurunkan PUE lebih lanjut adalah sangat rumitnya memprediksinya dengan tepat dalam berbagai skenario menggunakan pendekatan rekayasa tradisional, karena interaksi yang kompleks antar faktor (misalnya, angin dapat membantu mendinginkan pembangkit dan mengurangi kebutuhan akan pendinginan buatan). Di sisi lain, ia sangat menyadari besarnya kumpulan data yang dikumpulkan oleh timnya sebagai bagian dari operasi sehari-hari, berkat ribuan sensor yang disebarkan di seluruh komponen yang mengumpulkan jutaan titik data.

Gao meramalkan potensi penggunaan data ini untuk melatih model pembelajaran mesin (ML) yang mampu mengatasi keterbatasan termodinamika tradisional. Pendekatan pertamanya adalah membangun jaringan saraf tiruan sederhana (algoritma klasik yang digunakan untuk membangun model pembelajaran terawasi) yang dilatih untuk memprediksi PUE, dengan mempertimbangkan daftar fitur yang memengaruhinya. Gao menggunakan total 19 fitur, termasuk:

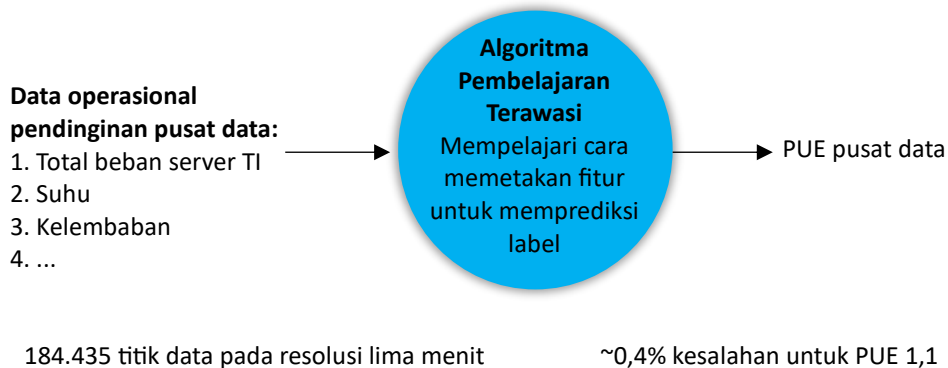
- Total beban TI server
- Total jumlah pompa air proses yang beroperasi
- Rata-rata titik setel suhu air keluar menara pendingin
- Total jumlah pendingin yang beroperasi
- Rata-rata suhu pendekatan penukar panas
- Suhu bola basah dan bola kering udara luar
- Kelembapan relatif udara luar, kecepatan angin, dan arah

Label dari masalah pembelajaran terawasi ini adalah PUE. Seperti yang dapat Anda lihat pada Gambar 2.6, untuk melatih modelnya, Gao menggunakan 184.435 sampel waktu dengan resolusi lima menit (sekitar dua tahun data operasional). Model akhir mampu memprediksi PUE DC dalam kisaran $0,004 \pm 0,0005$, dengan kesalahan sekitar 0,4%, untuk PUE sebesar 1,1. Gao berhasil membangun model bukti konsep (POC) pertama dengan cepat menggunakan kerangka kerja pengkodean sumber terbuka. Model akhir yang dihasilkan telah digunakan untuk tiga aplikasi utama:

- Peringatan dan pemecahan masalah kinerja otomatis, dengan membandingkan kinerja DC aktual versus prediksi untuk serangkaian kondisi tertentu
- Mengevaluasi sensitivitas PUE terhadap parameter operasional
- Menjalankan simulasi digital dengan konfigurasi berbeda tanpa melakukan perubahan fisik

Hasil kerja Gao berdampak besar pada perusahaan dan diakui publik oleh VP pusat data, Joe Kava, yang menjelaskan pekerjaan dan hasilnya dalam postingan blog resmi Google. "Pusat Data yang Lebih Baik Melalui Pembelajaran Mesin" menyoroti bagaimana model Gao mampu mengidentifikasi pola dalam data yang mustahil dikenali oleh manusia, menghasilkan model yang mampu memprediksi PUE dengan akurasi 99,6%.

Studi kasus



Gambar 2.6 Model pembelajaran mesin Gao

Gao dan timnya mulai menggunakan model tersebut untuk menemukan cara-cara baru dalam meningkatkan efisiensi. Ketika server dimatikan, hal itu diketahui menyebabkan penurunan kinerja di pusat data, misalnya. Berkat model Gao, tim Google DC mampu menjalankan simulasi perilaku pusat data dan menemukan cara baru untuk mengatasi penurunan kinerja, sehingga menghemat energi dan biaya.

Penerapan ML untuk Optimasi Pusat Data Mendapatkan Momentum

Kerja keras Gao menarik perhatian di jajaran atas perusahaan. Ia dipromosikan untuk memimpin tim lintas fungsi yang terdiri dari insinyur pembelajaran mesin, perangkat lunak, kelistrikan, mekanik, kontrol, dan operasi untuk mengembangkan solusi intelijen pusat data yang menyeluruh.

Keahlian ML dalam tim baru ini didatangkan oleh DeepMind, sebuah perusahaan Inggris yang berspesialisasi dalam algoritma AI mutakhir yang diakuisisi oleh Google pada tahun 2014 senilai Rp. 500 Triliun. Keahlian ML tambahan tersebut ternyata menjadi aset kunci dalam mencapai kinerja yang belum pernah terlihat sebelumnya dalam sejarah perusahaan, mengurangi biaya overhead PUE sebesar 15% dan menghasilkan pengurangan tagihan energi secara keseluruhan sebesar 40%.

Meskipun model DeepMind lebih kompleks dan akurat daripada pendekatan Gao dan memungkinkan Google memangkas tagihan energinya hingga 40%, model tersebut tetap menghasilkan rekomendasi yang perlu diverifikasi dan diimplementasikan oleh manusia. Langkah logis selanjutnya dari proyek ini adalah sepenuhnya mengalihkan pengelolaan daya DC ke solusi AI, dengan tetap menempatkan para ahli manusia sebagai pengawas.

Model DeepMind diterapkan pada Agustus 2017, dan telah menghasilkan penghematan energi sebesar 10%. Selama model baru berjalan, model tersebut menghasilkan data tambahan yang digunakan untuk melatih ulang model dan meningkatkan kinerjanya lebih lanjut. Dalam setahun, penghematan energi sebesar 10% meningkat menjadi 30%.

Sistem kontrol AI menggunakan cara-cara baru untuk mengelola pendinginan yang menghasilkan kinerja terbaik dengan cara yang tak terduga dan belum pernah dieksplorasi oleh operator DC ahli. Misalnya, Dan Fuenffinger, salah satu operator pusat data Google, berkomentar tentang keterkejutannya ketika sistem AI mampu secara otomatis memanfaatkan

kondisi musim dingin dan menghasilkan air pada suhu yang lebih rendah, sehingga mengurangi penggunaan energi.

PERTANYAAN KASUS

1. Meningkatkan PUE pusat data merupakan masalah yang menantang, dan Gao merasa bahwa ML dapat menjadi solusi untuk mengatasinya. Apa saja karakteristik tantangan ini yang membuatnya cocok untuk ML?
2. Apa saja elemen kunci yang memungkinkan Gao membangun POC pertama sendirian?
3. Bagaimana Google mengubah inisiatif seorang insinyur menjadi kesuksesan di seluruh perusahaan?

DISKUSI KASUS

Proyek pusat data Google adalah kasus optimasi klasik yang sangat cocok dengan ML. Proyek ini dimulai dengan indikator kinerja utama (KPI) yang ingin kami prediksi (PUE) oleh bisnis, dan serangkaian besar parameter yang memengaruhinya (suhu, tekanan, beban, dan sebagainya). Para insinyur telah mencoba membangun model yang dapat memetakan parameter-parameter ini ke PUE, tetapi metode tradisional yang mengandalkan termodinamika memiliki kekurangan yang signifikan sehingga tidak dapat memberikan hasil di atas ambang batas tertentu.

Seperti inilah seharusnya permasalahan bisnis pembelajaran mesin: hubungan yang kompleks dan tidak diketahui antara serangkaian variabel yang memengaruhi metrik yang berdampak langsung pada bisnis. Gao melakukan pekerjaan yang hebat dalam mengidentifikasi karakteristik ini dan menyadari bahwa pembelajaran mesin bisa menjadi palu yang tepat untuk mencapai tujuan tersebut.

Beberapa elemen memungkinkan Gao untuk berhasil. Yang pertama adalah ketersediaan data, yang tanpanya mustahil untuk mengejar solusi pembelajaran mesin. Gao mengetahui dari pengalaman kerjanya tentang keberadaan kumpulan data ini, dan berkat tata kelola data yang gesit di perusahaannya, ia dapat dengan mudah mengambil data yang dibutuhkannya untuk memulai eksperimen. Kita akan membahas tata kelola data nanti secara lebih rinci, tetapi jelas bahwa ketersediaan dan akses ke data sangat penting untuk memulai proyek pembelajaran mesin seperti ini. Perhatikan bahwa data ini telah tersedia selama bertahun-tahun, tetapi tidak ada yang menyadari lebih dari sekadar tujuan operasionalnya.

Faktor lain yang membantu Gao memulai adalah ketersediaan perangkat sumber terbuka yang memungkinkannya membangun POC dengan cepat. Hal ini memungkinkannya mendapatkan perhatian dan dukungan perusahaan yang dibutuhkannya untuk membawa pendekatannya ke tingkat berikutnya, dengan melibatkan para insinyur ML terampil dari DeepMind dan membangun berdasarkan intuisi pertamanya.

Terakhir, cara proyek ini dijalankan merupakan contoh sempurna dari strategi ML yang hebat. Kami akan membahas topik ini lebih detail di bagian kedua buku ini, tetapi perlu

diperhatikan bahwa proyek ini tidak dimulai dengan misi yang muluk-muluk dan solusi yang rumit. Langkah pertama cukup sederhana untuk dibangun oleh seorang insinyur dalam 20% waktunya, tetapi juga cukup kuat untuk membuktikan potensi pendekatan tersebut. Langkah kedua lebih kompleks, dan untuk mencapai kinerja yang lebih tinggi, diperlukan pengetahuan baru, pengujian, peningkatan momentum, dan transisi ke langkah berikutnya, yaitu kendali AI yang sepenuhnya otomatis. Google merencanakan proyek ini dibangun secara bertahap, masing-masing dengan pengembalian investasi (ROI) yang nyata dan membangun berdasarkan langkah sebelumnya, hingga mencapai kinerja mutakhir dan berdampak besar pada laba bersih.

Singkatnya, studi kasus ini memiliki poin-poin utama berikut:

- Masalah yang berpotensi cocok untuk ML harus memiliki karakteristik tertentu:
 - Terdapat KPI yang penting.
 - KPI dipengaruhi oleh serangkaian parameter yang terukur.
 - Parameter tersebut memengaruhi KPI melalui hubungan yang kompleks dan tidak diketahui.
- Kumpulan data operasional dapat memiliki nilai tersembunyi yang luar biasa.
- Teknologi sumber terbuka memungkinkan tim insinyur kecil untuk membangun POC yang memberikan hasil nyata bagi bisnis, dan membuktikan bahwa konsep tersebut dapat ditingkatkan dan ditingkatkan lebih lanjut.
- Saat membangun proyek ML, selalu merupakan ide yang baik untuk memulai dari yang kecil, mengidentifikasi tugas terkecil yang dapat memberikan dampak bisnis langsung, untuk membangun fondasi bagi proyek yang lebih kompleks.

Bagaimana Square Menggunakan AI untuk Meminjamkan Miliaran Rupiah Kepada Usaha Kecil

Studi kasus ini membahas bagaimana sebuah perusahaan rintisan fintech menantang bank-bank besar dengan meminjamkan modal kepada usaha kecil yang secara tradisional kurang terlayani. Dengan menggunakan pembelajaran mesin pada dataset transaksi ritelnya, Square Capital dapat secara proaktif menawarkan pinjaman kecil kepada bisnis bahkan sebelum mereka memintanya. Tingkat tunggakan Square adalah setengah dari rata-rata industri.

Square dan Square Capital

Square adalah perusahaan jasa keuangan yang berfokus pada kebutuhan usaha kecil. Perusahaan ini didirikan oleh pendiri Twitter, Jack Dorsey, pada tahun 2009, dan go public pada tahun 2015. Berkantor pusat di San Francisco, perusahaan ini memiliki kapitalisasi pasar sebesar Rp. 260 Triliun pada tahun 2019 dan mempekerjakan 2.300 orang. Produk pertama Square adalah pembaca kartu kredit yang terhubung langsung ke ponsel pintar untuk memungkinkan peritel kecil dan profesional menerima pembayaran kartu kredit. Perangkat keras pembaca tersebut didistribusikan secara gratis tanpa perlu berlangganan bulanan, yang melejitkan reputasi Square di kalangan peritel kecil. Desain perangkat keras yang ikonik dan aliran referensi yang stabil juga berkontribusi memperkuat reputasi Square (ekspansi internasional sejauh ini terbatas).

Pada tahun-tahun pertama operasinya, Square berfokus pada ceruk bisnis kecil dengan omzet kurang dari Rp. 1.250.000.000; bisnis-bisnis ini kurang terlayani oleh bank dan pemroses pembayaran konvensional. Biaya rendah Square dan situs web yang mudah digunakan membantu menarik para wirausahawan ke dalam ekosistem keuangan yang sedang berkembang pesat. Pada tahun-tahun menjelang pencatatan saham perdananya, Square melengkapi layanannya dengan meluncurkan aplikasi pembayaran peer-to-peer, platform manajemen hubungan pelanggan (CRM) dan pemasaran, serta perangkat keras point-of-sale untuk pedagang besar.

Pada tahun 2014, Square meluncurkan Square Capital, komponen lain dari ekosistem keuangannya. Seiring berkembangnya bisnis kecil, mereka sering mengalami masalah arus kas yang dapat diatasi dengan modal kerja tambahan. Namun, bank konvensional seringkali tidak peka terhadap kebutuhan nasabah ini, karena jumlah yang kecil tidak sebanding dengan biaya overhead yang dikeluarkan untuk pemasaran dan pemrosesan pinjaman. Square Capital menggunakan data transaksi titik penjualannya untuk secara proaktif menawarkan pinjaman kecil kepada pelanggannya.

Bagi pelanggan, pengalamannya lancar, karena mereka dapat menerima dan mengelola pinjaman langsung dari situs web manajemen yang sama dengan yang mereka gunakan untuk layanan lain dalam ekosistem. Pinjaman tidak memiliki jadwal pembayaran yang spesifik: dana secara otomatis dipotong dari transaksi kartu kredit yang diproses oleh Square. Kombinasi pengalaman yang lancar dan basis pelanggan yang ada telah memungkinkan Square Capital untuk berkembang pesat: 200.000 pedagang telah meminjam lebih dari Rp. 31 Triliun melalui platform ini hanya dalam kurun waktu 2014–2018.

Meskipun model risiko kredit telah menjadi bagian penting dalam industri keuangan selama beberapa dekade, Square memiliki akses ke kumpulan data yang jauh lebih luas, termasuk musim dan waktu setiap pembelian individu. Hal ini memungkinkan Square untuk membangun model dengan visibilitas yang baik terhadap posisi arus kas setiap bisnis dan menentukan besaran pinjaman yang ditawarkan dengan tepat. Bersama dengan opsi pembayaran yang mudah, hal ini menurunkan tingkat tunggakan pinjaman menjadi 4%, setengah dari rata-rata industri.

Kritik dan Persaingan

Dari sudut pandang keuangan, pinjaman yang ditawarkan oleh Square Capital bukanlah hal baru. Para pemberi pinjaman telah menawarkan uang muka tunai pedagang (MCA) untuk memenuhi kebutuhan kas jangka pendek bisnis sejak dahulu kala. Dibandingkan dengan MCA, yang tidak memiliki batas waktu yang ditentukan, pinjaman Square Capital harus dilunasi dalam waktu 18 bulan. Hal ini memungkinkan Square untuk mengalihkan pinjaman ke lebih banyak lembaga keuangan, yang lebih sulit menangani MCA tanpa tanggal pembayaran yang diketahui. Square beroperasi hanya sebagai pemrakarsa pinjaman, alih-alih pemberi pinjaman, dengan mempertahankan neraca yang menarik dan sebanding dengan perusahaan teknologi lainnya.

Meskipun sebagian besar lembaga keuangan menggunakan pembelajaran mesin untuk mengevaluasi risiko kredit, perusahaan fintech seperti Square saat ini juga dikritik karena

terlalu banyak menghilangkan pengawasan manusia dalam proses aplikasi kredit mereka. Misalnya, startup pinjaman peer-to-peer Prosper menjadi sorotan pada tahun 2015 setelah meminjamkan Rp. 250.000.000 kepada pasangan teroris di balik penembakan San Bernardino (California), yang menewaskan 24 orang.

Meskipun komunitas teknologi antusias dengan kemampuan fintech untuk mendisrupsi dan meningkatkan skala perbankan tradisional melalui penggunaan pembelajaran mesin, beberapa komunitas tidak seoptimis itu. Misalnya, usaha kecil yang sama yang menikmati waktu penyelesaian cepat untuk pinjaman Square Capital juga vokal tentang ketidakadilan yang dirasakan dan kurangnya transparansi ketika aplikasi pinjaman mereka ditolak. Karena Square tidak mempublikasikan detail apa pun tentang model evaluasi risikonya, beberapa pedagang terpaksa merevisi ulang keputusan algoritmik. Dari sudut pandang teknis, hal ini terkait dengan masalah akurasi (dinyatakan dalam bentuk positif benar/salah) dan bias model.

Berkat integrasinya dengan ekosistem Square, Square Capital dapat menawarkan fitur-fitur menarik dibandingkan bank dan MCA tradisional, yang memiliki akses lebih sedikit secara langsung ke alur transaksi pedagang. Namun, layanan lain seperti PayPal Working Capital juga memiliki akses ke sumber data ini, sehingga menyulitkan Square Capital untuk menonjol. Survei yang dilakukan terhadap pemilik usaha kecil menunjukkan bahwa mereka biasanya memilih pemberi pinjaman berdasarkan peluang persetujuan yang dirasakan, alih-alih jumlah atau jangka waktu pinjaman. Ini berarti Square Capital mungkin harus berfokus pada pengalaman pengguna dan integrasi dengan ekosistem lainnya, alih-alih mengoptimalkan model evaluasi risiko.

PERTANYAAN KASUS

1. Risiko kredit merupakan aplikasi pembelajaran mesin yang telah dieksplorasi dengan baik. Menurut Anda, apa yang berhasil dilakukan Square dibandingkan dengan sejarah masa lalu dan pesaingnya saat ini?
2. Pelajaran apa yang dapat Anda petik dari interaksi antara pembelajaran mesin dan organisasi regulator (perbankan, dalam hal ini)?
3. Dapatkah Anda mengembangkan strategi untuk mengumpulkan data pelatihan untuk aplikasi ini? Bagaimana Anda akan mengukur kinerja model yang dihasilkan?

DISKUSI KASUS

Kami menyajikan studi kasus tentang Square Capital ini karena merupakan kisah sukses tentang pengintegrasian model bisnis berbasis AI ke dalam ekosistem layanan yang sudah ada. Lebih lanjut, studi kasus ini juga mengeksplorasi isu-isu seputar pengenalan fitur berbasis ML ke dalam industri yang sangat teregulasi seperti perbankan.

Produk inti Square secara historis adalah sistem point-of-sale untuk usaha kecil. Keunggulannya dianggap lebih pada desain, pengalaman pengguna, dan layanan pelanggan, alih-alih teknologi mutakhir. Namun, data secara efektif merupakan inti dari bisnis Square, karena deteksi dan pencegahan penipuan merupakan aspek mendasar dari setiap perusahaan jasa keuangan. Hal ini menunjukkan bahwa setelah suatu organisasi membangun budaya yang kuat seputar data (termasuk pengumpulan, penyimpanan, dan penyediaannya bagi karyawan), wajar saja jika mereka memperkenalkan produk baru berdasarkan data tersebut. Ketika terdapat infrastruktur data yang dirancang dengan baik, model risiko kredit dapat diintegrasikan ke dalam platform yang sama yang sebelumnya digunakan untuk memindai umpan transaksi real-time guna mendeteksi penipuan, sehingga memberikan nilai tambah bagi investasi tersebut dalam teknologi inti.

Dari sudut pandang pengalaman pengguna, Square Capital adalah contoh yang baik dalam menghadirkan fitur-fitur berbasis AI secara mulus ke dalam platform yang sudah ada dengan cara yang sepenuhnya transparan dan alami bagi pengguna akhir. Berdasarkan kepercayaan yang telah diberikan para pedagang kepada Square, penawaran Square Capital bersifat sederhana dan sesuai untuk audiensnya. Tidak ada lagi bualan yang tidak perlu tentang "evaluasi risiko berbasis AI" atau pemasaran yang mengutamakan teknologi. Yang tersisa hanyalah fokus pada pengalaman pelanggan.

Dalam hal keuangan perusahaan, menarik untuk melihat bagaimana fleksibilitas yang diperoleh Square dengan menggunakan ML untuk evaluasi risiko kredit memungkinkannya bereksperimen dengan model bisnis baru. Menawarkan pinjaman berjangka tetap, alih-alih MCA tradisional, memungkinkan Square untuk fokus pada misi intinya daripada mengambil pinjaman itu sendiri. Hal ini memungkinkan Square untuk menciptakan pertumbuhan yang eksplosif (sekitar 50% per tahun) yang memanfaatkan keunggulannya tanpa menambah risiko pada neraca keuangannya.

Elemen menarik dari kasus ini adalah kelemahan ML dalam kasus-kasus tertentu. Kami telah mempertimbangkan pembelajaran otonom algoritma ML sebagai kekuatan, tetapi terkadang kami kesulitan untuk memahami logika yang diikuti oleh algoritma tersebut saat mengambil keputusan. Jika pengajuan pinjaman usaha kecil ditolak oleh algoritma ML, menjelaskan alasannya bisa jadi sulit atau bahkan mustahil. Kami akan membahas aspek ML ini lebih detail di bab terakhir.

Square berawal dari posisi yang kuat karena model bisnisnya memungkinkan perusahaan mengumpulkan data bisnis inti tentang semua pelanggannya: riwayat transaksi mereka. Aset data inilah yang menjadi pilar dasar Square Capital. Perhatikan bahwa data ini dikumpulkan secara alami untuk menawarkan layanan POS standar oleh Square, dan kemudian menjadi aset yang luar biasa ketika Square Capital diperkenalkan.

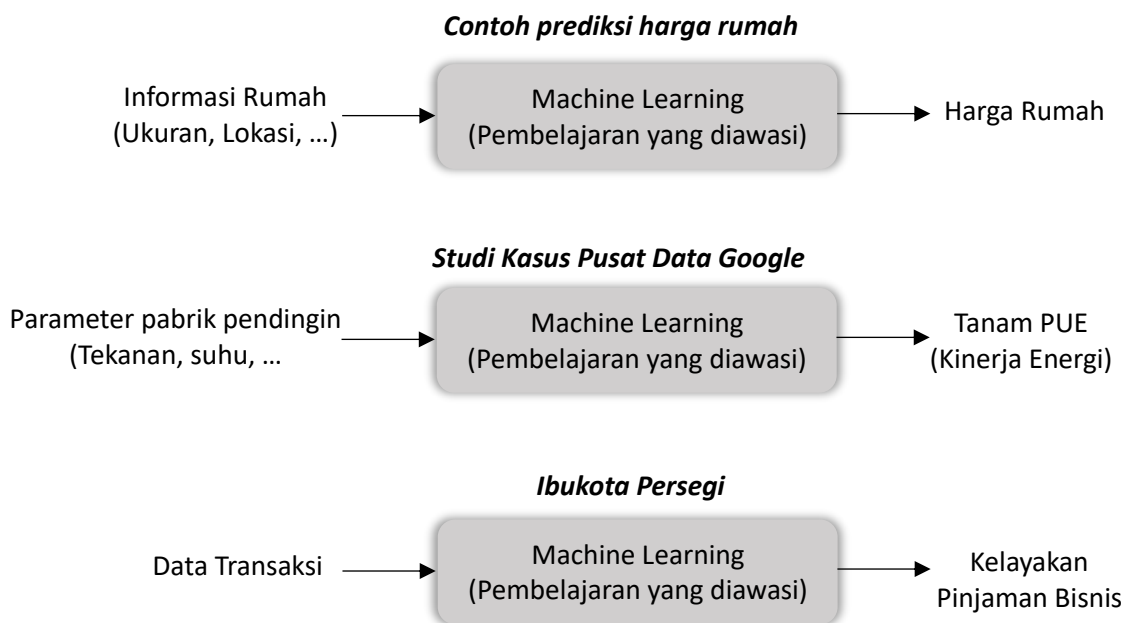
Pelajaran dari Studi Kasus

Baik Google maupun Square menciptakan aplikasi AI yang memberi mereka keunggulan besar dibandingkan pesaing, sekaligus memiliki perbedaan pada intinya. Google menggunakan AI untuk mengurangi biaya dan meningkatkan operasional, sementara Square menggunakannya untuk meluncurkan layanan yang benar-benar baru. Prinsip dasarnya tetap

sama: baik Square maupun Google menggunakan AI pada data bisnis inti, menciptakan nilai yang sangat besar dari apa yang dianggap sebagian besar perusahaan lain sebagai biaya hangus.

Perhatikan bahwa baik Google maupun Square menggunakan data yang awalnya tidak dikumpulkan dengan mempertimbangkan AI. Dalam kasus Google, praktik umum adalah menyimpan data operasional dari instalasi pendingin udara di pusat data, terutama untuk pelaporan dan pemantauan bahwa sistem beroperasi sesuai harapan. Gagasan Jim Gao untuk menggunakan ML guna meningkatkan kinerja sistem merupakan hal baru dan menunjukkan intuisi yang hebat dari seorang insinyur Google.

Hal yang sama berlaku untuk Square: mereka mengumpulkan data transaksi pelanggan untuk menawarkan layanan seperti analitik penjualan, tetapi menawarkan pinjaman bukanlah alasan awal pengumpulan data ini. Square bahkan tidak menawarkan layanan pinjaman bisnis sejak awal: layanan ini sepenuhnya baru yang dimungkinkan oleh pembelajaran mesin (ML) dan oleh kumpulan data kaya yang dimilikinya. Hal ini memungkinkan Square memasuki pasar baru yang sama sekali kurang terlayani, karena bank tradisional tidak memiliki sumber daya untuk mengevaluasi kelayakan pinjaman usaha kecil.



Gambar 2.7 Bagaimana contoh prediksi rumah dan studi kasus Square dan Google menggunakan pembelajaran terawasi

Aspek lain yang perlu Anda ingat dari kedua studi kasus ini adalah keduanya menggunakan metode pembelajaran terawasi. Google belajar memprediksi PUE instalasi pendingin berdasarkan parameter operasionalnya, sementara Square menentukan kelayakan usaha kecil berdasarkan riwayat transaksinya. Kesamaan dalam hal penetapan masalah ditunjukkan pada Gambar 2.7. Dalam kasus Google, pembelajaran terawasi sangat berguna dalam mengatasi keterbatasan solusi tradisional untuk masalah efisiensi energi. Aplikasi

Square digunakan untuk mengotomatiskan proses yang akan terlalu memakan waktu jika dilakukan oleh manusia.

2.4 MENGEVALUASI KINERJA DAN RISIKO

Sekarang saatnya untuk bertemu dengan pimpinan FutureHouse, untuk duduk dan membahas risiko dan kekhawatiran dari proyek yang telah kami jalankan. Kami telah mengidentifikasi kasus bisnis yang hebat untuk algoritma prediksi nilai properti, mengumpulkan data yang tepat, dan membangunnya. Sekarang, Anda hanya perlu menekan tombol merah yang akan menampilkan fitur tersebut di situs web bagi pengguna. Anda akan segera melanjutkan; semua orang di kantor memperhatikan Anda. Tangan Anda berkeringat dan gemetar. Anda mengambil risiko, menekan tombol, menutup mata, dan berharap algoritmanya akan berfungsi. Untungnya, hal ini tidak terjadi di dunia nyata. Menerapkan model tanpa mengukur kinerjanya secara menyeluruh adalah tindakan yang tidak bertanggung jawab.

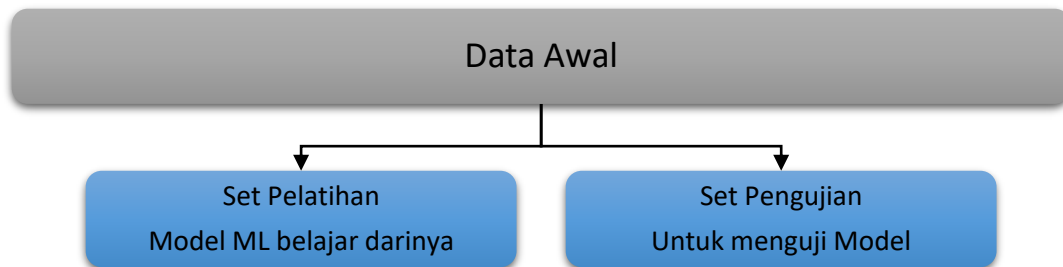
Hal pertama yang terpenting: pembelajaran mesin didasarkan pada data dan statistik, sehingga prediksinya kemungkinan besar tidak akan pernah 100% akurat. Namun, kami ingin prediksi menjadi berharga bagi bisnis, dengan tetap menghormati semua standar keselamatan dan etika. Sistem keseluruhan yang kami bangun harus toleran terhadap margin kesalahan ini karena alasan yang sama dengan alasan kami menggunakan pembelajaran mesin sejak awal: beberapa masalah ini begitu kompleks sehingga tidak mungkin data sebanyak itu dapat mencakup semuanya. Perusahaan yang menggunakan ML dalam situasi hidup-mati (seperti mobil tanpa pengemudi) harus jauh lebih berhati-hati dan konservatif daripada pemasar yang menggunakannya untuk mengoptimalkan pesan mereka kepada pelanggan. Untungnya, semua aplikasi yang kami bahas dalam buku ini termasuk dalam kategori yang terakhir.

Mengevaluasi kinerja sebuah model memiliki dua aspek utama. Cara yang baik untuk memahaminya mungkin sudah tidak asing lagi bagi Anda dari literatur manajemen: hal yang tidak diketahui yang diketahui dan hal yang tidak diketahui yang tidak diketahui.

Hal yang tidak diketahui yang utama adalah akurasi: sebuah pengukuran seberapa baik model dalam mencocokkan fitur dengan label. Dalam masalah prediksi nilai rumah, hal ini diterjemahkan menjadi seberapa dekat harga yang diprediksi dengan harga aktual setelah rumah terjual. Untuk masalah pembelajaran terawasi, ada cara cerdas untuk mengukur akurasi model. Insinyur biasanya secara acak membagi data pelatihan menjadi dua bagian (misalnya, 70% dan 30% dari data), seperti yang dapat Anda lihat pada gambar 2.8. Mereka melatih model hanya pada bagian pertama, dan menggunakan bagian kedua sebagai set pengujian untuk mensimulasikan skenario dunia nyata dari model yang berjalan pada data yang tidak diketahui. Karena set pengujian juga merupakan bagian dari set data awal, kami memiliki label untuknya.

Oleh karena itu, kami dapat membandingkan label yang diprediksi dengan yang sebenarnya dan mengukur akurasi model dalam situasi yang cukup cocok dengan dunia nyata. Menyisihkan set pengujian itu penting, karena pengujian pada data yang digunakan untuk pelatihan agak seperti memberi siswa ujian dengan pertanyaan yang telah mereka lihat solusinya. Dalam kasus prediksi harga rumah, Anda dapat membayangkan memilih satu atau

dua rumah di setiap blok, menghapusnya dari set data yang digunakan untuk pelatihan, dan menambahkannya ke set pengujian. Ini akan memberi kita gambaran yang baik tentang seberapa baik kinerja model.



Gambar 2.8 Data yang tersedia dibagi menjadi set pelatihan dan set pengujian. Set pertama digunakan untuk membuat model belajar, dan set kedua digunakan untuk menguji model pada data yang tidak diketahui, mensimulasikan aplikasi di dunia nyata.

Dalam hal penggunaan AI dalam organisasi, akurasi bukanlah tolok ukur keberhasilan. Model yang tidak akurat yang diterapkan pada masalah yang tepat bisa jauh lebih berharga bagi bisnis daripada model yang sangat akurat yang memecahkan masalah yang tidak signifikan. Aturan emas yang perlu diingat adalah dari ahli statistik George Box:

Semua model salah, tetapi beberapa berguna.

Jika akurasi adalah metrik yang baik untuk mengukur hal-hal yang tidak diketahui yang diketahui, bagaimana dengan hal-hal yang tidak diketahui (yang lebih menakutkan)? Dalam situasi ini, model memberikan jawaban yang salah atau tidak masuk akal. Sering kali, ini karena data pelatihan tidak lengkap, dan kita mencoba memprediksi target dalam situasi (kombinasi fitur) yang belum pernah terlihat sebelumnya dalam pelatihan. Meskipun inferensi dalam situasi yang tak terlihat adalah inti dari penggunaan pembelajaran mesin, kueri yang sepenuhnya tak terduga tetap dapat mengacaukan model. Misalnya, entri data yang salah dapat menyebabkan model mencoba menebak harga rumah dengan 50 kamar mandi padahal tidak ada rumah besar seperti itu dalam dataset pelatihan. Sebagian besar keluarga model tidak memunculkan tanda bahaya dalam situasi seperti ini, sehingga para insinyur harus menerapkan pemeriksaan kewarasan pada input dan output untuk memastikan keamanan.

Faktor risiko penting lainnya berputar di sekitar inti dari apa yang membuat pembelajaran mesin begitu hebat: ekstraksi otomatis pemetaan antara target dan fitur. Ini adalah kemampuan yang hebat, tetapi juga berarti bahwa para insinyur tidak dapat dengan mudah memahami mengapa model membuat prediksi tertentu, mengingat serangkaian fitur. Inilah mengapa pembelajaran mesin sering disebut kotak hitam: Anda tahu apa yang masuk (fitur) dan apa yang keluar (target), tetapi sulit untuk memahami mengapa model memilih output tersebut. Meskipun beberapa keluarga model lebih baik daripada yang lain, kenyataannya ini adalah salah satu aspek paling mengkhawatirkan yang terkait dengan

penggunaan AI skala besar. Jangan khawatir, karena kami telah membahas banyak materi tentang masalah ini di bagian kedua buku ini.

Kami telah membahas banyak hal dalam bab ini. Kami memperkenalkan Anda pada konsep-konsep utama pembelajaran terawasi dengan merancang fitur AI yang berhadapan langsung dengan pelanggan: prediksi nilai properti untuk perusahaan pialang real estat fiktif. Anda mempelajari nilai data dan bagaimana data tersebut digunakan untuk membangun model pembelajaran mesin. Studi kasus telah membantu Anda menerapkan pengetahuan ini ke dalam situasi industri, sehingga Anda akan mulai mampu mengenali peluang serupa di organisasi Anda sendiri.

Sebagai penutup, sebuah fakta menarik. Zillow adalah pasar real estat daring terkemuka di Amerika Serikat. Memprediksi nilai properti sangat penting bagi Zillow, dan pembelajaran mesin merupakan solusi yang begitu jelas, sehingga Zillow menyelenggarakan kompetisi untuk algoritma terbaik.

- Data bisnis inti adalah jenis data yang paling erat kaitannya dengan mesin penghasil nilai suatu organisasi. Oleh karena itu, data ini juga memiliki potensi tertinggi untuk dipengaruhi oleh AI.
- Pembelajaran terawasi adalah keluarga algoritma pembelajaran mesin yang memungkinkan komputer mempelajari cara memetakan input (fitur) dan output (label), dengan contoh yang cukup.
- Pembelajaran mesin dapat mengatasi banyak masalah sulit yang tidak dapat diatasi oleh rekayasa perangkat lunak konvensional, karena pembelajaran mesin didasarkan pada data historis, alih-alih pemahaman matematis.
- Meskipun model berbasis pembelajaran mesin tidak akan pernah 100% akurat, metrik numerik (seperti akurasi) dapat membantu melacak kinerjanya.

BAB 3

AI UNTUK PENJUALAN DAN PEMASARAN

Bab ini membahas

- Mengidentifikasi pelanggan mana yang paling mungkin meninggalkan layanan (churn)
- Menargetkan pelanggan yang paling tertarik untuk membeli lebih banyak (upselling)
- Menggunakan pembelajaran tanpa pengawasan untuk segmentasi pelanggan berbasis data
- Studi kasus: penggunaan AI pada data jaringan listrik dan penggalian analitik ritel

Pada bab sebelumnya, kita telah membahas peran data terstruktur dalam berbagai aplikasi bisnis. Meskipun penjualan dan pemasaran terkadang dapat masuk ke dalam kategori data bisnis inti, keduanya begitu penting dan unik sehingga layak mendapatkan bab tersendiri. Kita akan membahas berbagai masalah pemasaran dan mengeksplorasi bagaimana Anda dapat menggunakan kecerdasan buatan dan ilmu data untuk memperkuat dan meningkatkan hubungan antara organisasi Anda dan pelanggannya.

3.1 MENGAPA AI UNTUK PENJUALAN DAN PEMASARAN

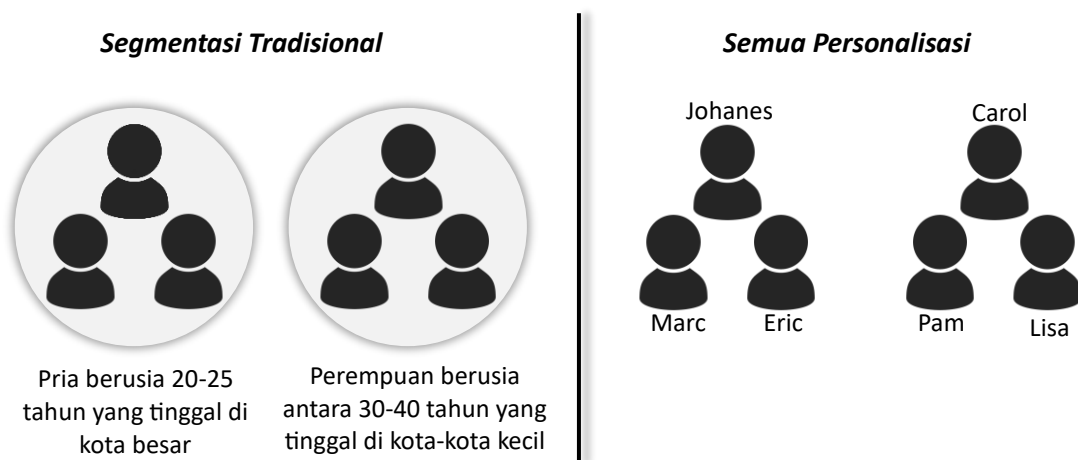
Salah satu tujuan utama pemasar adalah menemukan cara terbaik untuk menawarkan produk yang tepat kepada pelanggan yang tepat pada waktu yang tepat. Namun, meskipun miliaran dolar dipertaruhkan, pemasar telah menghadapi berbagai keterbatasan. Yang pertama adalah kurangnya data. Ketika dunia belum terhubung seperti sekarang, satu-satunya cara untuk mendapatkan jawaban atas pertanyaan seseorang adalah dengan berbicara langsung. Internet sebagian besar memecahkan masalah ini: kini lebih mudah dari sebelumnya untuk menjangkau khalayak luas, memaparkan mereka pada suatu pesan, dan mengukur reaksi mereka. Sisi lain dari hal ini adalah mudahnya menghasilkan data yang begitu besar dan terperinci sehingga terlalu banyak bagi manusia untuk memahami dan mengekstrak wawasan.

Kami ingin memulai bab ini dengan memberi Anda sedikit wawasan tentang mengapa AI mengubah segalanya. Setiap pemasar tahu bahwa tidak semua pelanggan sama, dan bahwa mereka merespons paling baik ketika mereka terlibat dengan pesan yang dipersonalisasi. Strategi pemasaran yang umum adalah membagi pelanggan ke dalam segmen-segmen berdasarkan demografi atau aspek serupa. Segmen sederhana bisa berupa "wanita kaya berusia antara 25 dan 30 tahun yang menghabiskan lebih dari Rp. 10.000.000 per tahun untuk hiburan." Seorang pemasar dapat membuat pesan khusus untuk menjangkau kategori orang ini, yang berbeda dari apa yang akan dilakukan untuk segmen lain. Meskipun teknik ini sudah ada, ini benar-benar yang terbaik yang bisa kita lakukan sebelum AI muncul.

Masalah dengan pendekatan ini adalah, betapa pun spesifiknya Anda dalam segmentasi Anda (pemasar berbicara tentang mikrosegmentasi), Anda akan selalu berakhir dalam situasi di mana dua pelanggan diperlakukan persis sama meskipun mereka pada dasarnya berbeda, hanya karena mereka termasuk dalam kategori yang sama. Ada batasan

jumlah kategori yang dapat dikelola oleh otak manusia. Bayangkan saja berapa banyak teman Anda yang memiliki karakteristik serupa dengan Anda di atas kertas (usia yang sama, lingkungan, pendidikan yang sama) tetapi memiliki selera yang sangat berbeda.

Seperti yang Anda lihat pada gambar 3.1, pendekatan segmentasi pemasaran tradisional tidak dapat menargetkan Marc hanya karena memang Marc. Targetnya akan selalu "pria berusia antara 25 dan 30 tahun yang tinggal di kota besar." AI mengubah aturan main karena dapat memproses lebih banyak informasi. Dengan AI, Anda dapat mencapai personalisasi dalam skala besar, mempelajari orang-orang melalui tindakan dan karakteristik spesifik mereka, serta menargetkan mereka berdasarkan jati diri mereka, bukan berdasarkan kategori yang mereka anggap remeh.



Gambar 3.1 Personalisasi AI versus segmentasi pemasaran tradisional

Apa arti kemampuan personalisasi yang begitu detail bagi bisnis? Perusahaan yang berspesialisasi dalam AI untuk pemasaran dapat menunjukkan beberapa metrik yang mencengangkan dan menjadi impian setiap pemasar. Contohnya adalah Amplero, perusahaan AS yang berspesialisasi dalam pemasaran berbasis AI. Berikut beberapa hasil yang dilaporkan dalam materi pemasarannya:

- Amplero membantu perusahaan telekomunikasi besar Eropa meningkatkan pendapatan rata-rata per pengguna dalam 30 hari pertama dari 0,32% menjadi 2,8%, peningkatan hampir 800%.
- Amplero mengurangi biaya akuisisi pelanggan (CAC) dari salah satu dari lima operator seluler teratas di Amerika Utara lebih dari 97%: dari Rp. 400.000 per pelanggan menjadi hanya Rp. 10.000.
- Amplero berhasil menargetkan ulang pelanggan yang paling tidak puas dari sebuah operator seluler besar Eropa tiga minggu sebelum mereka membatalkan paket, menciptakan pengalaman pelanggan yang lebih bermakna untuk melibatkan mereka kembali, dan meningkatkan tingkat retensi dari 2% menjadi 10%.

Angka-angka ini tidak dimaksudkan untuk menyombongkan hasil dari perusahaan pemasaran tertentu. Anda akan menemukan banyak perusahaan rintisan dan organisasi besar yang dapat

mencapai hasil serupa. Jika gagasan mencapai kinerja seperti ini di organisasi Anda membuat Anda merinding, Anda tidak sendirian. Mari kita lihat bagaimana hal ini dapat diwujudkan.

Pemasaran adalah fungsi yang kompleks, jadi alih-alih mencantumkan semua kemungkinan penerapannya, kami akan berfokus pada tiga masalah umum yang berlaku untuk sebagian besar bisnis:

- Mengidentifikasi pelanggan mana yang kemungkinan besar akan meninggalkan layanan Anda (churn)
- Mengidentifikasi pelanggan mana yang kemungkinan besar akan membeli layanan baru (upselling)
- Mengidentifikasi kelompok pelanggan yang serupa (segmentasi pelanggan)

3.2 MEMPREDIKSI PELANGGAN YANG CHURN

Salah satu metrik pemasaran terpenting adalah churn pelanggan (juga dikenal sebagai atrisi atau pergantian pelanggan). Churn didefinisikan sebagai persentase pelanggan yang meninggalkan bisnis selama periode waktu tertentu. Bukankah luar biasa jika mengetahui sebelumnya pelanggan mana yang paling tidak puas dan kemungkinan besar akan meninggalkan produk atau layanan dalam waktu dekat? Inilah tepatnya bagaimana AI dapat membantu Anda memecahkan masalah churn pelanggan: menggunakan pembelajaran mesin dan aset data organisasi, kami dapat menemukan pelanggan yang paling mungkin meninggalkan layanan Anda dan menghubungi mereka dengan pesan yang dipersonalisasi untuk meningkatkan kembali keterlibatan mereka. Selanjutnya, kami akan menunjukkan cara kerja prediktor churn, memberi Anda keyakinan untuk melihat peluang penerapan ini di organisasi Anda.

Dalam masalah pembelajaran mesin ini, kami memiliki dua kelas pelanggan: mereka yang cenderung churn dan mereka yang tidak. Oleh karena itu, label yang harus dipelajari oleh model ML kami untuk memprediksi apakah pelanggan tersebut termasuk dalam satu kelas atau yang lain (katakanlah pelanggan yang akan churn termasuk dalam kelas 1, dan yang lainnya termasuk dalam kelas 0). Misalnya, perusahaan telepon dapat memberi label "churn" untuk semua pelanggan yang keluar dari paket teleponnya, sementara "not churn" untuk semua pelanggan lain yang masih menggunakan paket mereka.

Setelah kita mendefinisikan label yang harus diprediksi oleh algoritma kita, mari kita lihat fitur apa saja yang bisa kita gunakan. Ingat bahwa fitur dalam permasalahan ML adalah parameter yang akan dilihat model untuk membedakan antar kelas. Ini bisa berupa atribut pengguna (misalnya, demografi) atau interaksinya dengan produk Anda (misalnya, jumlah penggunaan layanan tertentu dalam sebulan terakhir).

Apa yang baru saja kita jelaskan berbentuk permasalahan pembelajaran terawasi: sebuah algoritma ML diminta untuk mempelajari pemetaan antara serangkaian fitur (karakteristik pelanggan) dan label (berhenti berlangganan/tidak berhenti berlangganan) berdasarkan data historis. Mari kita rangkum langkah-langkah yang diperlukan untuk menyelesaikannya, seperti yang divisualisasikan pada Gambar 3.2:

1. Tentukan tugas ML mulai dari tugas bisnis (mengidentifikasi pelanggan yang kemungkinan akan meninggalkan layanan kita).
2. Identifikasi label dengan jelas: berhenti berlangganan atau tidak berhenti berlangganan.
3. Identifikasi fitur: elemen pelanggan yang kemungkinan memengaruhi kemungkinan churn. Anda dapat menemukan contoh-contoh yang memungkinkan dengan memikirkan apa yang akan Anda perhatikan jika Anda harus melakukan pekerjaan ini sendiri:
 - Usia
 - Berapa lama pelanggan telah menggunakan layanan
 - Uang yang dihabiskan untuk layanan
 - Waktu yang dihabiskan menggunakan layanan dalam dua bulan terakhir
4. Kumpulkan data historis pelanggan yang churn dan pelanggan aktif.
5. Latih model: model ML akan mempelajari cara memprediksi label, berdasarkan fitur-fiturnya.
6. Lakukan inferensi: gunakan model pada data baru untuk mengidentifikasi pelanggan Anda saat ini yang kemungkinan churn.

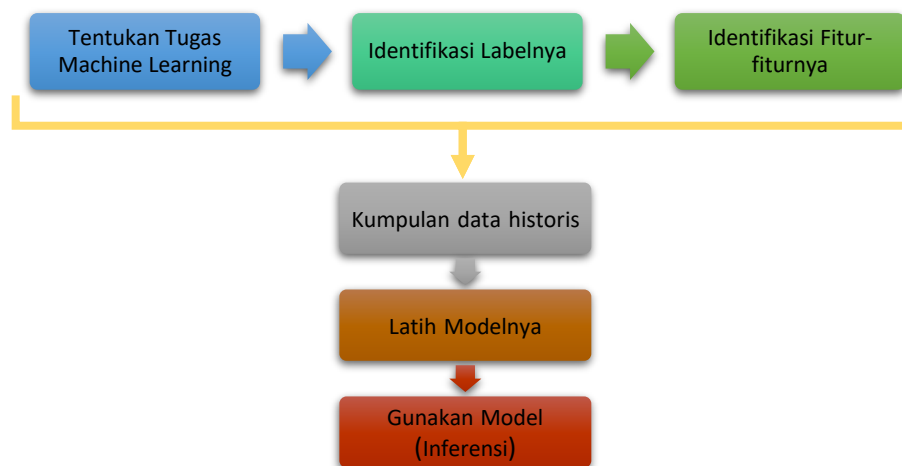
Perhatikan bahwa label harus ditemukan secara retroaktif dengan melihat catatan pelanggan sebelumnya. Mari kita pertimbangkan situasi yang paling mudah terlebih dahulu. Asumsikan Anda memiliki model bisnis berbasis langganan, seperti Netflix atau Spotify. Langganan biasanya diperpanjang secara otomatis, sehingga pelanggan harus secara aktif melakukan tindakan untuk membatalkan langganan: hubungi layanan pelanggan untuk perusahaan telepon, atau kunjungi situs web dan nonaktifkan perpanjangan otomatis untuk Netflix atau Spotify. Dalam situasi seperti ini, menemukan label Anda mudah: tidak ada keraguan apakah pelanggan masih berlangganan atau tidak, dan terdapat tabel basis data yang jelas yang dapat memberi tahu Anda secara tepat kapan hal itu terjadi.

Model bisnis lain lebih rumit untuk ditangani. Anggaplah Anda adalah manajer pemasaran sebuah supermarket, dan Anda menggunakan kartu loyalitas untuk melacak pelanggan Anda setiap kali mereka datang dan berbelanja. Kemungkinan besar, pelanggan yang menemukan supermarket yang lebih baik tidak akan menelepon Anda dan berkata, "Ngomong-ngomong, saya hanya ingin memberi tahu Anda bahwa saya tidak akan kembali berbelanja di supermarket Anda." Sebaliknya, orang ini kemungkinan besar tidak akan muncul lagi, dan hanya itu! Tidak ada jejak yang tersisa, tidak ada kolom Berhenti Berlangganan di basis data Anda, tidak ada label yang mudah. Apakah Anda masih dapat menemukan cara untuk memberi label kepada pelanggan seperti ini? Tentu saja bisa.

Seperti yang Anda lihat pada gambar 3.3, cara yang umum dan sederhana untuk melakukannya adalah dengan melihat pola pembelian dan melihat kapan pola tersebut berubah secara tiba-tiba. Anggaplah sebuah keluarga yang sangat loyal datang untuk membeli bahan makanan setiap hari Minggu. Namun, dalam sebulan terakhir, Anda belum melihat mereka. Anda mungkin berasumsi bahwa mereka telah memutuskan untuk tidak datang lagi, dan karena itu melabeli mereka sebagai "churn".

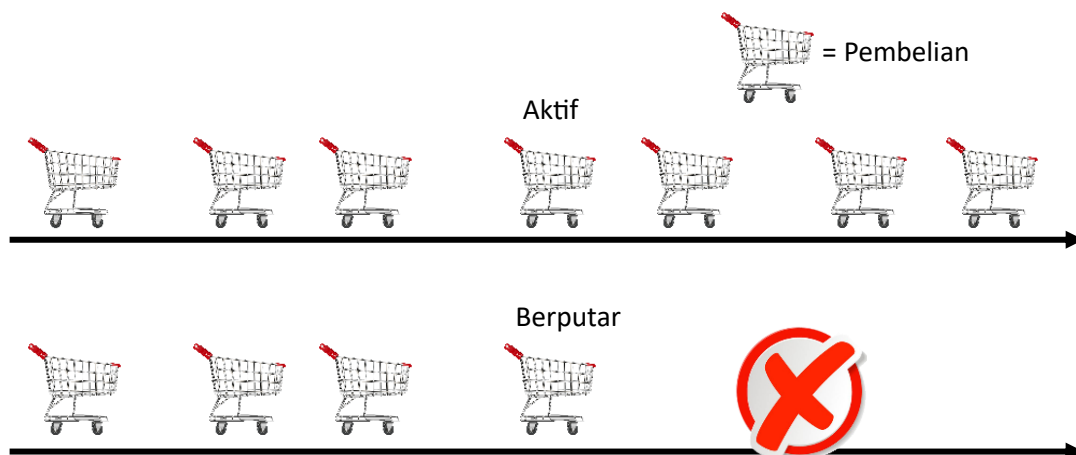
Apakah satu bulan merupakan ambang batas yang tepat? Sulit untuk mengatakannya tanpa konteks tambahan, tetapi untungnya, itu bukan tugas Anda: serahkan tugas untuk mencari tahu kepada ilmuwan data Anda. Yang penting adalah Anda memahami bahwa apa pun bisnisnya, jika Anda memiliki pelanggan yang kembali dan telah mengumpulkan data tentang interaksi mereka kemungkinan ada cara untuk mendefinisikan churn dan mengidentifikasi siapa yang telah pergi dan siapa yang masih aktif.

Setelah Anda menemukan beberapa label untuk membedakan "pelanggan yang puas" dari yang churn, situasinya menjadi serupa dengan contoh prediksi harga rumah yang telah kita lihat sebelumnya. Untungnya, data pelatihan untuk prediksi churn dapat dengan mudah diekstraksi dari sistem manajemen hubungan pelanggan (CRM) perusahaan. Lebih spesifik lagi, kita dapat mengekstrak data CRM hingga, katakanlah, 18 bulan yang lalu, dan kemudian memberi label apakah pelanggan telah churn dalam 6 bulan terakhir.



Gambar 3.2 Proses pembuatan dan penggunaan model ML, dari definisinya hingga penggunaannya dalam fase inferensi

Perilaku pembelian pelanggan aktif vs. pelanggan yang berhenti membeli



Gambar 3.3 Representasi grafis perilaku pola pembelian pelanggan yang churn dan aktif

Saat ini, Anda sudah jauh lebih percaya diri dan efektif dalam menentukan label untuk proyek prediksi churn dibandingkan kebanyakan manajer bisnis. Setiap ilmuwan data akan berterima kasih untuk itu, tetapi jika Anda benar-benar ingin membantu mereka, Anda perlu melakukan upaya ekstra: bantu mereka memilih fitur.

Jika ini terdengar seperti detail teknis bagi Anda, Anda kehilangan kesempatan besar untuk menunjukkan pengalaman dan pengetahuan domain Anda. Dalam permasalahan ML, ingatlah bahwa fitur adalah atribut dari fenomena yang kita coba modelkan yang memengaruhi hasilnya. Dengan asumsi Anda seorang pakar pemasaran, tidak ada seorang pun di dunia yang memiliki wawasan lebih baik tentang fitur-fitur yang relevan, dan keahlian Anda dapat membantu tim ilmu data Anda mengikuti jalur yang mengarah pada hasil yang sukses.

Untuk memberi Anda gambaran tentang seperti apa kontribusi Anda nantinya, tanyakan pada diri sendiri, "Jika saya harus menebak kemungkinan churn hanya dari satu pelanggan, parameter apa yang akan saya perhatikan?" Hal ini dapat menjadi dasar percakapan dengan seorang teknisi:

- Teknisi : Tahukah Anda apa yang memengaruhi churn pelanggan? Saya perlu memikirkan beberapa fitur yang relevan.
- Pemasar : Tentu, kita tahu bahwa pengaturan pembayaran sangat relevan dengan churn. Biasanya, seseorang yang memiliki kontrak, alih-alih kartu prabayar, cenderung tidak meninggalkan layanan karena mereka memiliki ikatan yang lebih kuat. Juga benar bahwa ketika kita mendekati tanggal kedaluwarsa kontrak, pelanggan mulai mencari pesaing, jadi itu faktor lainnya.
- Teknisi : Menarik. Tentu, saya akan menggunakan fitur dalam model yang menyatakan "kontrak" atau "prabayar." Fitur lainnya adalah jumlah hari hingga berakhirnya kontrak. Ada lagi?
- Pemasar : Tentu, kita tahu bahwa usia memainkan peran besar. Generasi milenial muda ini sering berganti perusahaan, sementara orang yang lebih tua lebih loyal. Selain itu, jika seseorang telah menjadi klien kita dalam waktu lama, itu merupakan indikator loyalitas yang baik.
- Teknisi : Bagus; Kita bisa melihat di CRM dan menambahkan fitur untuk "hari sejak mendaftar" dan satu lagi untuk usia. Apakah usia satu-satunya atribut demografis yang menarik?
- Pemasar : Saya rasa gender bukan satu-satunya; kami tidak pernah melihat dampaknya. Pekerjaan itu penting: kita tahu bahwa wiraswasta cenderung kurang bersemangat untuk mengubah rencana.
- Insinyur: Oke, saya akan coba periksa kembali apakah gender memiliki korelasi dengan churn. Mengenai pekerjaan, itu petunjuk yang bagus. Terima kasih!

Percakapan seperti ini bisa berlangsung sehari-hari, biasanya dengan diskusi terus-menerus antara insinyur dan Anda. Anda akan memberikan pengalaman dan pengetahuan domain Anda, dan insinyur akan menerjemahkannya menjadi sesuatu yang dapat dibaca oleh mesin.

Pada akhirnya, insinyur akan kembali dengan beberapa wawasan atau pertanyaan yang muncul dari analisis data dan yang membutuhkan bantuan Anda untuk menafsirkannya. Seperti yang Anda lihat, ini bukan latihan yang membosankan: ini adalah upaya tim antara perusahaan dan para pekerja yang membosankan.

3.3 OPTIMALKAN KONVERSI DAN UPSELLING DENGAN AI

Anda telah melihat bagaimana prediksi churn dapat menjadi aplikasi algoritma klasifikasi yang ampuh. Dalam kasus ini, kelas yang kita beri label pada pelanggan adalah "churn" atau "not churn". Dalam situasi lain, Anda dapat memberi label pelanggan dengan kelas yang relevan dengan departemen pemasaran Anda dan menggunakan algoritma ML untuk membuat prediksi. Salah satu prediksi yang paling umum adalah apakah pelanggan akan membeli layanan berdasarkan penjualan sebelumnya.

Bayangkan Anda memiliki corong pemasaran klasik: pelanggan berlangganan layanan gratis, lalu beberapa dari mereka meningkatkan ke layanan premium. Oleh karena itu, Anda memiliki dua kelas pelanggan:

- *Terkonversi*—Pelanggan yang membeli layanan premium setelah mencoba versi gratis
- *Tidak Terkonversi*—Pelanggan yang terus menggunakan layanan gratis

Perusahaan web mungkin akhirnya menginvestasikan jutaan dolar untuk memaksimalkan jumlah pengguna yang berkonversi ke produk berbayar. Metrik ini sangat penting bagi perusahaan perangkat lunak yang memiliki model bisnis Perangkat Lunak sebagai Layanan (SaaS): perusahaan yang menawarkan layanan yang dibeli dengan berlangganan. Bergantung pada tingkat konversi, bisnis langganan berbasis web dapat bertahan atau mati.

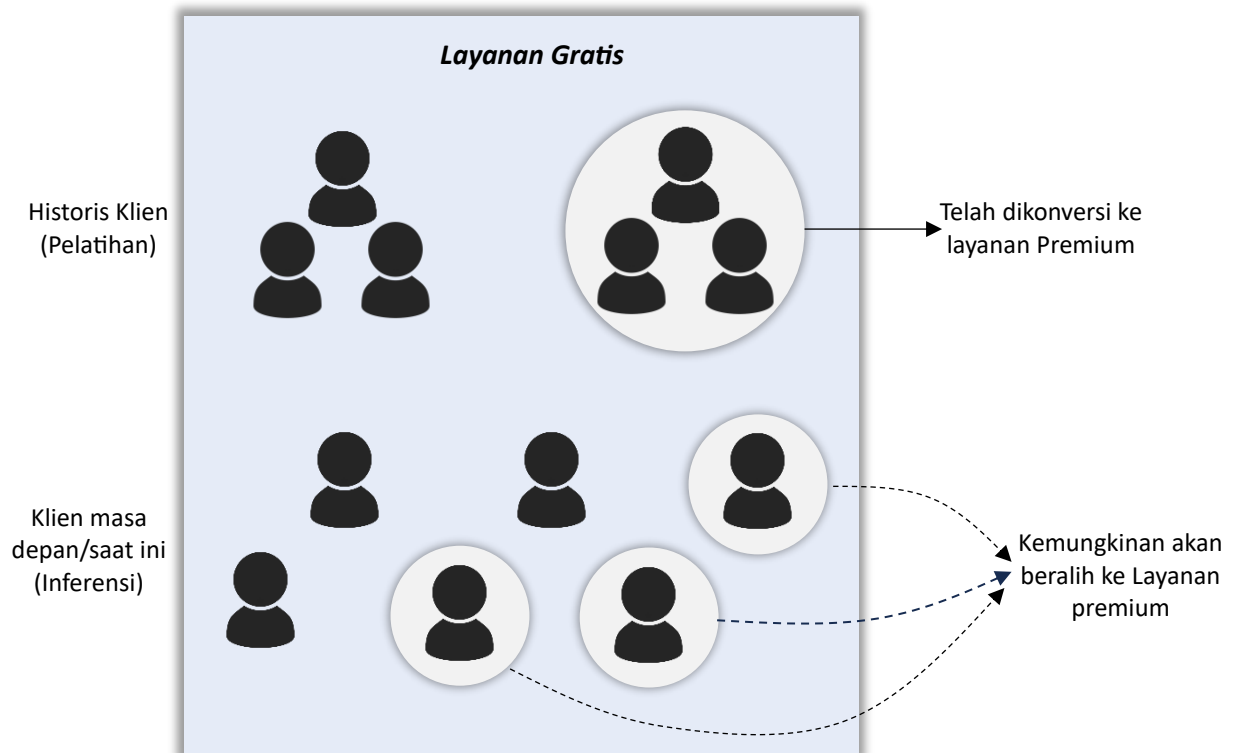
Cara paling naif untuk mencoba meningkatkan tingkat konversi pengguna berbayar adalah dengan menargetkan seluruh basis pengguna secara masif melalui aktivitas pemasaran: buletin, penawaran, uji coba gratis, dan sebagainya. Pemasar yang lebih canggih mungkin berpikir untuk menyiapkan strategi yang rumit guna menilai kemungkinan konversi dan menginvestasikan anggaran pemasaran dengan lebih bijak. Misalnya, kita mungkin percaya bahwa pengguna yang membuka buletin lebih tertarik membeli layanan premium daripada pengguna yang tidak pernah membukanya, dan menargetkan mereka dengan iklan Facebook (pernahkah Anda menerima spam di Facebook setelah mengunjungi situs web atau membuka buletin?).

Mengingat pentingnya topik ini dan jumlah uang yang bergantung padanya, mari kita lihat apakah kita dapat menggunakan ML untuk mengklasifikasikan pengguna berdasarkan kemungkinan konversi mereka, mengoptimalkan biaya pemasaran kita, dan mencapai hasil yang lebih baik. Jika Anda melihat masalahnya, Anda akan melihat bahwa ini sangat cocok untuk pembelajaran mesin. Anda telah melihat bahwa Anda memiliki tugas yang terdefinisi dengan jelas: mengidentifikasi pengguna yang dapat meningkatkan layanan dari gratis ke berbayar. Ini adalah tugas klasifikasi pembelajaran terawasi, dan Anda telah menyiapkan labelnya: katakanlah 1 untuk pengguna yang membeli layanan berbayar, dan 0 untuk pengguna yang tidak. Sekarang Anda perlu memikirkan fitur apa yang akan Anda gunakan untuk melatih pengklasifikasi Anda. Ingatlah bahwa titik awal yang baik untuk mengidentifikasi

fitur adalah dengan bertanya pada diri sendiri pertanyaan ini: "Jika saya harus menebak sendiri kemungkinan konversi, informasi apa yang saya perlukan?" Informasi ini mungkin mencakup hal-hal berikut:

- Penggunaan produk gratis. Ingatlah bahwa ini harus berupa angka aktual, jadi Anda harus menemukan cara yang berguna untuk mendeskripsikannya. Jika Anda menjual layanan seperti Dropbox, penggunaannya dapat dijelaskan dengan beberapa parameter:
 - Jumlah file yang disimpan
 - Jumlah perangkat yang digunakan pengguna untuk masuk (memberikan petunjuk tentang seberapa bermanfaat layanan tersebut bagi pengguna)
 - Jumlah akses per hari/minggu/bulan (menunjukkan seberapa sering pengguna mengandalkannya)
- Tingkat pembukaan buletin (Seberapa tertarik pengguna dengan pesan kami?)
- Berapa lama pengguna berlangganan
- Saluran akuisisi (Seseorang yang berlangganan setelah referensi teman mungkin lebih berharga daripada seseorang yang mengklik iklan Facebook.)

Variabel-variabel ini dapat bervariasi berdasarkan jenis bisnis yang Anda miliki, tetapi konsepnya umumnya sederhana: pikirkan faktor-faktor apa yang dapat mengisyaratkan kemungkinan konversi dan kemudian masukkan ke dalam algoritma pembelajaran mesin Anda. Perlu dicatat bahwa beberapa bisnis memiliki lebih banyak data daripada yang lain: misalnya, layanan internet yang menggunakan login Facebook akan dapat mengetahui semua minat penggunanya sebagai fitur untuk pengklasifikasi tersebut. Dengan asumsi Anda memiliki data historis tentang pelanggan lama yang berkonversi dan yang tidak, Anda dapat melatih algoritme untuk mengidentifikasi bagaimana fitur yang Anda pilih memengaruhi kemungkinan pengguna membeli layanan premium Anda. Setelah fase pelatihan selesai, algoritme Anda akhirnya siap menerapkan apa yang telah dipelajari dari pelanggan lama ke pelanggan lama Anda, memeringkat mereka dari yang paling mungkin berkonversi hingga yang paling kecil kemungkinannya. Seperti yang mungkin Anda ingat dari bab sebelumnya, fase ini disebut inferensi (membuat prediksi berdasarkan data baru setelah algoritme dilatih berdasarkan data lama). Gambar 3.4 mengilustrasikan proses pembelajaran dari perilaku pelanggan lama dan membuat prediksi tentang pelanggan baru.



Gambar 3.4 Untuk meningkatkan rasio konversi, algoritma ML dilatih dengan data pelanggan yang pernah membeli layanan premium. Kemudian, algoritma ini mengidentifikasi pelanggan mana yang paling mungkin melakukan hal yang sama.

Perhatikan bahwa kami menerapkan metodologi ini pada kasus sistem berbasis internet yang menggunakan model freemium (layanan gratis dan peningkatan berbayar), tetapi dapat diterapkan pada kasus lain di mana sekelompok pelanggan melakukan satu tindakan dan kelompok lain melakukan tindakan lain (atau tidak melakukan apa pun). Skenario ini umum terjadi, dan kami ingin mendorong Anda untuk mencari situasi seperti itu dan memikirkan apakah ada ruang untuk membangun pengklasifikasi ML untuk situasi tersebut.

Untuk memberi Anda inspirasi, berikut beberapa kasus lain di mana Anda dapat menerapkan metodologi ini:

- Anda memiliki produk dasar dan beberapa upsell (aksesori atau layanan tambahan, yang umum bagi perusahaan telekomunikasi). Anda dapat memberi label pelanggan dengan "telah membeli upsell X" atau "belum membeli upsell X" dan menggunakan penggunaan produk dasar mereka untuk menilai apakah upsell tersebut layak ditawarkan kepada pelanggan Anda.
- Anda memiliki buletin dan ingin mengoptimalkan rasio bukannya. Label Anda adalah "telah membuka buletin" atau "belum membuka buletin". Fitur yang Anda gunakan untuk pengklasifikasi dapat berupa waktu Anda mengirim email (hari, jam, dan sebagainya) dan beberapa fitur terkait pengguna, dan Anda juga dapat menandai email berdasarkan kontennya (misalnya, "informatif", "berita produk", atau "whitepaper").
- Anda memiliki toko fisik dengan kartu loyalitas (untuk melacak pelanggan mana yang membeli apa). Anda dapat menjalankan inisiatif pemasaran (buletin lagi atau juga iklan

fisik) dan mengklasifikasikan pengguna Anda berdasarkan apa yang menarik mereka ke toko Anda dan apa yang tidak. Seperti yang Anda lihat, metode yang baru saja kami jelaskan, yaitu membagi pengguna ke dalam dua kelas terpisah dan membangun pengklasifikasi ML yang dapat mengenali keduanya, cukup fleksibel dan dapat digunakan untuk banyak masalah. Cukup ampuh, bukan?

3.4 MELAKUKAN SEGMENTASI PELANGGAN OTOMATIS

Dalam pengantar bab ini, kami merujuk pada salah satu aktivitas utama yang harus dilakukan pemasar saat mengembangkan rencana pemasaran: segmentasi pelanggan. Segmentasi pasar berarti membagi pelanggan yang memiliki karakteristik dan perilaku serupa ke dalam kelompok-kelompok. Gagasan inti di balik upaya ini adalah bahwa pelanggan dalam kelompok yang sama akan responsif terhadap tindakan pemasaran yang serupa. Misalnya, peritel fesyen kemungkinan akan diuntungkan dengan memiliki segmen pasar terpisah untuk pria versus wanita, dan remaja versus dewasa muda versus profesional.

Segmen bisa lebih spesifik atau kurang spesifik, dan oleh karena itu lebih detail atau kurang terperinci. Berikut dua contohnya:

- *Segmen luas*—Pria muda berusia antara 20 dan 25 tahun
- *Segmen sangat spesifik*—Pria muda berusia antara 20 dan 25 tahun, sedang kuliah, tinggal di salah satu dari lima kota terbesar di AS, dan memiliki minat pada gim video first-person shooter

Banyak pemasar dapat secara intuitif melakukan tugas segmentasi ini di dalam otak mereka selama jumlah data terbatas, baik dari segi contoh (jumlah pelanggan) maupun fitur. Hal ini biasanya menghasilkan segmen pelanggan generik seperti yang pertama, yang dapat membatasi mengingat banyaknya variasi yang ada di antara kelompok-kelompok ini. Seorang pemasar dapat mencoba mendefinisikan segmen yang lebih spesifik seperti yang kedua, tetapi bagaimana mereka menemukannya? Berikut adalah pertanyaan-pertanyaan yang dapat diajukan selama sesi curah pendapat yang umum:

- Apakah menggunakan ambang batas usia 20 hingga 25 tahun merupakan ide yang baik, atau lebih baik menggunakan 20 hingga 28 tahun?
- Apakah kita yakin bahwa mahasiswa yang tinggal di kota besar pada dasarnya berbeda dengan yang tinggal di kota kecil? Tidak bisakah kita memasukkan semuanya ke dalam satu klaster?
- Apakah ada perbedaan mendasar antara pria dan wanita? Apakah kita benar-benar perlu membuat dua segmen, atau ini hanya klise?

Menjawab pertanyaan-pertanyaan ini dapat dilakukan dengan tiga cara:

- Ikuti firasat Anda. Kita tidak berada di tahun 1980, jadi jangan lakukan itu.
- Lihat datanya, dan gunakan insting pemasar untuk menafsirkannya. Ini lebih baik daripada firasat, tetapi pemasar kemungkinan akan memproyeksikan bias mereka ke dalam analisis mereka dan melihat apa yang ingin mereka lihat. Jadi hindari hal ini juga.

- Biarkan AI menghasilkan segmen pelanggan sendiri, sehingga pemasar tetap mendapatkan informasi terkini untuk menggunakan kreativitas dan pengetahuan konteks mereka.

Opsi 3 kemungkinan besar akan mengungguli yang lain. Mari kita lihat mengapa dan bagaimana.

Pembelajaran Tanpa Pengawasan (atau Pengelompokan)

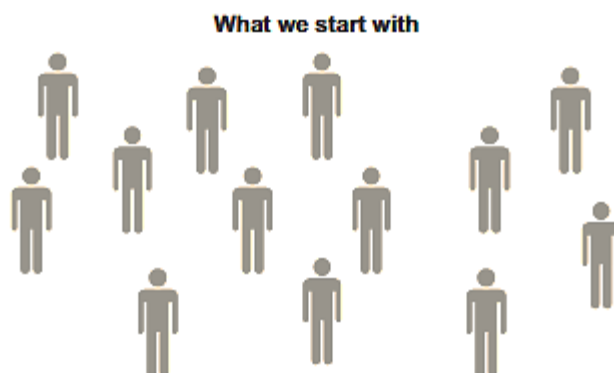
Mari kita lihat masalah yang baru saja kita jelaskan:

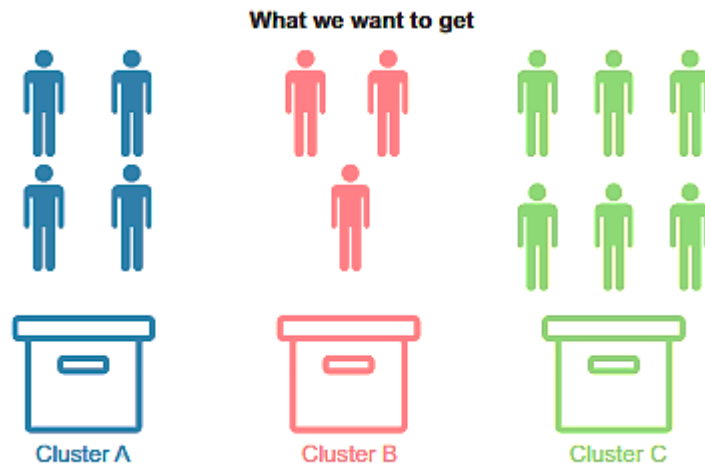
- *Yang kita mulai:* sekelompok pelanggan, dengan serangkaian elemen yang mencirikan mereka (usia, lokasi, minat, dan sebagainya)
- *Yang kita inginkan:* sejumlah segmen yang dapat kita gunakan untuk membagi pelanggan kita

Anda dapat membayangkan masalah ini seperti memiliki sekelompok pelanggan dan perlu menempatkan masing-masing ke dalam sebuah wadah, yang akan kita sebut kluster (gambar 3.5).

Karakteristik pelanggan yang akan kita gunakan menyerupai apa yang telah kita sebut fitur di bab-bab sebelumnya, jadi Anda mungkin berpikir bahwa kita sedang menangani jenis tugas yang sama dan dapat menggunakan alat yang sama yang telah kita jelaskan. Namun, masalahnya ada pada detailnya: karena kita tidak tahu sebelumnya kelompok yang ingin kita definisikan, kita tidak tahu label mana yang harus diterapkan. Sejauh ini dalam buku ini, kita telah menggunakan subset ML yang disebut pembelajaran terawasi. Resep umum untuk tugas pembelajaran terawasi adalah sebagai berikut:

1. Kami memiliki data sekelompok pelanggan yang dicirikan oleh fitur-fitur tertentu.
2. Pelanggan-pelanggan ini juga memiliki label: nilai target yang ingin kami prediksi (misalnya, apakah mereka berhenti berlangganan atau tidak).
3. Algoritma pembelajaran terawasi menelusuri data pelanggan, dan mempelajari pemetaan umum antara fitur dan label.



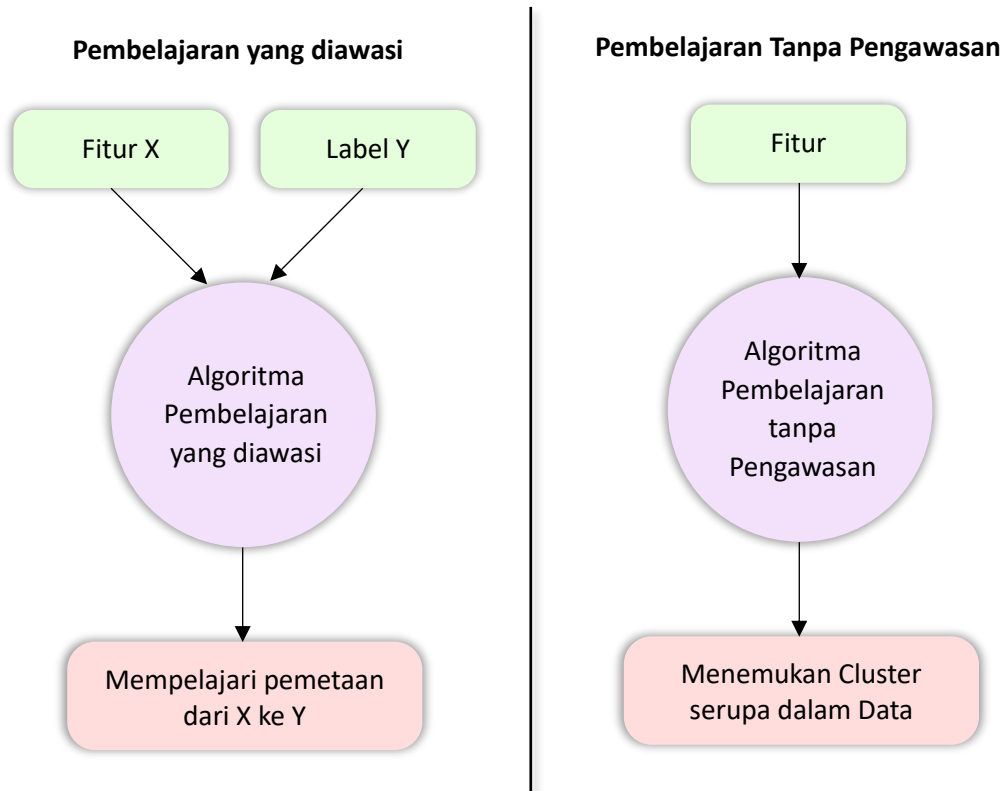


Gambar 3.5 Algoritma pengelompokan membagi sekelompok pelanggan seragam menjadi tiga klaster.

Dalam skenario baru kami, poin 2 hilang: kami tidak memiliki label yang melekat pada setiap pengguna. Inilah yang ingin kami temukan oleh algoritma baru kami. Oleh karena itu, beginilah tampilan tugas baru kami:

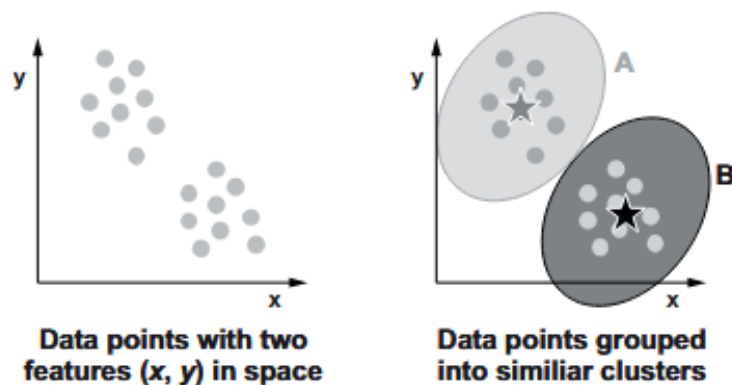
1. Seperti sebelumnya, kami memiliki data tentang sekelompok pelanggan yang dicirikan oleh fitur-fitur tertentu.
2. Kami ingin membagi pelanggan menjadi sejumlah segmen (klaster)—misalkan tiga.
3. Kami menjalankan semacam algoritma pembelajaran mesin (ML) yang, dengan melihat data, menentukan klaster terbaik yang dapat kami buat dan membagi pengguna ke dalam klaster tersebut.

Algoritma ML jenis baru ini disebut pengelompokan, atau pembelajaran tanpa pengawasan. Pembelajaran tanpa pengawasan adalah bentuk lain dari ML di mana suatu algoritma diberi serangkaian contoh tanpa label (hanya serangkaian parameter) dan diminta untuk membagi contoh-contoh tersebut ke dalam kelompok-kelompok yang memiliki beberapa kesamaan. Dalam hal ini, algoritma pembelajaran tanpa pengawasan menggunakan konsep kesamaan untuk mengatasi ketiadaan label yang telah ditentukan sebelumnya, dengan membagi contoh yang diberikan ke dalam kelompok-kelompok secara otonom. Inilah perbedaan utama antara pembelajaran terawasi dan tanpa pengawasan: algoritma pembelajaran terawasi mempelajari pemetaan antara serangkaian fitur dan label, sedangkan algoritma tanpa pengawasan hanya mengamati label dan mengelompokkan titik data berdasarkan klaster yang memiliki kesamaan tertentu, seperti yang ditunjukkan pada Gambar 3.6.



Gambar 3.6 Perbedaan masukan dan keluaran algoritma tersupervisi dan tak tersupervisi

Tugas menemukan kelompok serupa dalam kumpulan data cukup mudah ketika dimensi yang perlu kita perhitungkan terbatas. Perhatikan Gambar 3.7, dan Anda dapat melihat bahwa titik-titik tersebut secara alami mengembun menjadi dua kelompok yang terpisah dengan baik. Namun, apa yang terjadi ketika kita ingin mempertimbangkan sejumlah besar karakteristik pengguna? Jika kita ingin mempertimbangkan, katakanlah, 10 atribut, pikiran kita mustahil mengidentifikasi kelompok-kelompok yang serupa satu sama lain. Di sinilah algoritma pembelajaran tak tersupervisi bersinar: algoritma tersebut dapat menskalakan konsep kesamaan bahkan hingga ratusan dimensi tanpa masalah, dan kita dapat menggunakan hasilnya untuk mendapatkan wawasan yang bermanfaat.



Gambar 3.7 Efek pengelompokan pada sekumpulan titik sederhana dengan dua fitur

Saat ini, Anda seharusnya sudah memiliki firasat mengapa jenis pembelajaran mesin yang telah kita gunakan sejauh ini disebut tersupervisi: algoritma kita diminta untuk memetakan serangkaian fitur tertentu ke serangkaian label tertentu. Karena kami tidak memiliki label sejak awal, algoritma harus menemukannya sendiri, tanpa pengawasan.

Anda dapat membayangkannya seperti ini:

- Dalam pembelajaran terawasi, Anda sudah tahu apa yang Anda cari. Jika Anda dapat mengklasifikasikan pelanggan ke dalam kelas yang berbeda (misalnya, pelanggan yang berhenti berlangganan/tidak berhenti berlangganan seperti yang telah kita lakukan sebelumnya), algoritma pembelajaran mesin dapat mempelajari cara mengenali pelanggan yang termasuk dalam salah satu kelas tersebut.
- Dalam pembelajaran tanpa pengawasan, Anda tidak tahu persis apa yang Anda cari: Anda tidak dapat memberi label kepada pelanggan. Algoritma pembelajaran tanpa pengawasan akan mengenali kelompok pelanggan yang serupa dan memberi mereka label. Namun, algoritma tersebut tidak akan memberi tahu Anda apa arti label tersebut: algoritma hanya akan memberi tahu Anda bahwa pelanggan dengan label A serupa satu sama lain dan berbeda dari pelanggan dengan label B; Anda sendiri yang harus memahami alasannya.

Mari kita lihat bagaimana percakapan antara pakar ML dan pemasar yang telah membaca buku ini akan terungkap:

Pemasar : Saya sedang mencari cara agar ML dapat membantu tim kami meningkatkan segmentasi pelanggan.

Pakar ML : Bagaimana Anda melakukan segmentasi pelanggan sebelumnya?

Pemasar : Anda tahu, "cara lama yang ampuh": menggunakan gabungan survei, pengalaman, dan firasat. Saya tahu ada beberapa teknik ML yang dapat membantu.

Pakar ML : Ya, saya dapat menggunakan pembelajaran tanpa pengawasan untuk menghasilkan kluster secara otomatis. Mari kita mulai dengan sesuatu yang sederhana: apa tiga elemen teratas yang relevan untuk mensegmentasi pelanggan kita?

Pemasar : Tentu saja, demografi seperti usia dan jenis kelamin, dan saya akan menambahkan pengeluaran rata-rata bulanan ke dalamnya. Ini merupakan indikator yang baik untuk kemungkinan mereka membeli layanan baru dari kami.

Pakar ML : Bagus. Saya akan mengeksplor dimensi ini dari CRM kami untuk 1.000 klien dan akan menghubungi Anda kembali. Saya membutuhkan bantuan Anda untuk menginterpretasikan hasilnya. [Setelah beberapa saat]

Pakar ML : Saya telah melakukan pengelompokan awal, dan sepertinya kita memiliki tiga kelompok yang terdefinisi dengan baik: pria muda dengan pengeluaran rendah, wanita dengan pengeluaran tinggi berusia 30-an, dan satu kelompok di antaranya.

Pemasar : Menarik—kami tidak tahu bahwa wanita berusia 30-an merupakan segmen yang begitu menguntungkan bagi kami. Saya ingin menggali lebih dalam; bisakah kita menambahkan dimensi lain ke dalam pengelompokan ini? Saya tertarik untuk

mengetahui frekuensi pembelian mereka: kita tahu bahwa wanita lebih suka membeli daripada pria, dan saya ingin tahu apakah pembelajaran tanpa pengawasan dapat memahami sesuatu yang lebih mendalam.

Pakar ML : Tentu, mari kita definisikan label "rata-rata waktu antar pesanan". Saya akan melihat hasilnya.

Sebagai seorang pebisnis, penting bagi Anda untuk memulai percakapan ini dengan pengetahuan tentang pembelajaran tanpa pengawasan agar Anda dapat berdiskusi secara konstruktif dengan para ahli teknis, dan dengan pola pikir terbuka yang dapat menerima masukan yang mereka berikan kepada Anda. Cara yang baik untuk memvisualisasikan pengelompokan adalah dengan membayangkan bagaimana bintang-bintang tersebar di langit malam. Otak kita secara intuitif mengelompokkan bintang-bintang tetangga dan menetapkan tanda zodiak untuk mereka. Sebagian besar aplikasi di dunia nyata sedikit lebih kompleks, karena tiga alasan utama:

- Terkadang, titik data terdistribusi secara homogen, sehingga sulit untuk menentukan berapa banyak kluster yang perlu dipertimbangkan, apalagi bagaimana membagi pelanggan.
- Sebagai manusia, kita dengan mudah melakukan segmentasi pada dimensi yang terbatas, tetapi kita kesulitan ketika jumlah dimensinya bertambah. Pikirkan kembali contoh tanda zodiak: kita melihat langit sebagai kanvas dua dimensi; melakukan tugas yang sama dalam ruang tiga dimensi akan jauh lebih sulit. Dalam empat dimensi, hal itu mustahil. Bagaimana dengan 20 dimensi informasi pelanggan?
- Dari sudut pandang bisnis, segmentasi bukanlah kegiatan yang sia-sia, tetapi paling berguna hanya ketika berbagai segmen dapat dikaitkan dengan hasil bisnis, seringkali dalam hal nilai umur pelanggan, sensitivitas harga, atau preferensi saluran.

Di bagian selanjutnya, kita akan membahas masalah ini melalui sebuah contoh. Untuk saat ini, mari kita gali lebih dalam seluk-beluk pengelompokan. Salah satu keputusan penting pertama yang harus diambil saat menangani masalah pengelompokan adalah menentukan fitur (atau dimensi) mana yang akan digunakan untuk pengelompokan. Dalam contoh sederhana langit malam kita, pilihan dimensinya jelas: posisi horizontal dan vertikal setiap bintang. Namun, penerapannya di dunia nyata mungkin jauh lebih kompleks.

Pembelajaran tanpa Pengawasan untuk Segmentasi Pelanggan

Bagian ini akan memberikan detail lebih lanjut tentang pembelajaran tanpa pengawasan dan menjelaskan cara kerja internal algoritma ini. Konsep inti yang telah Anda pelajari sejauh ini cukup untuk membayangkan aplikasi baru pembelajaran tanpa pengawasan di organisasi Anda. Jika itu sudah cukup untuk tujuan Anda, silakan lewati bagian ini. Jika Anda ingin tahu lebih banyak tentang cara menggunakan teknik ini dalam praktik, teruskan membaca.

Mari kita gunakan contoh situs web e-commerce yang menjual sepatu dan memiliki akses ke riwayat pembelian pelanggannya. Setiap titik data mewakili pembelian atau pengembalian, dan berisi informasi tentang sepatu seperti harga, merek, ukuran, warna,

tanggal dan waktu transaksi, dan apakah sepatu tersebut dibeli bersama barang lain. Kita dapat memutuskan untuk menggunakan semua fitur ini untuk segmentasi atau membatasi analisis kita pada sebagian kecil saja. Misalnya, melihat warna semua sepatu yang dibeli pelanggan dapat membantu kita lebih memahami selera pelanggan tersebut, atau melihat waktu pembelian tersebut dapat memberikan saran tentang waktu terbaik untuk memberikan diskon.

Kita bahkan dapat mengekstrapolasi parameter seperti jumlah rata-rata yang dibelanjakan per pembelian dan jumlah sepatu yang dibeli dalam sebulan. Kedua informasi ini bersama-sama kemungkinan akan membantu algoritma pengelompokan menemukan perbedaan alami antara pelanggan frekuensi tinggi/nilai rendah dan pelanggan frekuensi rendah/nilai tinggi. Untuk menyederhanakan, mari kita asumsikan kita sedang membangun algoritma pengelompokan sederhana yang melihat tiga fitur setiap pelanggan:

- Usia
- Jenis Kelamin
- Rata-rata pengeluaran bulanan di e-commerce

Perlu diingat bahwa atribut yang tidak digunakan untuk pengelompokan tidak dibuang tetapi dapat digunakan untuk pembuatan profil. Ini berarti mendeskripsikan karakteristik setiap kelompok untuk menginformasikan keputusan pemasaran. Tabel 3.1 menunjukkan data kita untuk lima pelanggan pertama.

Tabel 3.1 Lima fitur pelanggan sebelum dimasukkan ke algoritma pembelajaran tanpa pengawasan

ID Pelanggan	Usia	Jenis Kelamin	Pengeluaran Bulanan rata-rata (RP. X00.000)
1	18	L	14.67
2	21	L	15.67
3	28	L	18.02
4	27	P	34.61
5	32	P	30.66

Dalam beberapa algoritma pengelompokan yang paling umum digunakan, langkah selanjutnya adalah menentukan jumlah kluster yang kita cari. Hal ini seringkali berlawanan dengan intuisi: Lagipula, bukankah Anda ingin algoritma memberi tahu Anda berapa banyak kelompok pengguna yang ada? Namun, jika dipikir-pikir, ada banyak cara untuk mengiris dan membagi populasi, dan memilih jumlah kluster di awal adalah satu-satunya cara untuk mengarahkan algoritma. Mari kita sederhanakan untuk saat ini dan katakanlah kita ingin mendapatkan tiga kluster. Algoritma pengelompokan akan menemukan cara untuk membagi pengguna sedemikian rupa sehingga:

- Pelanggan dalam kluster yang sama serupa satu sama lain.
- Pelanggan di kluster yang berbeda berbeda satu sama lain.

Dengan cara ini, kita secara matematis yakin bahwa ketika kita memilih tindakan untuk menargetkan pelanggan dari kluster tertentu, kita memaksimalkan kemungkinan mereka akan merespons dengan cara yang sama. Dua keluaran dari algoritma pengelompokan ini menarik untuk dicermati:

- Kluster yang akan dikaitkan dengan setiap pengguna, seperti yang ditunjukkan pada tabel 3.2.
- Pusat-pusat kluster. Setiap kluster memiliki sebuah pusat, yang dapat dianggap sebagai "stereotipe" dari jenis pengguna yang termasuk dalam kluster tersebut. Seorang pemasar akan menyebutnya sebagai persona pembeli.

Tabel 3.2 Menambahkan kolom Kluster, yang diidentifikasi oleh algoritma pembelajaran tanpa pengawasan

ID Pel	Usia	Jenis Kelamin	Pengeluaran Bulanan rata-rata (RP. X00.000)	Kelompok
1	18	L	14.67	1
2	21	L	15.67	1
3	28	L	18.02	3
4	27	P	34.61	2
5	32	P	30.66	2

Melihat pusat-pusat kluster sangatlah penting, karena memberikan informasi kuantitatif tentang temuan algoritma, yang kemudian dapat diinterpretasikan untuk mendapatkan wawasan. Setiap pusat akan dicirikan oleh tiga fitur yang sama yang telah kita gunakan sebelumnya untuk mendeskripsikan pengguna (meskipun kita mungkin akan menyusunnya kembali agar lebih bermakna). Biasanya, kita juga akan menambahkan hitungan jumlah pengguna yang termasuk dalam setiap kluster. Tabel 3.3 menunjukkan seperti apa pusat-pusat kluster tersebut, dengan asumsi kita memulai dengan data dari 1.000 pelanggan.

Tabel 3.3 Ringkasan karakteristik ketiga kluster yang ditemukan oleh algoritma kami

Nomor Kluster	Usia	% Perempuan	Pengeluaran Bulanan rata-rata (RP. X00.000)	#Pelanggan
1	18.2	20 %	15.24	290
2	29.3	90 %	28.15	120
3	22	40 %	17.89	590

Tabel yang tampaknya tidak berbahaya ini sarat dengan informasi bermanfaat. Mari kita luangkan waktu untuk mencoba mengekstrapolasi wawasan dan memahami segmen kita:

- Kluster 1 sebagian besar terdiri dari kaum muda (usia rata-rata: 18,2), didominasi laki-laki, yang tidak terlalu banyak berbelanja (rata-rata pengeluaran bulanan Rp. 1.524.000). Ini adalah kluster berukuran rata-rata (29% pengguna ada di sini).
- Kluster 2 cenderung didominasi perempuan yang lebih tua (rata-rata 29,3 tahun) yang menghabiskan jauh lebih banyak daripada kluster lainnya (rata-rata pengeluaran Rp.

2.815.000 dibandingkan dengan Rp. 1.524.000 dan Rp. 1.789.000 untuk klaster 1 dan 3). Ini adalah segmen yang agak kecil, dengan 12% pengguna termasuk di dalamnya.

- Klaster 3 hampir terbagi rata antara laki-laki dan perempuan. Mereka tidak sehemat dan muda klaster 1, tetapi jelas menghabiskan lebih sedikit daripada klaster 2 dan jauh lebih muda (22 tahun versus 29,3 tahun). Pemasar dapat memanfaatkan informasi ini dengan baik dan merancang strategi yang dipersonalisasi untuk memasarkan produk kepada setiap kelompok orang. Misalnya, pelanggan di klaster 1 dan 3 dapat ditawarkan produk yang lebih murah dibandingkan dengan yang ada di klaster 2. Menambahkan fitur baru seperti warna sepatu yang dibeli dapat memberi kita wawasan lebih lanjut untuk mengambil keputusan yang lebih rinci.

Perhatikan bagaimana proses dimulai dengan data dan temuan algoritma tentangnya, tetapi berakhir dengan manusia yang melihat hasilnya dan menafsirkan pusat klaster untuk mengekstrak informasi bermanfaat yang dapat ditindaklanjuti. Aspek ini merupakan kunci bagi setiap algoritma ML, tetapi sangat penting untuk pengelompokan: nilai sebenarnya dicapai melalui simbiosis antara apa yang dapat dilakukan AI dan apa yang dapat dikembangkan oleh manusia yang ahli dan ingin tahu.

3.5 PENGUKURAN KINERJA PREDIKSI BERBASIS MACHINE LEARNING

Karena prediksi berbasis ML memiliki dampak langsung pada hasil bisnis, mengevaluasi kinerjanya merupakan keterampilan penting yang harus dimiliki. Peneliti dan insinyur akan sering mengembangkan model untuk Anda, dan melaporkan kinerja algoritma dengan menggunakan berbagai metrik. Meskipun metrik-metrik ini cukup baik dalam menggambarkan kinerja statistik model, metrik-metrik ini tidak menceritakan keseluruhan cerita. Faktanya, hubungan antara angka-angka rumit ini dan hasil bisnis bisa jadi samar. Tugas Anda adalah memahami metrik ML secara memadai agar dapat membuat keputusan yang tepat tentang bagaimana akurasi model dapat memengaruhi tujuan bisnis Anda.

Algoritma Klasifikasi

Bagian penting dari bekerja dengan pembelajaran mesin adalah membiasakan diri dalam menangani kesalahan. Bahkan algoritma dengan kinerja terbaik pun tidak akan 100% sempurna, dan akan terjadi kesalahan dalam mengklasifikasikan beberapa contoh. Ingatlah bahwa proses membangun aplikasi ML adalah melakukan pelatihan pada data historis terlebih dahulu, lalu menggunakannya "di dunia nyata": penting untuk mengetahui berapa banyak kesalahan yang mungkin akan dibuat algoritma setelah diterapkan, dan jenis kesalahannya.

Cara paling sederhana dan paling naif untuk mengevaluasi suatu algoritma adalah dengan metrik yang disebut akurasi, yang mewakili persentase tebakan yang benar terhadap total prediksi:

$$Akurasi = \frac{Prediksi\ Benar}{Total\ Prediksi}$$

Namun, tidak semua kesalahan sama. Dalam kasus prediktor churn kami, ada dua kemungkinan kesalahan:

- Seorang pelanggan salah diberi label oleh algoritma sebagai pelanggan yang churn, tetapi sebenarnya masih aktif. Kasus ini disebut positif palsu (FP).
- Seorang pelanggan salah diberi label oleh algoritma sebagai pelanggan aktif, tetapi sebenarnya akan membatalkan langganannya mereka. Kasus ini disebut negatif palsu (FN).

Perbedaan ini bukanlah sesuatu yang hanya dipikirkan oleh para ahli: perbedaan ini berdampak langsung pada bisnis, dan penting bagi Anda untuk memahami bagaimana hal ini akan memengaruhi keputusan bisnis Anda. Seorang ilmuwan data yang baik seharusnya tidak merepresentasikan kinerja dengan angka akurasi yang sederhana, dan Anda pun seharusnya tidak. Ide yang lebih baik adalah menggunakan tabel yang lebih informatif seperti tabel 3.4.

Tabel 3.4 Visualisasi kemungkinan kombinasi prediksi algoritma dan kebenaran dasar

Sebenarnya (Apa yang dikatakan data)		Diprediksi (Apa yang diprediksi oleh algoritma)	
		Benar-benar positif (Pelanggan yang diidentifikasi oleh algoritma sebagai pelanggan yang berhenti berlangganan, dan ternyata benar)	Negatif palsu (Pelanggan yang diidentifikasi oleh algoritma sebagai tidak churn, tetapi ternyata salah)
		Positif palsu (Pelanggan yang diidentifikasi oleh algoritma sebagai pelanggan yang berhenti berlangganan, tetapi ternyata salah)	Negatif yang sebenarnya (Pelanggan yang diidentifikasi oleh algoritma sebagai tidak churn, dan itu benar)

Menyajikan hasil dalam tabel jenis ini umum untuk tugas klasifikasi biner (di mana label hanya memiliki satu dari dua hasil dalam hal ini, churn/tidak churn), jadi mari kita luangkan waktu untuk menjelaskan cara membacanya. Tujuan tabel ini adalah untuk mengukur kinerja algoritma dalam satu snapshot, baik dari segi tebakan yang benar maupun kesalahan.

Sel kiri atas dan kanan bawah relatif terhadap tebakan yang benar. Sel kiri atas mewakili jumlah pelanggan yang churn dan diklasifikasikan dengan benar oleh algoritma kami (positif sejati, atau TP). Sel kanan bawah untuk pelanggan aktif (tidak churn) yang diidentifikasi dengan benar oleh algoritma (negatif sejati, atau TN).

Sel kiri bawah dan kanan atas relatif terhadap kesalahan yang telah kami sebutkan sebelumnya. Pelanggan yang aktif tetapi salah diklasifikasikan oleh algoritma sebagai pelanggan yang berhenti berlangganan berada di kiri bawah (positif palsu, atau FP), dan pelanggan yang aktif tetapi salah diklasifikasikan sebagai pelanggan yang berhenti berlangganan berada di kanan atas (negatif palsu, atau FN).

Para peneliti dan teknisi menghabiskan banyak waktu untuk mempelajari tabel seperti tabel 3.4, karena angka-angka yang dikandungnya sering kali memberikan wawasan tentang

bagaimana kinerja algoritma di dunia nyata. Sebagai seseorang yang bisnisnya terdampak oleh angka-angka ini, penting bagi Anda untuk memahami nuansa yang tersembunyi di baliknya.

Pertama, jumlah positif palsu dan negatif palsu saling terkait, dan mudah untuk menukar salah satunya tanpa perubahan substansial pada model atau pelatihan tambahan. Misalnya, pertimbangkan model yang sangat naif yang selalu memprediksi bahwa pelanggan sedang berhenti berlangganan: apa pun masukan yang diterimanya, model tersebut selalu menghasilkan "ya, pelanggan akan berhenti berlangganan." Metrik untuk hasil positif benar dan hasil negatif salah akan menggembirakan (100% untuk hasil positif benar, dan 0% untuk hasil negatif salah apakah Anda mengerti maksudnya?), tetapi hasil positif salah dan hasil negatif benar akan sangat buruk.

Situasi Rumit dan Cara Menghindari Tipuan

Bayangkan Anda sedang membangun prediktor churn berdasarkan kumpulan data sampel berisi 1.000 pengguna, yang dibagi sebagai berikut:

- 900 di antaranya aktif.
- 100 di antaranya churn.

Jenis data seperti ini disebut tidak seimbang, karena jumlah contoh yang mewakili satu label jauh lebih tinggi daripada jumlah contoh yang mewakili label lainnya. Sekarang, katakanlah Anda membangun sebuah algoritma yang memiliki akurasi prediksi 90%. Kedengarannya bagus, tetapi benarkah? Jika Anda membangun algoritma untuk memprediksi "Aktif" setiap saat, hasilnya akan akurat 90% (karena 90% pengguna aktif). Dalam kasus ini, kita akan mendapatkan hasil berikut:

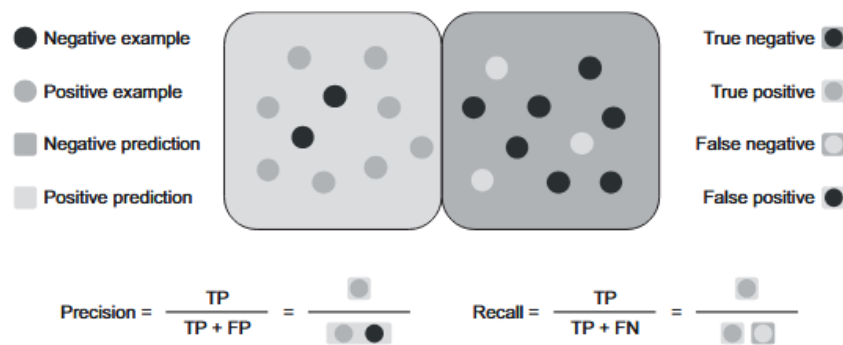
- 100% negatif benar (Semua pelanggan yang tidak churn diidentifikasi dengan benar.)
- 0% positif benar (Algoritme tidak pernah mendeteksi satu pun pelanggan yang churn.)

Kenyataannya, algoritma Anda berkinerja buruk, meskipun akurasinya 90% dari waktu. Seperti yang Anda lihat, melihat positif benar dan negatif palsu adalah strategi yang jauh lebih berguna untuk menghindari tertipu oleh kinerja akurasi tinggi.

Positif palsu, negatif palsu, positif benar, dan negatif benar adalah metrik absolut. Sebaiknya selalu laporkan ke metrik absolut yang tidak sensitif terhadap jumlah sampel yang digabungkan. Dua metrik dapat membantu kita dalam hal ini, yang disebut presisi dan recall. Berikut informasi yang diberikan oleh metrik tersebut:

- *Presisi*—Dari semua pelanggan yang diprediksi oleh algoritma akan berhenti berlangganan (positif benar), berapa banyak yang benar-benar akan berhenti berlangganan?
- *Recall*—Dari semua pelanggan yang berhenti berlangganan, berapa banyak yang diprediksi oleh algoritma?

Gambar 3.8 menunjukkan bagaimana presisi dan recall dihitung berdasarkan prediksi algoritma.



Gambar 3.8 Representasi grafis dari positif benar, negatif benar, positif salah, dan negatif salah, serta bagaimana gabungan metrik ini membentuk presisi dan recall.

Anda dapat membayangkan algoritma dengan presisi tinggi dan recall rendah sebagai penembak jitu: ia ingin memastikan sebelum menembak, sehingga ia tetap konservatif dan tidak menembak kecuali 100% yakin. Ini berarti ia melewatkan beberapa target, tetapi setiap kali menembak, Anda yakin akan mengenai target yang tepat. Di sisi lain, algoritma dengan recall tinggi dan presisi rendah seperti senapan mesin: ia menembak banyak dan mengenai banyak target yang seharusnya, tetapi juga mengenai beberapa target yang salah di sepanjang jalan. Tabel 3.5 merangkum perbedaan antara presisi, recall, dan akurasi, serta apa yang ditunjukkan oleh metrik-metrik ini.

Tabel 3.5 Akurasi, presisi, dan recall

Metrik	Definisi	Apa yang metrik ini sampaikan kepada Anda
Akurasi	$\frac{\text{True Positive} + \text{True Negative}}{\text{Total Example}}$	Berapa banyak kesalahan yang dibuat algoritma? Tidak ada perbedaan dalam salah mengklasifikasikan sesuatu yang positif dari yang negatif.
Presisi	$\frac{\text{True Positive}}{\text{True Positive} + \text{False positive}}$	Seberapa besar Anda harus memercayai algoritma saat menemukan contoh positif?
Recall	$\frac{\text{True Positive}}{\text{True positive} + \text{False Negative}}$	Dari sekian banyak contoh positif, berapa banyak yang berhasil ditemukan oleh algoritma?

Algoritma Pengelompokan

Dalam pembelajaran tanpa pengawasan (unsupervised learning), mengevaluasi kinerja menjadi rumit karena tidak ada metrik "kehebatan" yang objektif: karena kita tidak memiliki label untuk dibandingkan, kita tidak dapat menentukan apakah keluaran algoritma tersebut "benar" atau "salah". Ingat juga bahwa sebagian besar algoritma pengelompokan mengharuskan Anda menentukan jumlah kluster yang ingin diidentifikasi, jadi pertanyaan lain yang harus Anda jawab adalah apakah pilihan Anda untuk jumlah kluster sudah tepat. Bagaimana kita keluar dari situasi yang tampaknya membingungkan ini?

Pertama-tama, beberapa alat matematika dapat memberi tahu ilmuwan data apakah pengelompokan telah dilakukan dengan baik. Sayangnya, memiliki algoritma yang berkinerja baik dari sudut pandang matematika belum tentu merupakan tanda bahwa algoritma tersebut berguna untuk tujuan bisnis. Jika Anda membeli buku ini, kami berasumsi bahwa Anda tidak

menggunakan ML untuk menerbitkan makalah ilmiah, melainkan untuk membantu organisasi Anda. Jika ini benar, trik matematika tidak akan menarik bagi Anda. Yang seharusnya Anda lakukan adalah melihat hasil Anda dan bertanya pada diri sendiri pertanyaan-pertanyaan ini:

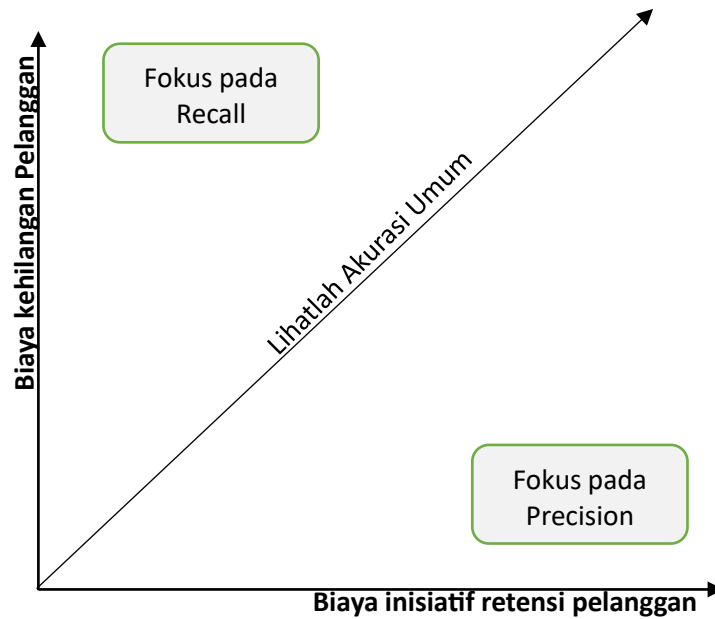
1. Apakah hasilnya dapat diinterpretasikan? Dengan kata lain, apakah pusat kluster dapat diinterpretasikan sebagai persona pembeli yang masuk akal? Jika jawabannya ya, lanjutkan ke pertanyaan 2.
2. Apakah hasilnya dapat ditindaklanjuti? Dengan kata lain, apakah kluster saya cukup berbeda sehingga saya dapat menemukan strategi yang berbeda untuk menargetkan pelanggan dari pusat yang berbeda?

Jika jawaban atas pertanyaan-pertanyaan ini adalah ya, selamat: Anda dapat mulai menguji hasilnya di dunia nyata, mengumpulkan data, dan melanjutkan dengan mengulangi dan menyempurnakan algoritma Anda ketika Anda memiliki lebih banyak data, atau dengan menggunakan pengetahuan baru untuk mendesain ulang pendekatan Anda. Untungnya, algoritma tanpa pengawasan yang berkinerja buruk biasanya lebih rendah risikonya dibandingkan algoritma pembelajaran terawasi yang berkinerja buruk, karena konsep prediksi "benar" atau "salah" lebih kabur. Karena alasan ini, Anda tidak perlu terlalu khawatir tentang metrik, melainkan tentang metodologi pengujian yang harus Anda miliki untuk mengevaluasi dampak bisnis dari proyek Anda.

3.6 MENGHUBUNGKAN METRIK ML DENGAN HASIL DAN RISIKO BISNIS

Setelah Anda terbiasa dengan metrik ML umum, mari kita lihat apa implikasinya dalam skenario bisnis. Mari kita asumsikan Anda telah menerapkan prediktor churn yang hebat yang telah mengidentifikasi sekelompok pelanggan yang kemungkinan akan meninggalkan layanan Anda, dan Anda ingin menghubungi mereka melalui panggilan telepon yang dipersonalisasi.

Jika tim ilmu data Anda telah membangun model dengan presisi tinggi tetapi daya ingat rendah, Anda tidak akan membuang banyak panggilan telepon: setiap kali Anda menelepon, Anda akan berbicara dengan pengguna yang benar-benar mempertimbangkan untuk meninggalkan Anda. Di sisi lain, beberapa pengguna meninggalkan layanan, tetapi algoritma Anda gagal mendeteksi mereka. Algoritma dengan daya ingat tinggi dan presisi rendah justru akan membuat Anda melakukan banyak panggilan telepon sehingga Anda akan menghubungi sebagian besar (bahkan semua) pelanggan yang berencana meninggalkan layanan Anda, tetapi Anda akan membuang-buang panggilan telepon ke pengguna lain yang sama sekali tidak berencana berhenti berlangganan.



Gambar 3.9 Lembar contekan untuk memutuskan apakah akan berfokus pada presisi tinggi, daya ingat tinggi, atau hanya akurasi umum berdasarkan metrik bisnis utama.

Kompromi antara presisi dan daya ingat jelas merupakan fungsi dari bisnis. Jika setiap pelanggan memiliki nilai tinggi dan biaya menghubunginya murah, pilihlah daya ingat yang tinggi. Jika pelanggan Anda tidak melakukan pembelian mahal, tidak ingin diganggu secara tidak perlu, dan biaya menelepon mereka mahal, pilihlah daya ingat yang tinggi. Strategi yang lebih canggih lagi adalah dengan mengalokasikan tindakan yang lebih mahal seperti panggilan telepon untuk pelanggan dengan nilai lebih tinggi atau kemungkinan churn yang lebih tinggi, dan menggunakan tindakan yang lebih murah seperti email untuk pelanggan lainnya. Seperti yang Anda lihat pada gambar 3.9, Anda dapat memutuskan apakah akan berfokus pada daya ingat atau presisi berdasarkan dua parameter: biaya kehilangan pelanggan, dan biaya tindakan yang Anda lakukan untuk mempertahankan mereka.

Masalahnya menjadi lebih serius dalam aplikasi yang kritis terhadap keselamatan. Jika kita memiliki pengklasifikasi yang digunakan untuk diagnosis medis, biaya positif palsu dan negatif palsu sangat berbeda. Misalkan algoritma kita mendeteksi kanker paru-paru. Pada tabel 3.6, Anda dapat melihat arti setiap kesalahan dan implikasinya.

Tabel 3.6 Implikasi dari positif benar, negatif benar, positif palsu, dan negatif palsu

Metrik	Apa maksudnya	Implikasi
Benar-benar positif	Algoritma tersebut memprediksi bahwa pasien menderita kanker, dan ternyata benar: pasien tersebut benar-benar sakit.	Pasien segera mendapatkan perawatan yang dibutuhkan.
Benar-benar negatif	Algoritma memperkirakan pasien itu sehat, dan memang benar pasien itu sehat.	Pasien pulang dengan perasaan lega.
Positif palsu	Algoritma tersebut meramalkan bahwa pasien menderita kanker (maka hasilnya positif), tetapi pasien sebenarnya sehat (maka hasilnya salah).	Pasien menjadi takut dan mungkin memulai percobaan, tetapi sebenarnya baik-baik saja.

Negatif palsu	Algoritma memperkirakan pasien sehat (negatif), tetapi pasien malah sakit (salah).	Pasien pulang dengan perasaan lega tetapi sebenarnya sakit dan tidak menerima perawatan yang dibutuhkan.
---------------	--	--

Seperti yang Anda lihat, biaya yang terkait dengan kesalahan sangat berbeda, tergantung pada apakah kita salah mengklasifikasikan pasien sehat atau sakit. Hasil positif palsu dapat menyiratkan pemeriksaan tambahan, atau mungkin perawatan yang tidak perlu, tetapi hasil negatif palsu akan menghilangkan terapi yang sangat dibutuhkan pasien untuk bertahan hidup.

Dalam skenario seperti itu, Anda jelas melihat betapa berbahayanya mengoptimalkan metrik yang salah. Memiliki akurasi yang tinggi di sini tidak akan menunjukkan algoritma yang baik, karena akan memberi bobot yang sama pada hasil negatif palsu dan positif palsu. Yang kita inginkan dalam kasus ini adalah daya ingat yang tinggi: kita ingin jumlah pasien sakit terbanyak teridentifikasi, meskipun itu berarti terdapat beberapa hasil positif palsu.

Hasil positif palsu dapat menyebabkan pemeriksaan tambahan yang tidak perlu dan membuat beberapa keluarga takut, tetapi hal itu akan memastikan jumlah orang sakit terbanyak yang terdeteksi dan dirawat. Seperti yang Anda lihat, setiap algoritma memiliki indikator kinerjanya, dan Anda perlu menilai metrik mana yang paling penting bagi Anda, agar tim ilmu data Anda dapat memaksimalkannya. Kita telah melihat bahwa pilihan terbaik tidak selalu jelas. Dari sudut pandang bisnis, ide paling ampuh adalah mengaitkan setiap kesalahan klasifikasi dengan nilai tertentu.

Jika asumsi Anda benar, hal ini secara otomatis akan menghasilkan hasil bisnis terbaik. Perusahaan telekomunikasi imajiner kita dapat mengembangkan rencana retensi untuk pelanggan yang tidak puas yang memberi mereka bonus Rp. 1.000.000 jika mereka tetap menggunakan perusahaan selama 12 bulan ke depan, sehingga mempertahankan pelanggan yang mungkin akan beralih ke pesaing dalam waktu dekat.

Di sisi lain, mari kita asumsikan bahwa setiap pelanggan yang hilang membuat perusahaan kehilangan keuntungan sebesar Rp. 5.000.000. Setelah kita memiliki semua angka ini, kita dapat dengan mudah menghitung berapa banyak kerugian yang ditimbulkan oleh negatif palsu dan positif palsu. Setiap negatif palsu (pelanggan yang beralih ke pesaing sebelum kami dapat membujuk mereka dengan diskon) mengakibatkan kerugian pendapatan sebesar Rp. 5.000.000. Setiap positif palsu (pelanggan setia yang tidak akan berhenti berlangganan tetapi tetap menerima hadiah) mengakibatkan kerugian pendapatan sebesar Rp. 1.000.000. Sekarang, kita dapat menggunakan beberapa perhitungan dasar untuk menghubungkan kinerja model dengan nilai moneter:

$$\begin{aligned} \text{Total Biaya} = & (Rp. 5.000.000 \times FN) + (Rp. 0 \times TN) + (Rp. 1.000.000 \times FP) \\ & + (Rp. 1.000.000 \times TP) \end{aligned}$$

Di sini, FN adalah jumlah negatif palsu, TN adalah negatif benar, FP adalah negatif palsu, dan TP adalah positif benar.

Anda dapat menggunakan ide yang sama ini untuk situasi lain di mana Anda menggunakan model klasifikasi biner. Misalnya, katakanlah Anda sedang mengembangkan sistem kendali mutu otomatis untuk lini manufaktur. Anda dapat menyesuaikan model untuk membiarkan lebih banyak barang masuk (meningkatkan rasio negatif palsu) atau langsung menolak lebih banyak barang (meningkatkan rasio positif palsu). Strategi pertama mungkin akan membiarkan lebih banyak barang cacat masuk, sehingga menyebabkan lebih banyak masalah di hilir proses manufaktur. Strategi kedua mungkin mengakibatkan pemborosan material yang tidak perlu, karena produk yang sangat baik ditolak. Bagaimanapun, Anda dapat menyesuaikan rumus sebelumnya dengan situasi bisnis spesifik Anda.

Apa pun strategi Anda, Anda tetap harus merasa nyaman dengan kenyataan bahwa tidak ada algoritma pembelajaran mesin yang sempurna (jika sempurna, pasti ada yang salah). Anda akan membuat kesalahan: pastikan kesalahan ini dapat diterima dan Anda serta algoritma Anda belajar darinya.

3.7 STUDI KASUS

Bagian ini menyajikan dua studi kasus dari perusahaan nyata yang menggunakan AI untuk upaya pemasaran mereka. Yang pertama adalah Opower, sebuah perusahaan energi yang menggunakan pengelompokan untuk mengelompokkan pengguna berdasarkan kebiasaan konsumsi energi mereka. Yang kedua adalah inisiatif Target: toko ritel tersebut menggunakan data pelanggannya untuk mengidentifikasi tanda-tanda kehamilan sejak dini dan mulai mengiklankan popok.

AI untuk Menyempurnakan Penargetan dan Pemosisian: Opower

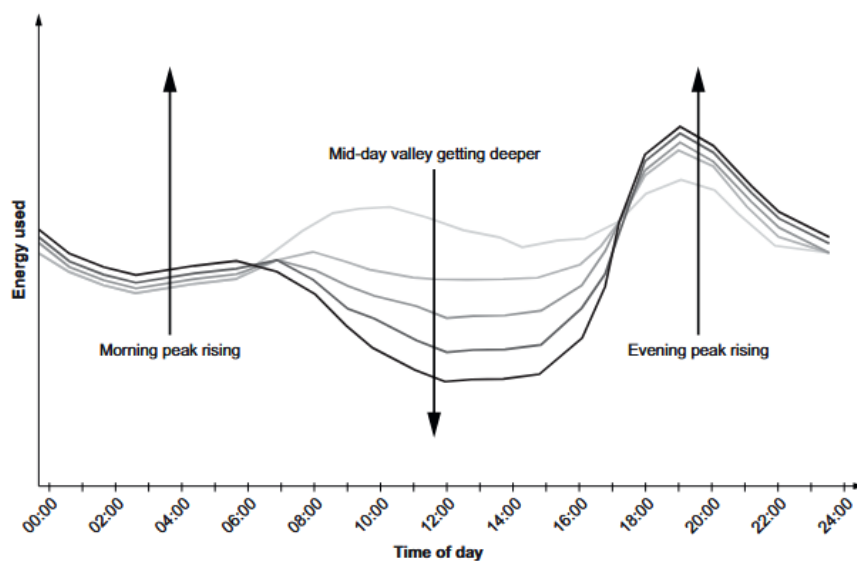
Hubungan antara perusahaan energi dan pelanggan cukup unik: meskipun perusahaan konsumen biasanya dibayar saat pelanggan melakukan pembelian, dalam hal energi, terdapat jeda waktu: orang-orang pertama-tama mengonsumsi energi untuk jangka waktu tertentu dan kemudian ditagih. Selain itu, orang tidak membeli energi hanya demi energi itu sendiri: energi merupakan pengeluaran yang diperlukan untuk menikmati barang-barang lain (seperti menyalakan TV).

Kedua faktor ini menyulitkan setiap perusahaan utilitas untuk mengenal dan berinteraksi dengan pelanggannya. Hal ini bukan hanya menjadi masalah bagi departemen pemasaran perusahaan energi, tetapi juga bagi keseluruhan biaya dan profitabilitas mereka. Faktanya, memproduksi energi tidak selalu membebani perusahaan utilitas dengan harga yang sama: hal ini bergantung pada jenis pembangkit listrik yang mereka gunakan untuk memproduksinya, sebuah pilihan yang bergantung pada seberapa banyak energi yang perlu mereka produksi dan seberapa jauh mereka merencanakannya.

Skenario terbaiknya adalah permintaan yang benar-benar datar: utilitas dapat mulai memproduksi energi dalam jumlah X menggunakan pembangkit listrik terbaik mereka dan tidak pernah berubah. Tentu saja, ini tidak pernah terjadi, karena permintaan harian justru berbentuk bebek, dengan dua puncak yang biasanya terjadi saat orang bangun tidur dan saat mereka pulang kerja. Untuk memenuhi permintaan selama puncak ini, utilitas perlu menggunakan pembangkit listrik yang mudah dinyalakan dan dimatikan serta dimodulasi,

tetapi fleksibilitas ini ada harganya: pembangkit ini lebih mahal untuk dioperasikan, mengonsumsi lebih banyak energi, dan membakar bahan bakar yang lebih mahal.

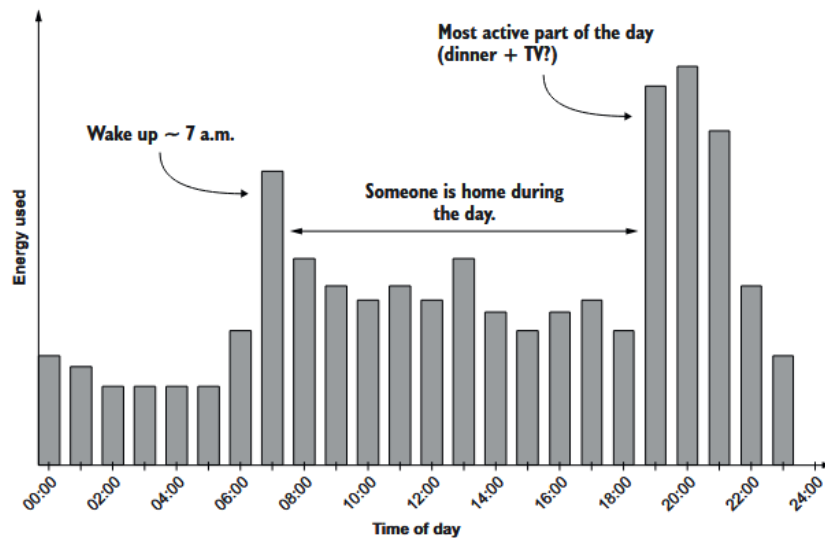
Seperti yang ditunjukkan pada gambar 3.10, masalah ini semakin parah: kebiasaan saat ini membuat kurva permintaan energi kurang datar, dengan lembah yang lebih besar di siang hari dan puncak yang lebih tajam di pagi dan sore hari. Impian setiap perusahaan utilitas adalah dapat membentuk kebiasaan pelanggan mereka dengan memindahkan puncak mereka ke waktu yang berbeda dalam sehari, sehingga mencapai permintaan yang hampir datar secara keseluruhan.



Gambar 3.10 Kurva beban tipikal untuk konsumsi energi rumah tangga di California

Strategi Opower

Opower didirikan pada tahun 2007 dengan misi menyediakan perangkat bagi perusahaan energi untuk menargetkan pengguna dan mengajak mereka mengubah kebiasaan konsumsi, baik untuk menghemat energi maupun mencapai permintaan yang lebih stabil secara keseluruhan. Perusahaan ini mendapatkan data meter pintar dari pelanggan utilitas, dan dapat mengirimkan laporan berisi detail konsumsi energi mereka serta kiat-kiat untuk meningkatkan kebiasaan mereka, melalui surat, email, portal internet, pesan teks, dan terkadang tampilan energi di rumah.

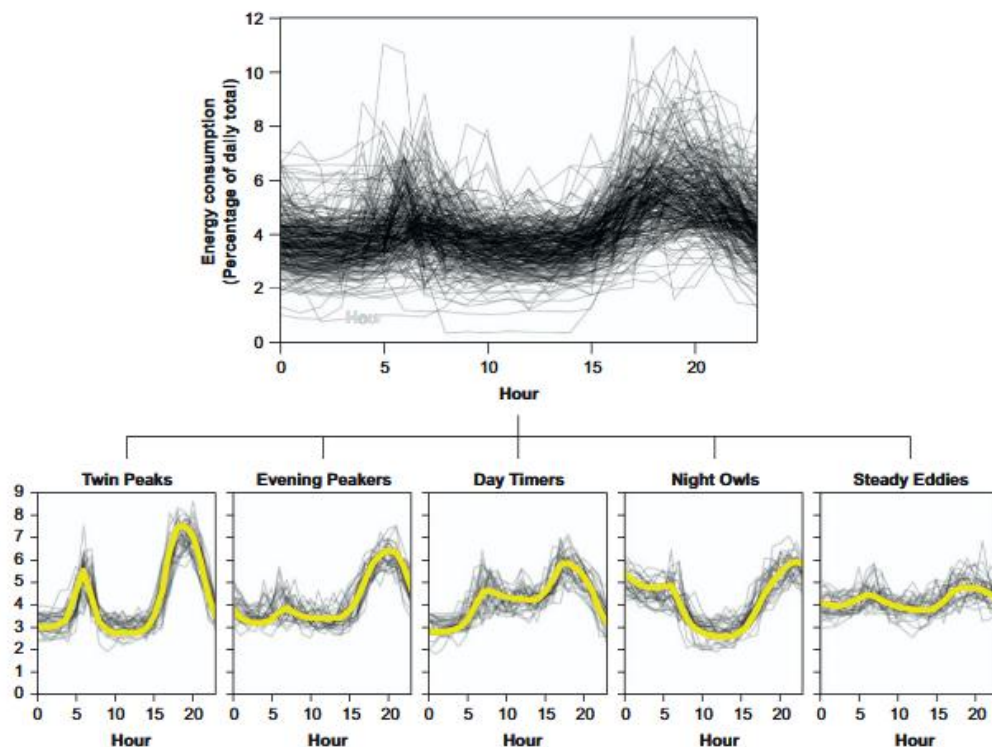


Gambar 3.11 Nilai konsumsi energi rumah tangga pada waktu yang berbeda dalam sehari

Rahasia perusahaan ini adalah perpaduan antara ilmu perilaku dan ilmu data. Opower beralih ke salah satu aplikasi inti pembelajaran mesin pembelajaran tanpa pengawasan yang digunakannya untuk mengidentifikasi kelompok pengguna tertentu yang dapat ditargetkan secara serupa.

Elemen paling khas dari pengguna energi adalah pola konsumsi mereka: berapa banyak energi yang mereka gunakan pada jam-jam yang berbeda dalam sehari. Oleh karena itu, kita dapat melihat setiap pengguna seperti pada Gambar 3.11: kombinasi 24 angka yang mewakili jumlah energi yang dikonsumsi untuk setiap jam dalam sehari. Angka-angka ini menggambarkan gambaran akurat tentang kebiasaan pengguna, seperti halnya jumlah pembelian, pengeluaran rata-rata, dan barang yang dibeli dalam skenario e-commerce.

Menemukan berbagai kelompok pengguna berdasarkan pola konsumsi mereka bukanlah tugas yang mudah. Jika kita merepresentasikan 1.000 pengguna dengan cara ini dan memplot semuanya pada grafik, kita mendapatkan apa yang para ilmuwan data sebut sebagai bola rambut: campuran garis yang rumit, satu di atas yang lain, yang mustahil dibedakan oleh mata manusia. Di sisi lain, algoritma tanpa pengawasan paling berguna dalam mengidentifikasi pola; dengan menggunakan algoritma ini, Opower berhasil menemukan lima kluster pengguna. Kami mengumpulkan data dari internet dan menjalankan algoritma tanpa pengawasan untuk meniru hasil yang dicapai Opower; Anda dapat melihat bola rambut dan kluster terakhir pada Gambar 3.12.



Gambar 3.12 Mengubah "hairball" berisi 1.000 pengguna menjadi lima kluster

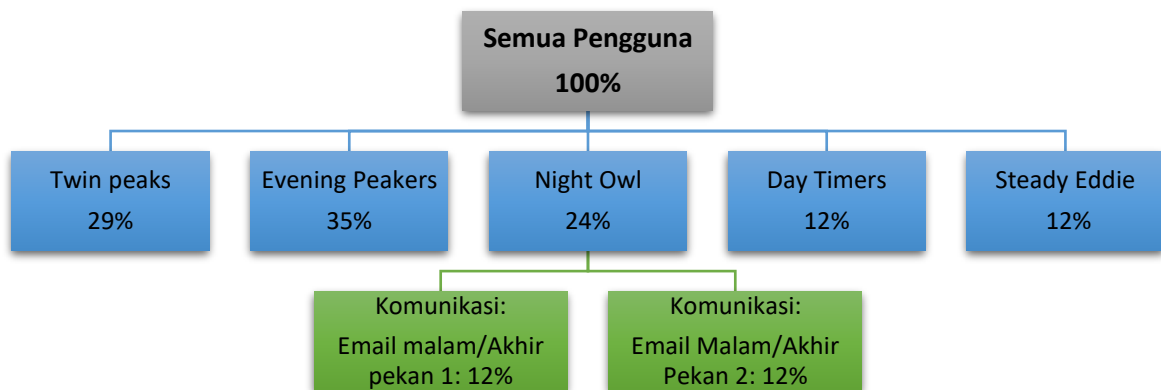
Melihat kluster-kluster ini, kita dapat mencoba menafsirkan masing-masing kluster:

- *Twin Peaks* — Mereka mengonsumsi banyak energi di pagi hari, jauh lebih banyak daripada pengguna dari kluster lain. Mereka juga mengalami puncak energi di malam hari sekitar waktu makan malam, tetapi sedikit lebih rendah daripada puncak energi di pagi hari.
- *Evening Peakers* — Mereka memiliki puncak energi yang tinggi di malam hari yang dicapai dengan mulus, dimulai dari pagi hari. Puncak energi pertama di pagi hari tidak ada.
- *Day Timers* — Mereka tidak memiliki puncak energi klasik seperti yang dimiliki semua kluster lain. Sebaliknya, mereka memiliki konsumsi energi yang tinggi di siang hari yang meningkat dan menurun secara mulus.
- *Steady Eddies* — Mereka adalah yang paling stabil; dua puncak energi klasiknya sangat mulus.
- *Night Owls* — Mereka memiliki puncak energi klasik di pagi hari dan sore hari, tetapi cenderung mengonsumsi energi di malam hari.

Opower menggunakan hasil pengelompokan untuk menargetkan pengguna yang berbeda dengan berbagai cara. Infrastrukturnya memungkinkan pembagian setiap kluster ke dalam kelompok-kelompok dan pengujian pesan yang berbeda. Dari analisis hasil dan penggunaan ilmu perilaku, Opower mampu terus meningkatkan penargetannya. Contoh segmentasi ini direpresentasikan pada Gambar 3.13: Opower dapat memilih kluster tertentu (dalam hal ini Night Owls), dan membaginya menjadi dua kelompok yang ditargetkan dengan email yang

berbeda. Pendekatan ini memungkinkan Opower untuk mengoptimalkan pesannya dan menargetkan orang-orang dengan cara terbaik.

Berkat pendekatannya, Opower mampu membantu konsumen menghemat 10% tagihan energi mereka, sekaligus membuat utilitas lebih menguntungkan, berkat distribusi total permintaan yang lebih baik. Sebagai hasil dari kinerjanya, Opower diakuisisi oleh Oracle pada tahun 2016 seharga Rp. 5,32 Triliun.



Gambar 3.13 Pohon keputusan yang digunakan Opower untuk menargetkan penggunanya

PERTANYAAN KASUS

Bagaimana Anda bisa membujuk sekelompok orang untuk mengubah perilaku mereka?

DISKUSI KASUS

Pikirkan sejenak tentang kompleksitas misi Opower: misi ini harus memungkinkan jutaan orang untuk mengubah cara mereka mengonsumsi energi dengan mengirimkan semacam laporan. Anda akan naif jika berpikir pesan yang sama dapat berhasil untuk semua orang.

Setiap orang merespons insentif dan pesan yang berbeda, jadi idealnya Anda harus menargetkan setiap orang dengan pesan yang berbeda. Hal ini, tentu saja, tidak praktis atau terukur, tetapi Anda masih dapat mencapai hasil yang signifikan dengan membagi audiens Anda ke dalam kelompok-kelompok yang memiliki kesamaan. Bagi Opower, langkah pertama ini dilakukan menggunakan pembelajaran tanpa pengawasan. Ingatlah bahwa pembelajaran tanpa pengawasan adalah kelas algoritma pembelajaran mesin yang mencari pola dalam data, menemukan kelompok titik data yang serupa berdasarkan fiturnya.

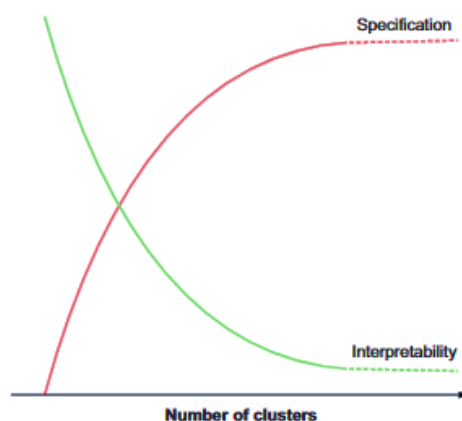
Dalam kasus Opower, karakteristik terpenting adalah kebiasaan konsumsi energi pengguna, sehingga Opower membutuhkan cara untuk memperhitungkan hal ini. Ide Opower untuk mengekspresikan kebiasaan konsumsi energi ke dalam fitur (sekelompok angka) adalah dengan mempertimbangkan jumlah energi yang dikonsumsi oleh pengguna tertentu setiap jam dalam sehari. Hasilnya adalah representasi setiap orang yang terdiri dari 24 angka, yang cukup mudah dipahami manusia, tetapi sangat mudah dipahami komputer: setelah algoritma

pembelajaran tanpa pengawasan diberi data ini, ia dapat dengan mudah membandingkan dua pengguna dan menemukan kesamaan.

Pembelajaran tanpa pengawasan memberi Opower perbedaan pertama dan terpenting di antara pengguna: konsumsi energi mereka. Langkah kedua adalah mengukur efek dari berbagai pesan, yang dilakukan Opower dengan menggunakan ilmu perilaku dan pengujian A/B (mengirim pesan yang berbeda kepada orang yang berbeda dan mengukur hasilnya).

Catatan implementasi terakhir tentang jumlah kluster yang diidentifikasi oleh algoritma di sini. Opower memiliki lima, tetapi dari mana angka ini berasal? Dengan beberapa algoritma yang banyak digunakan, Anda perlu menentukan jumlah kluster yang harus dicari oleh algoritma, dan algoritma akan mengoptimalkannya. Alat matematika memberikan petunjuk kepada ilmuwan data tentang angka berapa yang akan menghasilkan hasil terbaik dari sudut pandang matematika, tetapi angka tersebut belum tentu merupakan solusi terbaik untuk kebutuhan bisnis Anda, jadi tidak ada solusi yang "benar" atau "salah". Karena tidak ada skrip untuk menentukan jumlah kluster, menentukan angka ini terkadang terasa seperti berbelanja bahan makanan: "Apakah kita ingin lima kluster? Kita buat enam saja; lebih baik punya sedikit lebih banyak, untuk berjaga-jaga jika kita kehabisan." Kita bisa berargumen bahwa membuat keputusan sebenarnya jauh lebih sederhana.

Bayangkan ini: setiap kali Anda meminta algoritma untuk menemukan satu kluster lagi, ia akan melihat pola baru yang lebih halus dalam data. Dengan asumsi kita memiliki 1.000 pengguna, batas atasnya adalah menemukan 1.000 kluster, di mana setiap pengguna memiliki klasternya sendiri. Sebelum mencapai situasi ekstrem tersebut, terkadang Anda akan menyadari bahwa Anda tidak dapat benar-benar memperoleh pengetahuan baru dari menambahkan kluster baru. Inilah saat yang tepat untuk berhenti. Menambahkan kluster akan membuat hasil algoritma lebih spesifik, tetapi juga kurang dapat diinterpretasikan. Gambar 3.14 mengilustrasikan kompromi ini.



Gambar 3.14 Hubungan antara jumlah kluster dan interpretabilitas serta spesifikasi hasil

Proses idealnya adalah memulai dengan jumlah kluster yang relatif kecil; idealnya, jumlah ini dapat disediakan oleh ilmuwan data Anda, yang akan menggunakan perangkat matematika

untuk mencari tahu apa yang masuk akal dari sudut pandang teknis. Dari sana, Anda dapat bergerak ke bawah, dengan memadatkan berbagai klaster, atau ke atas, dengan meningkatkan granularitas pengelompokan. Setiap kali Anda menambah atau mengurangi jumlah klaster, cobalah menafsirkan hasilnya dan tanyakan pada diri sendiri apakah jumlah klaster yang lebih tinggi membantu Anda lebih memahami fenomena tersebut. Jika jawabannya tidak, kami dapat berargumen bahwa Anda telah menemukan jumlah klaster yang optimal. Ini mungkin bukan jumlah optimal teknis yang ditemukan oleh para ilmuwan data, tetapi akan menjadi jumlah yang optimal untuk bisnis.

AI untuk Mengantisipasi Kebutuhan Pelanggan: Target

Kelahiran seorang anak tidak hanya bisa menjadi peristiwa yang membahagiakan bagi pasangan, tetapi juga awal dari aliran biaya baru. Tiba-tiba, orang tua perlu memasukkan produk baru yang mahal ke dalam belanja rutin mereka, seperti popok, makanan bayi, kereta dorong bayi, dan sebagainya.

Peristiwa semacam ini sangat menguntungkan bagi para peritel, yang berusaha menarik orang tua baru ke toko mereka untuk menjadi rujukan terpercaya dalam membeli semua kebutuhan bayi baru mereka. Strategi yang biasanya digunakan peritel adalah menawarkan diskon dan promosi kepada keluarga yang baru terbentuk, dan semakin cepat peritel memberikan kuponnya, semakin besar keuntungan yang diperolehnya atas pesaing dalam mendapatkan pelanggan baru. Di antara semua pengeluaran baru yang harus ditanggung keluarga, popok merupakan salah satu yang sangat penting.

Menurut laporan Nielsen tahun 2016, keluarga Amerika menghabiskan rata-rata lebih dari Rp. 10.000.000 per tahun per bayi baru lahir untuk popok. Pendapatan bukan satu-satunya alasan popok penting bagi peritel, karena popok juga memainkan peran strategis sebagai penggerak arus barang: produk yang tak terelakkan bagi sebuah keluarga sehingga menarik ibu dan ayah muda ke toko, yang kemudian pada akhirnya mereka juga menghabiskan uang untuk barang-barang lain. Nilai ekonomi dan strategis dari akuisisi pelanggan yang baru saja melahirkan bayi menghasilkan "perang kupon" antar peritel: semuanya bersaing untuk menjadi yang pertama memberikan penawaran kepada ibu dan ayah baru, dan menarik mereka untuk membeli popok dan produk lainnya di toko mereka.

Pendekatan pertama yang dicoba peritel untuk memenangkan pelanggan baru ini adalah dengan mendapatkan data kantor catatan sipil yang tersedia secara terbuka untuk bayi baru lahir dan mengirimkan penawaran yang ditargetkan kepada para ayah dan ibu baru, memberi insentif kepada mereka untuk membeli popok pertama mereka di toko mereka. Namun, banyak peritel segera mulai menggunakan strategi ini dan mulai bersaing untuk mendapatkan perhatian dari keluarga bayi baru lahir tepat setelah bayi tersebut lahir.

Untuk mengungguli para pesaingnya, Target, peritel terbesar kedelapan di Amerika Serikat, memutuskan untuk menyelidiki strategi baru yang berani: memberikan penawaran kepada keluarga sebelum bayi lahir. Jika jaringan ritel tersebut dapat mengetahui bahwa ada pasangan yang sedang menantikan kelahiran bayi, mereka dapat mengungguli para pesaingnya dan menjangkau keluarga tersebut terlebih dahulu. Raksasa ritel tersebut menggunakan data kartu loyalitasnya untuk mengidentifikasi pola kebiasaan belanja ibu hamil,

dan menyiapkan serangkaian kupon dan promosi khusus yang akan menarik kategori pembeli tersebut untuk memperluas pilihan barang yang mereka pilih untuk dibeli di Target. Dalam artikel Forbes tahun 2012 (<http://mng.bz/PO6g>), kita mendapatkan beberapa indikasi hasil yang ditemukan oleh ahli statistik Target, Andrew Pole:

Saat komputer Pole menelusuri data, ia dapat mengidentifikasi sekitar 25 produk yang, ketika dianalisis bersama, memungkinkannya untuk memberikan skor "prediksi kehamilan" kepada setiap pembeli. Lebih penting lagi, ia juga dapat memperkirakan tanggal persalinannya dalam rentang waktu yang singkat, sehingga Target dapat mengirimkan kupon yang disesuaikan dengan tahap kehamilan yang sangat spesifik.

Target dilaporkan mampu meningkatkan penjualan berkat strategi barunya. Namun, peningkatan pendapatan ini harus dibayar dengan harga mahal: orang-orang mulai merasa dimata-matai dan tidak senang dengan hal itu. Tanda peringatan pertama bahwa masyarakat umum mungkin tidak menyukai teknik penargetan intensif semacam itu muncul ketika seorang ayah yang kesal mengeluh tentang kupon produk bayi yang diterima putrinya yang masih SMA dari Target. Seperti yang dilaporkan dalam artikel Forbes yang sama, ia mengeluh kepada manajer toko sambil berteriak:

Putri saya menerima ini melalui pos! Dia masih SMA, dan Anda mengirimkan kupon untuk pakaian bayi dan tempat tidur bayi? Apakah Anda mencoba mendorongnya untuk hamil?

Ternyata, gadis itu benar-benar hamil, dan Target telah mengetahuinya sebelum ayahnya. Manajer Target dilaporkan menelepon ayah yang marah itu beberapa saat kemudian, dan mendapat tanggapan seperti yang dilaporkan oleh Forbes:

Saya sudah bicara dengan putri saya. Ternyata ada beberapa kegiatan di rumah saya yang tidak sepenuhnya saya sadari. Dia akan melahirkan bulan Agustus. Saya berutang permintaan maaf kepada Anda.

Sebagaimana dilaporkan dalam artikel Forbes, Target akhirnya mencampur rekomendasi personal yang diperoleh melalui pembelajaran mesin (ML) dengan barang-barang yang tampak acak dari katalog (misalnya, menambahkan kupon untuk mesin pemotong rumput tepat di samping kupon untuk popok), untuk menghindari membuat pelanggan merasa seperti sedang dibuntuti.

PERTANYAAN KASUS

Apakah diskon khusus merupakan cara yang efektif untuk memenangkan pelanggan dibandingkan pesaing?

DISKUSI KASUS

Penawaran khusus dan kupon merupakan strategi yang sangat umum digunakan peritel untuk menarik pelanggan. Namun, ketika semua orang

tahu seseorang sedang mencari produk tertentu yang dapat dijual oleh semua orang, dampaknya biasanya berupa perang promosi, yang menggerogoti margin keuntungan peritel. Hal ini mungkin baik bagi pelanggan, tetapi jelas tidak bagi peritel. Kami berpendapat bahwa menawarkan diskon kepada pelanggan dapat menjadi ide pemasaran yang valid jika hal-hal berikut ini benar:

1. Peritel tahu kepada siapa kupon harus dikirim (ini efektif dalam menargetkan penggunaanya).
2. Peritel lain tidak menargetkan orang tersebut pada saat yang sama, untuk menghindari perang harga yang telah disebutkan sebelumnya. Strategi awal yang diadopsi para peritel cukup cerdas: menggunakan data kantor catatan sipil yang tersedia secara terbuka untuk bayi baru lahir guna menyasar keluarga segera setelah bayi lahir. Strategi ini efektif pada poin 1 (menargetkan pelanggan yang tepat), tetapi gagal pada poin 2: seperti namanya, data terbuka bersifat terbuka, dan setiap peritel dapat menggunakannya.

Kasus Target adalah contoh sempurna tentang bagaimana Anda dapat menggunakan data dan pembelajaran mesin untuk memenangkan persaingan, menargetkan pelanggan yang tepat sebelum orang lain. Target memulai dari data penjualan bisnis intinya: kebiasaan belanja pelanggan, yang dikumpulkan melalui kartu loyalitas. Target juga mampu membangun label "sedang menantikan kelahiran" / "tidak menanti kelahiran" untuk pelanggan sebelumnya, dengan melihat apakah pelanggan ini benar-benar mulai membeli popok di masa mendatang. Kami harap Anda sudah memahami bahwa ini tampak seperti masalah klasifikasi pembelajaran terawasi klasik:

- Fitur-fiturnya adalah kebiasaan belanja pelanggan (biarkan ilmuwan data mencari tahu cara membangunnya secara tepat).
- Labelnya adalah apakah pelanggan tersebut sedang menantikan kelahiran, berdasarkan fitur pembelian barang-barang perlengkapan bayi.
- Setelah model dibangun berdasarkan data masa lalu, model tersebut dapat diterapkan pada pelanggan saat ini, mengidentifikasi mereka yang sedang menantikan kelahiran bayi.

Perhatikan bahwa label tersebut tidak langsung tersedia: Pelanggan Target tidak menghubungi penjual untuk memberi tahu bahwa bayi telah lahir. Namun, label ini dapat diberikan kepada pelanggan sebelumnya dengan melihat perilaku pembelian mereka di masa lalu: jika seorang pelanggan mulai membeli popok atau kereta dorong bayi enam bulan yang lalu, ia pasti hamil dalam sembilan bulan sebelumnya dan dapat diberi label sebagai contoh positif.

Bayangkan betapa ampuhnya pendekatan ini. Terdapat pola dalam kebiasaan pembelian ibu hamil: mereka mungkin berhenti membeli makanan tertentu seperti sosis atau sushi, makan makanan yang lebih sehat, atau berhenti merokok. Masalahnya adalah ketika

tidak ada indikator yang jelas (misalnya, pembelian vitamin prenatal atau popok), sulit bagi manusia untuk menemukan pola dengan melihat pembelian pelanggan.

Namun, hal ini tidak sulit bagi algoritma pembelajaran mesin (ML), yang dapat mengetahui kebiasaan pembelian mana yang menunjukkan kehamilan. Hal lain yang perlu Anda ingat adalah bahwa apa yang mungkin terdengar seperti cerita yang bagus, dan langkah hebat dari Target untuk meningkatkan penjualan, juga bisa menjadi bom waktu PR yang terus berdetak karena penargetan yang agresif di domain sensitif tersebut.

Singkatnya, Anda harus mengingat hal-hal berikut dari kasus ini:

- Pembelajaran mesin (ML) dapat membantu Anda memprediksi pembelian di masa mendatang, mengantisipasi pesaing.
- Data bisnis inti memiliki nilai yang jauh lebih tinggi daripada data yang tersedia secara terbuka.
- Waspada faktor yang menyeramkan: ML bisa begitu kuat sehingga orang mungkin merasa seperti Anda memata-matai mereka.

Ringkasan

- Alasan inti di balik efektivitas AI dalam pemasaran dan penjualan adalah kemampuan untuk menargetkan pelanggan tertentu dalam skala besar dan membuat keputusan berdasarkan data.
- Model prediksi churn adalah aplikasi umum AI dalam pemasaran, dan dapat mengidentifikasi pelanggan mana yang paling mungkin berhenti menggunakan layanan Anda.
- Demikian pula, pembelajaran mesin (ML) dapat membantu pemasar menemukan peluang upselling dengan mengidentifikasi pelanggan mana yang lebih cenderung membeli lebih banyak.
- Pembelajaran tanpa pengawasan adalah bagian dari pembelajaran mesin yang memungkinkan komputer menemukan struktur dalam data, menemukan kluster titik data yang serupa. Pengelompokan juga memiliki kegunaan dalam pemasaran.
- Terdapat berbagai cara dan metrik untuk mengevaluasi kinerja algoritma klasifikasi (kami membahas akurasi, perolehan kembali, dan presisi). Masing-masing metrik memberi tahu kita sesuatu yang berbeda tentang algoritma tersebut, dan Anda dapat melihat tujuan bisnis Anda untuk memilih yang paling tepat.
- Opower menggunakan pembelajaran tanpa pengawasan untuk membagi konsumen energi ke dalam kluster dan kemudian menargetkan mereka secara individual untuk mengoptimalkan pola konsumsi energi mereka.
- Target menggunakan pembelajaran terawasi untuk mengidentifikasi pelanggan mana yang akan memiliki bayi, lalu menawarkan kupon untuk popok dan produk bayi.

BAB 4

ARTIFICIAL INTELLIGENCE UNTUK MEDIA

Bab ini membahas

- Menggunakan AI untuk memahami konten gambar
- Menggunakan pembelajaran mendalam untuk memecahkan tantangan dalam visi komputer
- Mendeteksi dan memberi label pada wajah manusia
- Menciptakan karya seni orisinal dengan pembelajaran mendalam
- Menggunakan AI dengan data audio
- Studi kasus: mengotomatiskan pemilahan sayuran dengan robot seharga Rp. 10.000.000

Gambar dan video mungkin merupakan sumber data terkaya yang digunakan manusia untuk memahami dunia. Sayangnya, keduanya sebagian besar terlarang bagi komputer hingga tahun 2010an. Meskipun kamera telah semakin baik selama beberapa dekade, komputer masih belum dapat memahami konten gambar. Momen penting adalah pengembangan bidang baru pembelajaran mesin yang disebut pembelajaran mendalam (DL).

Keluarga algoritma ini memiliki daya pembelajaran yang tak tertandingi (untuk saat ini) yang dapat mencapai akurasi super dalam tugas-tugas penting seperti mendeteksi objek dan wajah dalam gambar. Kemampuan ini memungkinkan bisnis untuk meningkatkan produk mereka yang sudah ada dengan kemampuan yang akan dianggap ajaib hanya 10 tahun yang lalu, mulai dari membuat album foto secara otomatis berdasarkan konten gambar, hingga mengotomatiskan proses industri dengan robot.

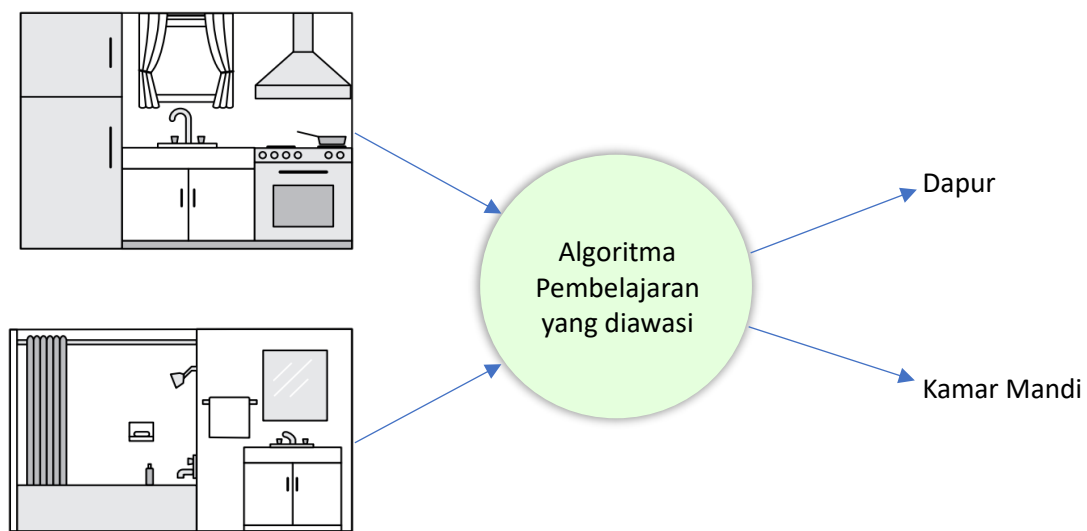
4.1 MENINGKATKAN PRODUK DENGAN VISI KOMPUTER

Alat prediksi harga rumah FutureHouse yang kami kembangkan di bab 2 meraih kesuksesan besar. Pelanggan sangat puas dengan saran harga yang mereka dapatkan dari model AI dan bergegas mendaftarkan rumah mereka di situs web. Rumah yang baru terdaftar terjual dengan cepat dengan harga yang sangat mendekati saran algoritma. Bisnis berjalan baik, tetapi hal ini justru memicu ambisi Anda.

Dalam sesi curah pendapat lainnya, tim pengalaman pengguna menampilkan data analitik perilaku untuk pasar dan menyampaikan beberapa berita yang mengkhawatirkan. Sebagian besar pengguna berhenti mengikuti alur langkah demi langkah untuk mendaftarkan properti mereka. Penurunan terbesar terjadi ketika mereka diminta untuk mengunggah gambar dan secara manual memilih ruangan mana di rumah yang akan mereka gambarkan (kamar mandi, ruang tamu, dan sebagainya). Proses ini panjang dan membosankan, dan tim Anda termotivasi untuk memulai proyek baru guna menyederhanakan pengalaman pengguna: kita dapat menggunakan AI untuk memberi label pada gambar secara otomatis dan menghilangkan semua pekerjaan manual. Pengalaman pengguna akan jauh lebih mudah:

mereka cukup mengunggah beberapa gambar, yang kemudian secara otomatis ditandai sebagai kamar mandi/dapur/ruang tamu/dll. Alur keseluruhan akan menjadi jauh lebih cepat dan menyenangkan.

Yang kami rancang adalah sebuah algoritma yang memiliki input berupa gambar sebuah ruangan, dan output berupa jenisnya (kamar mandi, dapur, dan sebagainya), seperti yang ditunjukkan pada gambar 4.1. Perhatikan bahwa AI kami harus mempelajari cara memetakan input ini ke target tertentu (kategori ruangan); oleh karena itu, kita berurusan dengan masalah pembelajaran terawasi. Lebih spesifiknya, ini adalah masalah klasifikasi, karena targetnya adalah kategori, bukan angka. Untuk alasan yang jelas, sebagian besar teknisi menggunakan nama umum untuk tugas ini: klasifikasi gambar.



Gambar 4.1 Gambaran umum input dan output pengklasifikasi gambar

Setelah Anda mengetahui jenis masalah yang kita hadapi, Anda juga tahu jenis data pelatihan apa yang perlu kita kumpulkan: kumpulan gambar berlabel yang telah dikelompokkan berdasarkan ruangan tempat gambar tersebut diambil. Untungnya, kita tahu tempat yang tepat untuk mendapatkannya: pengguna telah melalui proses yang rumit mengunggah gambar ruangan dan menandainya dengan kategori. Sekarang, kita dapat mengambil semua data berlabel ini dan membangun model yang menandai gambar secara otomatis, sehingga pengguna di masa mendatang tidak perlu melakukan semua pekerjaan membosankan ini. Pola ini cukup umum ketika membangun aplikasi AI dalam bisnis: proses manual telah menghasilkan banyak data berlabel, yang dapat digunakan untuk mengembangkan otomatisasi berbasis AI.

Jika kita harus mengikuti cetak biru di bab sebelumnya, langkah selanjutnya adalah mengidentifikasi fitur-fitur yang relevan agar algoritme dapat mempelajari elemen mana yang membentuk kamar mandi dan elemen mana yang membentuk dapur. Ingat: fitur adalah aspek apa pun dari setiap item input yang Anda yakini akan berguna bagi algoritme dalam mempelajari cara mengklasifikasikan item. Tabel 4.1 menunjukkan contoh fitur yang mungkin ingin Anda gunakan.

Tabel 4.1 Pengetahuan domain seringkali merupakan alat terbaik untuk memilih fitur.

Tujuan	Fitur yang mungkin
Memprediksi harga rumah	Luas persegi, lokasi, tahun renovasi
Memutuskan apakah suatu tumor jinak atau ganas	Ukuran tumor, usia, hasil tes darah
Memutuskan apakah pelanggan akan berhenti berlangganan	Penggunaan produk, usia, pengeluaran bulanan, produk yang dibeli

Bagian pembelajaran dari pembelajaran mesin baru dimulai setelah para insinyur menentukan fitur mana yang akan digunakan untuk masalah tertentu. Manusia cukup mahir melakukan tugas ini dengan menggunakan intuisi dan pengalaman, sementara algoritma pembelajaran mesin tradisional perlu diberi tahu angka mana yang harus digunakan. Pertanyaan yang harus Anda ajukan untuk mengidentifikasi fitur yang relevan adalah, "Jika saya harus mengklasifikasikan gambar ini sendiri, apa yang akan saya cari?" Wajar jika menggunakan fitur seperti ini:

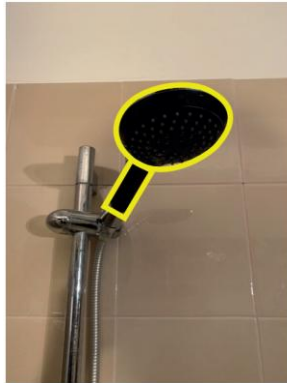
- Apakah ada toilet di gambar?
- Apakah ada kulkas di gambar?
- Apakah ada pancuran di gambar?

Tugas ini tampak begitu mudah sehingga kita bahkan tidak memerlukan pembelajaran mesin untuk menyelesaikannya: jika ada pancuran, kita tahu kita sedang melihat kamar mandi. (Kami sarankan untuk mengobrol langsung dengan arsitek Anda jika dapur Anda memiliki pancuran!) Sayangnya, dari sudut pandang komputer, semuanya tidak semudah itu. Ketika programmer menulis kode yang memproses gambar digital (bidang ini disebut visi komputer), yang harus mereka kerjakan hanyalah daftar panjang angka yang mewakili warna untuk setiap titik dalam gambar. Gambar 4.2 menunjukkan contoh angka yang membentuk gambar dengan beberapa lusin piksel (titik individual). Gambar yang diambil dengan ponsel pintar Anda akan memiliki jutaan piksel berwarna ini (inilah arti megapiksel).

Mengingat data yang kita miliki hanyalah kumpulan angka yang tampaknya acak, bagaimana kita dapat mendefinisikan fitur? Untuk pertama kalinya dalam buku ini, kita kesulitan melakukannya. Jika kita bisa mengenali beberapa bentuk yang terdapat dalam gambar, pekerjaan kita akan mudah; tetapi, sayangnnya, data kita hanyalah sekumpulan angka mentah. Pendekatan pertama yang dicoba oleh para peneliti visi komputer untuk memecahkan masalah ini adalah menemukan cara cerdas untuk mengekstrak fitur secara manual. Mereka akan membangun alat matematika kompleks yang disebut filter yang dapat menemukan bentuk dasar dalam gambar, dan kemudian mencoba menggabungkannya untuk mendeteksi elemen. Pendekatan ini dapat digunakan untuk objek dasar. Jika Anda memikirkan bentuk-bentuk kepala pancuran, Anda mungkin akan membayangkan lingkaran dan silinder yang mewakili gagangnya. Hal itu bisa diterapkan pada banyak kepala pancuran, tetapi bagaimana dengan desain seperti yang ada di sebelah kanan pada gambar 4.3?



207	202	136	89	119	103	117	179	216	213	210	212	216	215	215	220	217	219	224	218	215	222	215	202	211	224	216	219	219	221	215	219	219	215	211	206	198	195	193	189
202	206	130	81	105	104	130	212	215	198	202	214	213	216	221	215	216	216	214	220	228	216	184	155	196	220	217	217	216	222	216	215	207	199	192	190	190	187	178	
204	203	147	70	105	94	174	216	211	208	204	209	219	217	214	221	221	223	214	219	215	187	172	168	176	215	220	219	216	221	214	206	194	193	191	189	191	192	188	
201	202	154	74	86	114	202	180	172	202	213	209	214	218	215	214	211	217	216	224	203	169	175	186	165	213	222	222	216	215	206	197	195	198	202	201	196	197	197	194
200	215	155	68	77	178	202	170	160	144	176	219	218	211	217	216	222	222	220	215	188	171	179	168	166	213	217	218	210	204	198	195	200	198	197	198	199	199	190	176
205	208	174	25	128	207	201	187	194	146	138	177	208	217	217	212	200	208	214	187	155	167	181	161	170	213	209	209	203	197	196	198	197	196	197	197	197	188	173	
203	196	171	31	179	195	203	192	197	201	160	125	154	189	181	162	128	145	170	152	137	163	167	149	169	213	206	206	203	200	202	203	198	200	202	200	198	198	191	179
204	213	143	111	188	203	205	191	177	194	185	149	117	99	101	119	113	105	121	119	136	167	148	125	164	212	207	209	209	209	210	206	211	205	199	197	200	202	190	170
203	204	175	171	193	197	196	202	189	176	162	155	132	98	64	125	115	130	78	137	119	141	147	113	145	207	208	208	209	212	208	208	205	205	208	201	203	198	188	156
200	205	182	177	195	197	196	201	194	158	161	150	130	133	110	115	113	106	81	127	130	96	122	91	102	205	219	203	209	213	202	207	203	203	205	196	196	192	184	151
199	205	188	180	192	195	195	201	205	134	129	136	143	143	114	91	67	102	114	102	119	111	135	127	121	193	216	214	213	216	212	212	211	207	204	194	195	194	186	145
199	202	189	177	184	192	196	202	194	145	147	165	172	152	134	100	99	93	133	136	122	114	88	120	135	177	206	213	210	212	213	210	214	208	203	196	199	196	179	122
198	199	190	176	179	191	197	199	206	200	185	177	154	137	161	136	113	117	165	128	107	146	102	43	101	180	224	214	215	215	205	207	205	203	202	198	198	190	166	95
195	196	193	178	176	190	196	196	200	188	124	129	130	108	129	155	129	130	137	104	127	120	124	92	52	132	207	207	205	210	210	209	203	203	202	196	193	187	169	95
195	196	196	178	168	183	192	193	213	180	118	151	149	124	109	144	158	145	97	94	135	94	104	121	94	129	202	213	203	204	209	202	206	204	200	196	194	189	168	85
196	196	197	174	159	176	189	193	198	168	159	178	111	130	135	130	174	124	62	122	146	128	89	84	122	150	200	202	210	207	202	207	201	199	196	199	199	187	150	48
193	192	191	179	153	170	183	193	197	207	179	158	128	160	165	134	121	85	73	138	150	170	149	124	109	142	170	163	180	192	196	200	197	199	190	202	195	176	134	43
197	188	195	183	144	158	180	193	191	203	166	160	171	189	162	151	100	36	147	116	137	188	148	138	139	143	77	81	42	82	130	160	189	203	203	184	190	178	131	46
195	188	191	183	148	154	179	186	197	206	187	171	170	200	172	135	146	116	138	128	174	127	116	122	110	41	100	119	30	6	39	50	104	140	193	193	191	171	116	30
189	193	182	184	160	148	176	177	189	196	205	179	136	164	180	169	172	168	163	167	143	112	96	93	81	105	105	110	119	25	43	118	73	60	102	170	181	164	116	52
182	188	176	188	151	112	159	170	174	183	196	202	183	166	155	172	196	190	167	133	63	92	83	75	113	113	87	122	138	80	4	81	136	68	3	93	150	155	103	73
186	183	183	193	143	91	139	166	184	198	183	178	193	198	191	190	181	199	174	125	48	51	101	93	52	0	111	129	106	104	14	16	120	111	0	17	104	150	118	116
183	174	179	174	160	122	119	140	171	217	210	182	174	189	206	187	206	178	192	123	50	70	67	9	6	83	134	87	130	104	64	20	71	135	71	0	54	110	126	130
154	150	149	129	171	156	88	90	159	201	187	178	179	168	190	199	190	205	184	108	39	50	67	96	109	129	71	59	122	149	86	22	38	105	116	22	34	62	104	96
113	107	84	61	109	133	83	76	128	182	135	82	105	108	135	152	178	188	138	57	68	108	105	109	108	105	80	68	99	109	90	79	0	30	134	77	4	27	46	81
144	141	144	128	132	145	130	130	161	202	178	119	97	57	34	29	110	130	98	103	112	77	76	86	120	81	51	96	112	86	92	79	47	3	99	122	54	24	26	38
194	187	195	192	179	182	186	184	184	203	209	178	156	132	111	99	78	54	110	119	77	83	87	89	59	65	96	104	85	80	68	58	61	15	37	133	109	10	24	27
200	198	191	194	196	193	193	190	188	191	198	173	133	124	131	120	106	107	89	102	97	71	38	4	9	44	81	57	37	34	5	36	41	46	2	77	150	56	11	21
198	206	197	197	202	198	199	199	195	201	202	178	118	98	102	65	76	44	43	94	49	0	16	64	103	128	101	107	117	75	20	25	61	88	45	4	113	100	0	43
197	200	201	200	197	198	203	198	195	193	188	190	162	155	128	52	51	128	135	111	139	164	131	169	172	148	90	95	121	101	38	34	8	59	66	3	82	103	0	25
202	194	199	199	195	200	199	193	198	191	186	190	186	206	177	113	151	197	170	165	135	147	156	99	111	64	35	68	96	86	26	42	28	49	52	27	79	99	41	12
195	189	194	191	191	192	184	190	186	193	198	168	133	156	153	143	186	173	155	117	49	80	85	31	84	62	8	52	79	68	41	28	30	21	45	40	78	100	55	26
184	184	185	186	185	184	182	181	185	197	185	142	84	70	86	65	154	156	106	109	73	36	91	83	104	65	3	50	79	73	48	26	37	30	26	11	70	81	0	3
207	207	208	208	208	207	206	205	203	210	176	122	73	53	49	25	101	121	92	93	69	54	104	92	99	95	48	51	50	52	28	1	0	1	3	6	55	81	57	73
208	209	209	210	209	209	208	207	218	201	124	43	3	0	2	0	0	50	65	58	17	2	48	55	47	26	0	1	0	2	15	46	74	90	99	114	132	141	147	148
205	205	206	206	206	205	204	204	200	205	174	141	128	119	114	110	117	126	129	136	131	125	127	112	108	110	111	121	121	131	131	148	157	164	162	168	165	163	173	164
208	208	209	209	209	208	208	208	209	213	208	205	205	203	204	203	203	203	208	208	204	201	202	206	196	201	199	188	185	199	190	189	185	182	177	175	174	175	177	174
207	207	207	207	207	207	207	207	209	204	204	209	210	210	211	208	212	208	207	202	207	214	208	207	210	202	202	209	208	202	195	202	197	196	198	194	194	193	184	185
207	207	207	207	207	207	207	207	208	206	205	206	206	207	207	203	199	205	210	211	205	207	209	203	203	207	207	202	202	202	202	208	207	208	206	203	201	200	199</	



Kepala Shower A

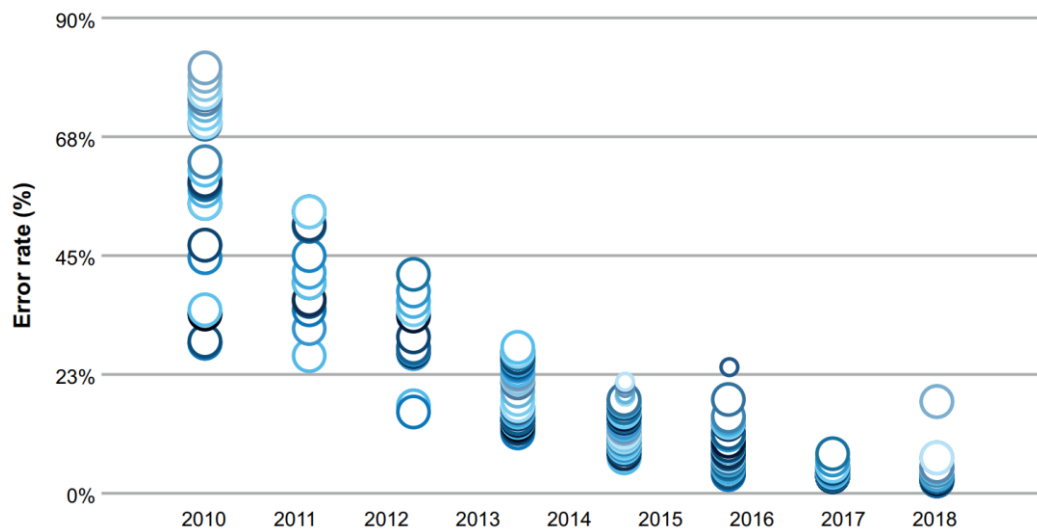


Kepala Shower B



Kepala Shower C

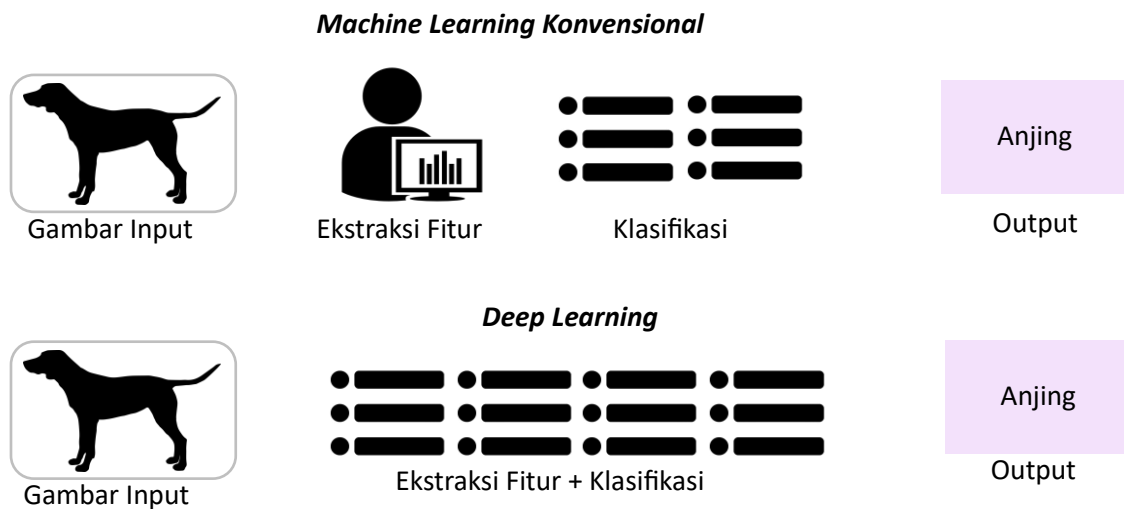
Gambar 4.3 Meskipun memungkinkan untuk mengidentifikasi A dan B sebagai kepala pancuran karena bentuk dasarnya, mengidentifikasi C lebih sulit.



Gambar 4.4 Pembelajaran mendalam telah mendominasi kompetisi ImageNet sejak kemunculannya pada tahun 2013.

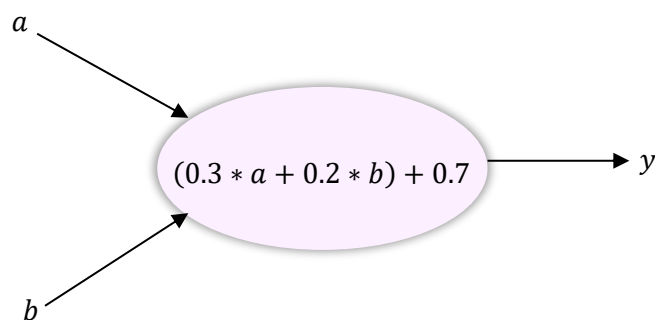
4.2 PEMBELAJARAN MENDALAM UNTUK KLASIFIKASI GAMBAR

Pertama-tama, algoritma pembelajaran mendalam tidak sepenuhnya berbeda dari algoritma pembelajaran mesin yang selama ini kita gunakan. Algoritma pembelajaran mendalam pada intinya masih merupakan algoritma pembelajaran terawasi: ia mempelajari pemetaan antara masukan dan keluaran dari data pelatihan. Perbedaan utamanya terletak pada jenis data masukan yang digunakannya: sementara algoritma pembelajaran mesin tradisional membutuhkan fitur yang disempurnakan (bagian atas gambar 4.5), algoritma pembelajaran mendalam bekerja langsung dengan angka mentah dan dapat memahami fitur yang relevan secara otomatis (bagian bawah gambar 4.5).



Gambar 4.5 Dalam pembelajaran mesin tradisional, para insinyur harus mengembangkan algoritma untuk mengekstrak fitur yang dapat dimasukkan ke dalam model. Model pembelajaran mesin tidak memerlukan langkah awal yang rumit ini.

Mari kita pahami mengapa pembelajaran mendalam memiliki kemampuan yang begitu ajaib. Pembelajaran mendalam didasarkan pada kelas algoritma khusus yang disebut jaringan saraf dalam. Anda mungkin berpikir bahwa keajaibannya terletak pada kata neural, yang memunculkan gambaran otak buatan yang mahakuasa. Sebenarnya, neuron buatan sama sekali tidak seperti neuron biologis. Mereka sangat sederhana: mereka hanya dapat melakukan perkalian dan penjumlahan.

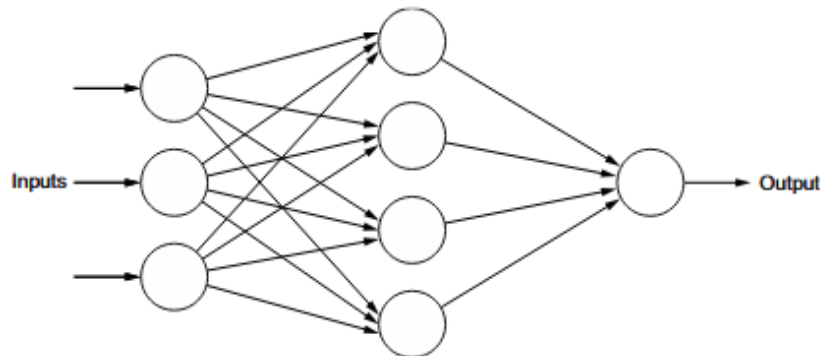


Gambar 4.6 Sebuah neuron buatan tunggal menerima dua angka (a dan b), melakukan operasi matematika sederhana, dan mengeluarkan angka lain, y.

Seperti yang ditunjukkan pada Gambar 4.6, setiap neuron menerima angka masukan, melakukan operasi matematika sederhana pada angka tersebut, dan menghasilkan angka keluaran. Sebuah neuron sendiri sederhana dan, sejujurnya, tidak berguna. Namun, ketika Anda menghubungkan banyak neuron bersama-sama untuk membentuk sebuah jaringan, mereka menjadi mampu melakukan komputasi yang canggih.

Sebuah jaringan sederhana ditunjukkan pada Gambar 4.7: kolom pertama neuron menerima masukan awal dan meneruskan informasi yang telah diproses ke kolom tengah

neuron, dan seterusnya. Akhirnya, neuron tunggal di kolom kanan menghasilkan keluaran akhir jaringan.



Gambar 4.7 Hal-hal mulai menjadi menarik ketika Anda menghubungkan beberapa neuron bersama-sama, karena mereka menjadi mampu mengekspresikan operasi matematika yang lebih kompleks.

Kembali ke klasifikasi gambar, yang ingin kita lakukan adalah membuat jaringan yang menerima piksel gambar, melakukan beberapa komputasi internal, dan mengeluarkan kelas yang (semoga) tepat untuk gambar masukan. Gambar 4.8 menunjukkan bagaimana semua ini saling terkait. Tentu saja, jaringan di dunia nyata akan jauh lebih kompleks daripada yang ada di Gambar 4.7: model kompetitif untuk ImageNet memiliki sekitar setengah juta neuron yang semuanya bekerja bersama!

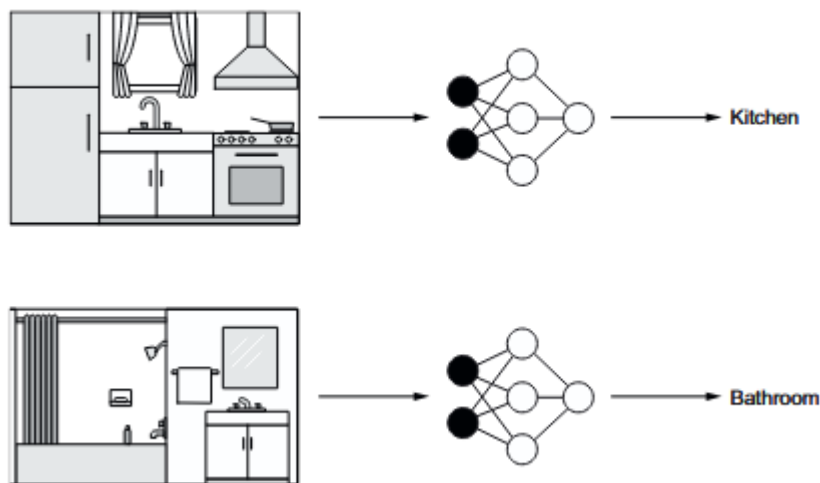
Dalam jaringan saraf dalam, neuron diatur dalam kolom (atau lapisan), dan struktur inilah yang membuatnya mampu mengekstraksi fitur dari gambar tanpa menentukannya secara manual. Mengapa dan bagaimana?

Anda dapat membayangkan setiap lapisan sebagai langkah dalam jalur produksi: angka mentah yang menyusun gambar masuk ke dalam lapisan pertama, tempat neuron di dalamnya memprosesnya dengan menerapkan operasi matematika mereka (perkalian dan penjumlahan). Hasilnya adalah angka-angka mentah ini dimodifikasi dan diubah menjadi angka-angka lain yang kemudian diteruskan ke lapisan kedua. Lapisan kedua menerapkan transformasi lain dan mengirimkan hasilnya ke lapisan ketiga. Proses ini diulang untuk setiap lapisan hingga lapisan terakhir berkata, "Gambar ini berisi dapur" atau "Gambar ini berisi kamar mandi."

Analogi dengan lini produksi ini masuk akal karena tidak ada mesin ajaib yang dapat mengubah bahan mentah seperti baja dan kulit menjadi mobil baru yang mengkilap dalam satu langkah. Jika Anda pernah menonton video lini produksi, Anda akan melihat bahwa proses pembuatan mobil dibagi menjadi langkah-langkah kecil: bahan mentah pertama-tama dibentuk oleh mesin menjadi komponen-komponen dasar kecil termasuk kenop, roda gigi, dan sebagainya. Komponen-komponen ini kemudian diproses oleh mesin lain yang membentuk tuas persneling, roda kemudi, radio, dan sebagainya. Baru setelah komponen-komponen kompleks ini dibuat, komponen-komponen tersebut akhirnya dirakit untuk membentuk sebuah mobil; mustahil memiliki satu mesin raksasa yang mengubah bahan mentah menjadi

mobil dalam satu langkah. Setiap lapisan jaringan saraf dalam pada dasarnya melakukan hal yang sama, hanya saja ia bekerja pada angka, bukan baja.

Transformasi inilah yang menjadi kunci kekuatan magis ekstraksi fitur otomatis. Ketika jaringan saraf pertama kali dilatih pada sejumlah besar gambar, akurasi awalnya sangat buruk, karena neuron di berbagai lapisan belum menemukan cara untuk memproses angka-angka ini secara bermakna. Seiring jaringan melihat lebih banyak gambar dan pelatihannya berlanjut, teknik matematika khusus yang disebut backpropagation mendorong setiap neuron untuk menyesuaikan transformasi yang diterapkannya sehingga jaringan secara keseluruhan meningkatkan akurasi klasifikasinya. Semakin banyak gambar yang dilihat jaringan, semakin banyak neuronnya akan mempelajari transformasi yang bermakna.



Gambar 4.8 Jaringan saraf dalam memproses gambar masukan (di sebelah kiri) dan menghasilkan klasifikasi yang konsisten dengan label pelatihan (di sebelah kanan).

Namun, transformasi apa sajakah ini? Kedengarannya samar, dan pada masa-masa awal pembelajaran mendalam, transformasi ini juga agak misterius bagi para peneliti yang membangun teknologi ini. Setelah jaringan saraf dalam mencapai kinerja yang luar biasa, para peneliti mulai menggali lebih dalam cara kerja internalnya. Apa yang mereka temukan sangat menarik: ketika neuron di lapisan pertama mengoptimalkan transformasinya, mereka melakukannya dengan cara yang memungkinkan mereka mengenali garis, tepi, dan sudut sederhana dari piksel mentah, serupa dengan filter buatan tangan yang coba dibuat oleh para insinyur visi komputer pertama.

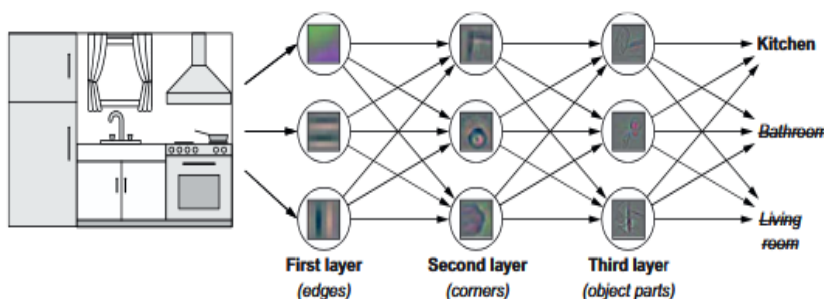
Lapisan berikutnya akan menerima informasi yang jauh lebih kaya daripada angka mentah awal: angka-angka baru ini merepresentasikan keberadaan bentuk di berbagai bagian gambar, alih-alih warna satu piksel. Lapisan kedua ini akan membangun informasi ini, dan menerapkan transformasi lain yang memungkinkannya mengenali bentuk yang lebih kompleks seperti lingkaran dan persegi, dimulai dari garis, tepi, dan sudut yang lebih sederhana. Ke depannya, suatu lapisan pada akhirnya akan mengenali bentuk-bentuk yang kita sukai dalam contoh kamar mandi versus dapur: pancuran, oven, toilet, microwave, dan sebagainya. Gambar 4.9 menunjukkan seperti apa bentuk-bentuk yang semakin kompleks ini. Terakhir,

lapisan terakhir harus membuat keputusan akhir tentang kelas gambar tersebut, berdasarkan elemen yang diekstraksi oleh lapisan sebelumnya.

Keajaiban pembelajaran mendalam terletak pada kenyataan bahwa semua neuron dan lapisan ini telah belajar secara otonom untuk menguraikan jutaan angka yang membentuk suatu gambar dari bentuk dasar menjadi bentuk kompleks, lalu membuat prediksi berdasarkan angka-angka tersebut. Pada dasarnya, struktur modular jaringan saraf memungkinkannya memproses informasi secara bertahap, dan berkat formulasi matematika yang cerdas, jaringan belajar menyesuaikan langkah-langkah ini untuk mengenali ciri-ciri yang lebih khas dari gambar yang dipelajarinya: dalam kasus kita, wastafel, lemari es, oven, dan sebagainya.

Bentuk-bentuk ini tampak persis seperti fitur yang kita cari di awal bab ini. Berkat keajaiban pembelajaran mendalam, mereka telah diidentifikasi secara otomatis: tidak ada yang memberi tahu lapisan pertama cara mengenali garis dan sudut, tidak ada yang memberi tahu lapisan berikutnya cara menggabungkannya untuk membentuk lingkaran dan persegi panjang, dan tidak ada yang menemukan cara cerdas untuk mengajari lapisan terakhir cara mengenali elemen kompleks seperti kepala pancuran. Yang kita lakukan hanyalah mengambil jaringan saraf dengan beberapa lapisan, memberinya gambar dapur dan kamar mandi, dan membiarkan matematika di balik backpropagation mengerjakan sisanya.

Apakah semua keajaiban pembelajaran mendalam ini gratis? Sayangnya, tidak. Kita meminta jaringan kita untuk melakukan banyak hal, dan para Dewa Pembelajaran Mendalam menuntut pengorbanan: banyak daya komputasi dan banyak data.



Gambar 4.9 Representasi internal yang dikembangkan oleh jaringan saraf dalam yang dilatih untuk mengenali mobil, orang, dan hewan

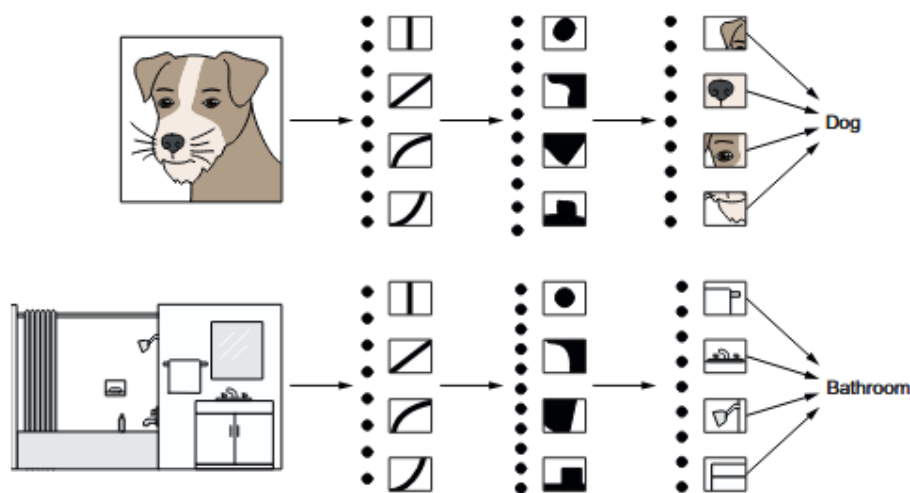
Meskipun daya komputasi semakin murah, memperoleh lebih banyak data berlabel tidaklah semudah itu. Dalam kasus pialang rumah, Anda harus menunggu untuk mencapai ribuan gambar ruangan rumah, dan menunggu mungkin menyakitkan. Apakah ini berarti bahwa jika bisnis Anda tidak memiliki kumpulan data yang besar, Anda harus melepaskan pembelajaran mendalam? Untungnya, teknologi yang disebut transfer learning merupakan aset berharga yang memungkinkan Anda membangun algoritma mutakhir, mulai dari kumpulan data kecil dan kode sumber terbuka. Mari kita lihat caranya.

4.3 PEMBELAJARAN TRANSFER DENGAN SET DATA KECIL

Pembelajaran transfer adalah semacam trik yang memanfaatkan struktur berlapis jaringan saraf dalam untuk mentransfer "pengetahuan" mereka dari satu tugas ke tugas lainnya. Seperti yang Anda lihat pada gambar 4.10, jaringan saraf dalam bekerja dengan

mengenali bentuk-bentuk dasar dalam gambar. Misalnya, untuk mengenali oven, Anda perlu mengenali lingkaran dan persegi panjang, lalu menggabungkannya untuk mengenali kenop, gagang, gelas, dan sebagainya.

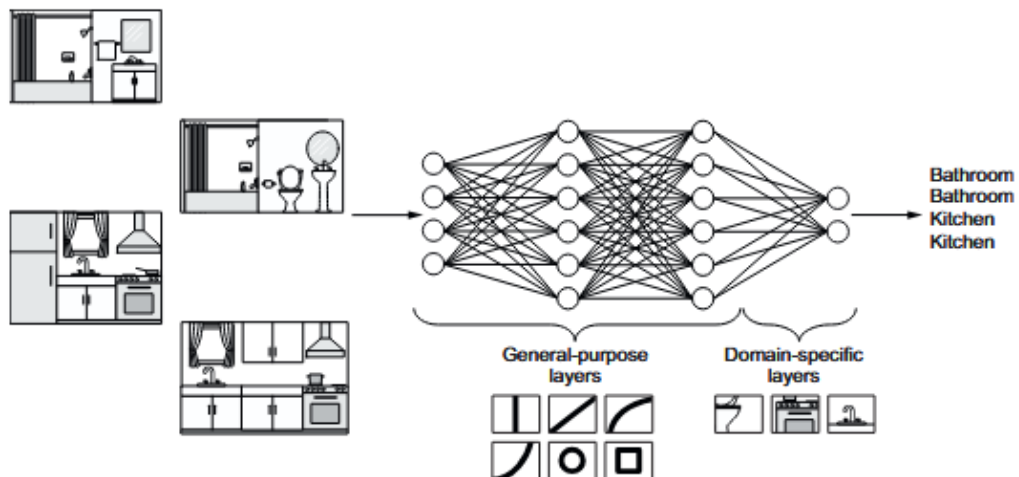
Jika dipikir-pikir, lingkaran, persegi panjang, dan bentuk dasar lainnya bukanlah hal yang unik dalam masalah pelabelan ruangan. Jika Anda harus membedakan anjing dari kucing, Anda tetap perlu mengenali lingkaran, segitiga, dan persegi panjang sebelum Anda dapat mengenali mata, telinga, ekor, dan sebagainya. Lagipula, lingkaran dan garis terlihat sama, terlepas dari apakah mereka muncul dalam gambar anjing atau kamar mandi. Hanya lapisan terakhir yang dikhususkan untuk tugas klasifikasi tertentu, seperti yang ditunjukkan pada gambar 4.10.



Gambar 4.10 Kesamaan antara pengklasifikasi berbasis DL untuk ruangan dan hewan.

Lapisan pertama belajar mengidentifikasi pola geometris umum yang sama.

Karena internet penuh dengan gambar kucing dan anjing, dapatkah kita menggunakan semua data ini untuk melatih jaringan saraf tiruan yang besar dan menggunakan kembali data yang kita miliki untuk tugas kategorisasi ruangan kita? Kita bisa, dan inilah konsep inti dari pembelajaran transfer: fitur geometris dasar yang dipelajari di beberapa lapisan pertama sebagian besar sama di sebagian besar tugas klasifikasi gambar, seperti yang ditunjukkan pada Gambar 4.11. Hal ini memungkinkan Anda memanfaatkan kumpulan data yang sangat besar (seperti ImageNet) untuk melakukan pra-pelatihan jaringan saraf tiruan Anda, dan kemudian mengkhususkannya untuk klasifikasi ruangan dengan melatihnya pada kumpulan data properti.



Gambar 4.11 Dalam pembelajaran transfer, lapisan-lapisan umum digunakan kembali untuk beberapa tugas, sedangkan lapisan-lapisan khusus domain dilatih untuk masalah tertentu.

Kabar baiknya tidak berhenti di sini: banyak rilis jaringan saraf sumber terbuka telah dilatih terlebih dahulu pada kumpulan data besar seperti ImageNet. Kembali ke tugas awal kita, berikut langkah-langkah yang akan diikuti oleh ilmuwan data Anda untuk membangun pengklasifikasi ruangan kita:

1. Unduh implementasi jaringan saraf sumber terbuka yang telah dilatih terlebih dahulu pada kumpulan data generik seperti ImageNet.
2. Pertahankan lapisan pertama, karena lapisan tersebut mampu mengenali bentuk dasar (lingkaran dan persegi panjang), dan abaikan lapisan terakhir, karena lapisan tersebut telah disetel untuk mengenali gambar-gambar spesifik yang terdapat dalam ImageNet (dari truk hingga macan tutul).
3. Ambil kumpulan data gambar ruangan Anda yang relatif kecil (ratusan gambar, bukan ribuan), dan latih ulang lapisan-lapisan terakhir jaringan. Lapisan-lapisan ini akan menggunakan informasi tentang bentuk-bentuk dasar yang terdapat dalam gambar dan belajar mengidentifikasi elemen-elemen seperti kepala pancuran, oven, dan sebagainya. Seperti yang Anda lihat, prosesnya relatif efisien, tetapi keuntungan teknologi ini bagi organisasi mana pun cukup signifikan:
 - Anda bisa mendapatkan kinerja mutakhir tanpa kumpulan data besar, hanya dengan beberapa ratus gambar dari domain spesifik Anda.
 - Anda tidak memerlukan banyak daya komputasi, karena jaringan dasar yang Anda gunakan telah dilatih oleh pihak lain yang merilisnya sebagai sumber terbuka (umumnya, perusahaan besar seperti Google atau Facebook).
 - Melatih jaringan saraf besar pada kumpulan data besar tidaklah mudah; dengan menggunakan versi sumber terbuka ini, Anda membangun pengalaman para peneliti raksasa teknologi yang berpenghasilan jutaan dolar.

4.4 MELATIH KOMPUTER UNTUK PENGENALAN WAJAH

Film penting RoboCop dirilis pada tahun 1987. Di masa depan distopia ini, kejahatan melonjak di Detroit, dan departemen kepolisian kekurangan dana, sehingga memberi ruang bagi perusahaan raksasa *Omni Consumer Products* (OCP) untuk menyediakan produk eksperimental berteknologi tinggi bagi penegak hukum. Petugas Alexander James "Alex" Murphy disiksa secara brutal dan, dinyatakan meninggal, menjadi tempat uji coba teknologi sibernetik OCP baru yang mengubahnya menjadi cyborg bernama RoboCop.

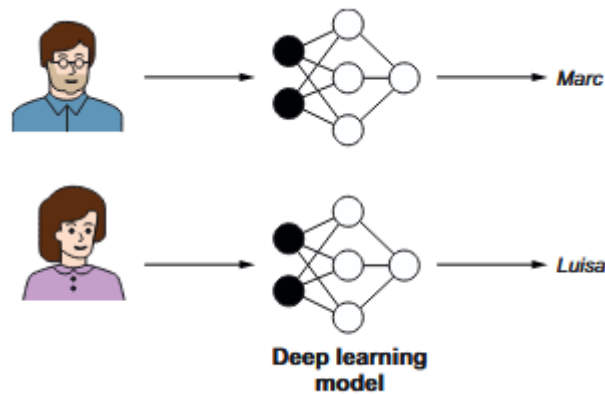
Masih dihantui oleh kehidupan masa lalunya sebagai manusia, suatu hari RoboCop melihat sekelompok orang yang ia kenali terkait dengan eksekusinya dan mengambil foto salah satu dari mereka. Ia mengunggah foto tersebut ke komputer polisi, yang kemudian memindai catatan polisi dan menemukan identitas orang tersebut. Aplikasi ini disebut Pengenalan Wajah, dan pada tahun 1987 menjadi subjek film fiksi ilmiah.

Jika Anda baru saja membeli ponsel pintar, teknologi ini mungkin sudah ada di saku Anda. Satu-satunya perbedaannya adalah, jika di film-film lama digunakan untuk mencari penjahat, kini penggunaannya yang paling umum adalah untuk... membuka kunci ponsel. Pengenalan wajah adalah aplikasi lain dari pembelajaran mendalam (Gambar 4.12) yang tampak seperti masalah klasifikasi tradisional, tetapi menambahkan beberapa aspek menarik.

Misalnya, Anda mencoba membuat kunci pintu pintar yang menggunakan gambar dari kamera untuk membuka kunci pintu depan rumah Anda berdasarkan siapa yang berdiri di depannya. Jika Anda, pasangan Anda, atau putri Anda mengetuk pintu, kamera akan mengambil gambar wajah tersebut, dan jika mengenali orang tersebut sebagai salah satu dari ketiganya, kamera akan membuka pintu. Jika tidak mengenali siapa pun, pintu akan tetap terkunci. Bagaimana kita membangun sesuatu seperti ini? Mari kita mulai dengan menguraikan masalahnya.

Data masukan kita adalah gambar wajah. Tugasnya adalah memahami siapa pemiliknya: Anda, pasangan Anda, putri Anda, atau orang lain. Sama seperti contoh dapur/kamar mandi di awal bab ini, ini adalah tugas klasifikasi terawasi. Seperti yang telah kita lihat, pembelajaran mendalam adalah alat terbaik untuk memecahkan sebagian besar masalah klasifikasi gambar. Kita bahkan dapat memanfaatkan pembelajaran transfer dengan melatih ulang lapisan terakhir jaringan yang dilatih di ImageNet.

Seperti biasa, langkah selanjutnya adalah menemukan data pelatihan: Anda membuka album keluarga dan mengambil beberapa ratus foto keluarga Anda. Penting bagi Anda untuk mendapatkan sebanyak mungkin, agar algoritme memiliki contoh wajah Anda dalam kondisi cahaya yang berbeda, potongan rambut yang berbeda, dan sebagainya. Untuk melengkapi kumpulan data, Anda mengunduh beberapa ratus foto orang acak dari internet dan melabelinya sebagai "lainnya". Ini agar model akan belajar mengenali fitur paling khas pada wajah keluarga Anda, dan mengabaikan atribut apa pun yang dimiliki oleh semua wajah manusia.



Gambar 4.12 Pengenalan wajah merupakan aplikasi spesifik klasifikasi citra yang inputnya berupa wajah manusia.

Pendekatan ini sepertinya berhasil, dan memang berhasil. Namun, inilah tantangannya: putri Anda masih SMA, dan temannya, Catherine, sering berkunjung ke rumah Anda untuk mengerjakan PR bersama putri Anda. Akan lebih baik jika Anda menambahkan Catherine ke daftar orang terpercaya agar pintunya juga bisa dibuka. Sekarang Anda memiliki lima kelas, bukan empat:

- Anda
- Pasangan
- Putri
- Catherine, teman belajar
- Lainnya

Anda harus meminta Catherine beberapa ratus foto wajahnya dan melatih model yang sama sekali baru. Baik Anda, Catherine, maupun orang tuanya tidak akan senang melakukan itu hanya untuk mempercepat proses pembukaan pintu.

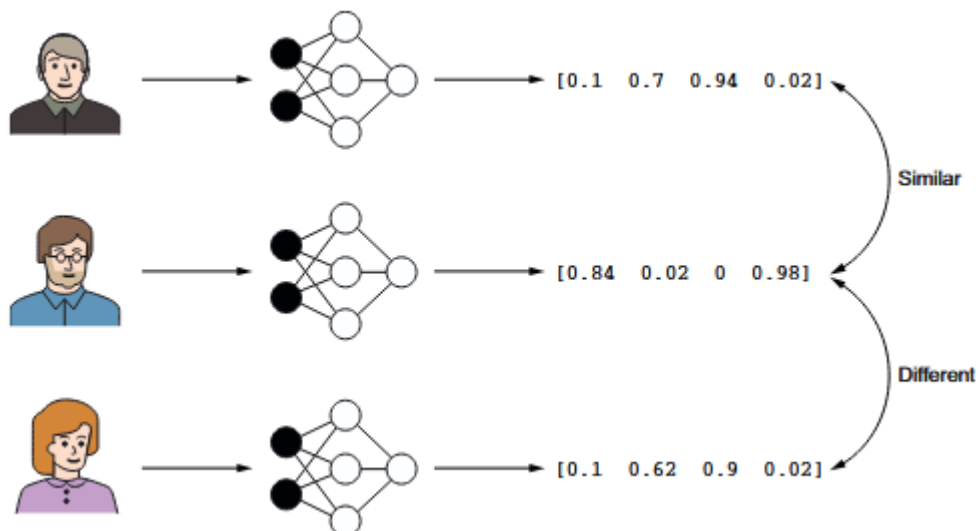
Masalah ini mirip dengan teknologi pengenalan wajah yang kini hadir di banyak ponsel pintar. Ketika Apple merilis iPhone X dengan Face ID, teknologi ini berfungsi dengan baik tanpa perlu mengunggah ratusan foto dan menunggu beberapa jam untuk melatih ulang jaringan saraf. Anda hanya perlu menatap kamera selama beberapa detik, dan selesai. Apakah ini sihir, atau apa?

Ternyata itu bukan sihir; melainkan teknologi yang disebut pembelajaran sekali pakai: trik lain yang memungkinkan jaringan saraf dalam khusus belajar mengklasifikasikan item hanya dengan menggunakan satu contoh pelatihan, alih-alih ribuan yang biasanya dibutuhkan.

Semuanya bermuara pada kemampuan jaringan saraf untuk mengenali bentuk dan pola dalam gambar. Jika Anda melatih jaringan untuk mengenali wajah, ia akan mengikuti langkah-langkah yang sama yang kami soroti untuk tugas "dapur versus kamar mandi", mengenali garis dan tepi terlebih dahulu, lalu secara bertahap bentuk yang semakin kompleks. Apa saja bentuk kompleks ini dalam pengenalan wajah? Bentuk-bentuk tersebut akan berupa ciri-ciri wajah: berbagai bentuk yang dapat diambil oleh hidung, dahi, atau bibir.

Di dalam perangkat lunak setiap iPhone, jaringan saraf dilatih pada ribuan wajah. Saat Anda mengatur Face ID dan pertama kali melihat wajah Anda, Face ID akan menguraikan citra wajah Anda untuk mengenali bentuk dasar dan kemudian ciri-ciri wajah utama, yang akan dikodekan menjadi daftar kecil angka (dari ratusan hingga ribuan). Daftar angka ini disebut *embedding*.

Setiap kali Anda membuka kunci ponsel, Face ID akan mengambil gambar wajah Anda, menjalankannya melalui jaringan sarafnya, mendapatkan *embedding*-nya, dan membandingkannya dengan yang disimpan saat pertama kali Anda mengaturnya. Jika angkanya mendekati, artinya ciri-ciri wajah pada gambar tersebut juga mirip dan oleh karena itu milik pemilik ponsel, dan ponsel akan terbuka (dua baris teratas pada gambar 4.13). Jika *embedding* sangat berbeda dari yang "emas", pasti orang yang berbeda (dua baris terbawah pada gambar 4.13). Selesai! Sederhana, cepat, efisien, dan akurat.



Gambar 4.13 Dalam pengenalan wajah, swafoto diterjemahkan menjadi *embedding* (daftar angka) yang dapat dibandingkan dengan gambar yang sudah dikenal.

Embedding (atau *encode*) disebut demikian karena *embedding* tersebut menanamkan informasi dalam gambar asli ke dalam beberapa angka, mengodekan sejumlah besar informasi yang terkandung dalam gambar mentah menjadi sesuatu yang lebih ringkas, praktis, dan bermakna (dari jutaan angka dari kamera depan menjadi hanya beberapa ratus).

Embedding membuka banyak peluang, dan Anda baru saja melihat bagaimana pembelajaran satu langkah merupakan cara yang hebat untuk mengimplementasikan fitur-fitur seperti pengenalan wajah dengan beberapa contoh, dan tanpa harus terus-menerus melatih ulang model. Banyak aplikasi menarik lainnya yang dapat dibangun menggunakan *encode*, dan Anda akan segera mempelajari lebih lanjut.

4.5 MENGGUNAKAN PEMBUATAN KONTEN DAN TRANSFER GAYA

Pembuatan konten dan transfer gaya adalah aplikasi AI yang sedang naik daun yang akhirnya mendorong AI dari tahap pemahaman ke tahap penciptaan. Aplikasi pertama yang

akan kita bahas adalah pembuatan konten otomatis. Sebelumnya, telah disebutkan bahwa jaringan saraf akan mencari tahu elemen-elemen relevan apa saja yang membentuk mobil, sepeda motor, kapal, wajah, dan sebagainya. Ketika dilatih dengan wajah yang berbeda, jaringan saraf akan berfokus pada tugas spesifik mengenali wajah, dan oleh karena itu akan menciptakan representasi bentuk pipi, dahi, hidung, dan sebagainya yang berbeda. Apa yang terjadi jika kita memberi jaringan saraf ribuan gambar orang yang tersenyum? Nah, jaringan saraf akan mempelajari apa itu senyuman (kombinasi warna dan bentuk apa yang membentuk kesan wajah tersenyum). Apa yang terjadi jika kita melatihnya dengan ribuan gambar orang berambut hitam keriting? Nah, sekarang jaringan saraf akan mempelajari seperti apa rupa rambut hitam keriting, murni dari data masukan.

Sekarang, jika jaringan saraf mengetahui apa itu senyuman, dapatkah kita menemukan cara untuk membuat jaringan ini menghasilkan senyuman yang benar-benar baru? Atau lebih baik lagi, bisakah kita mengambil wajah sedih dan meminta jaringan saraf untuk memodifikasi pikselnya sehingga menghasilkan senyuman? Ternyata kita bisa melangkah lebih jauh. Jika jaringan saraf yang dilatih dengan gambar senyuman memahami elemen mana yang membentuk senyuman, apa yang terjadi ketika kita melatihnya dengan ribuan gambar selebritas tertentu, katakanlah John Oliver?



Gambar 4.14 Dalam klip yang dihasilkan AI ini, ciri-ciri wajah John Oliver diterapkan pada gambar Jimmy Fallon. (Dihasilkan oleh Gaurav Oberoi, <https://goberoi.com/>.)

Tentu saja, jaringan akan mengodekan ciri-ciri wajah yang membentuk wajahnya. Bagaimana jika kita mengisi jaringan dengan video? Nah, asalkan kita memiliki cukup data, jaringan akan mulai mempelajari tidak hanya elemen mana yang membuat wajah Oliver unik, tetapi juga gerakan dan ekspresi wajahnya yang paling khas. Bisakah kita juga mencoba mengambil ciri-ciri ini dan mengubahnya menjadi orang lain? Kita tentu bisa, dan hasilnya menakjubkan, seperti yang Anda lihat pada gambar sebelum/sesudah pada gambar 4.14.

Sayangnya, beberapa orang telah menggunakan teknologi canggih ini dengan cara yang tidak etis. Anda akan melihat lebih lanjut tentang risiko yang terkait dengan pembuatan konten

otomatis di bab terakhir buku ini. Untuk saat ini, mari kita kembali ke sisi teknologi ini yang lebih positif. Anda telah melihat bagaimana jaringan dapat mempelajari karakteristik yang relevan dari berbagai kelas gambar dan mentransposisikannya dari satu gambar ke gambar lainnya, dan kita berfokus pada wajah orang. Aplikasi apa lagi yang dapat dibangun dengan teknologi yang sama?

Salah satu contohnya adalah transfer gaya: mengambil gaya dari sebuah gambar dan menerapkannya ke gambar lain. Pada Gambar 4.15, model tersebut membuat zebra terlihat seperti kuda (baris atas), dan kuda terlihat seperti zebra (baris bawah). Baris atas tampak hebat: bahkan bulu anak kuda pun ikut terangkat. Namun, tampaknya model tersebut kesulitan membedakan kulit manusia dari bulu kuda: wanita di baris bawah juga akhirnya "di-zebra-kan".

Ini memang menarik, tetapi apa yang bisa Anda lakukan dengannya? Nah, selain menghasilkan gambar-gambar yang indah, sejauhny kita belum tahu! Namun, jangan terlalu cepat mengabaikan aspek artistik dari teknologi ini, karena beberapa perusahaan telah berhasil menciptakan produk yang mampu melakukan hal tersebut.



Gambar 4.15 Transfer gaya beraksi, menerapkan pola zebra pada kuda (dan manusia). Gambar dihasilkan menggunakan kode dari "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks" oleh Jun-Yan Zhu dkk.,
<https://arxiv.org/abs/1703.10593>.

Contoh yang bagus adalah Prisma, sebuah aplikasi yang memungkinkan Anda mengambil gambar dan menerapkan berbagai gaya menggunakan teknik yang baru saja kami sebutkan. Seminggu setelah debutnya di App Store iOS pada Juni 2016, aplikasi ini diunduh lebih dari 7,5 juta kali dan memiliki lebih dari 1 juta pengguna aktif. Hanya sebulan kemudian, aplikasi ini telah diinstal di lebih dari 12,5 juta perangkat dengan lebih dari 1,5 juta pengguna aktif di seluruh dunia, dan masuk dalam 10 aplikasi teratas di App Store di 77 negara.

4.6 YANG PERLU DIWASPADAI

Model DL adalah jenis AI paling kompleks yang pernah kita lihat sejauh ini, dan ini berarti model ini juga memiliki jebakan yang lebih halus. Salah satunya, kekuatan jaringan saraf

dalam terletak pada kemampuannya untuk mengekstrak dan menghitung fitur secara otomatis dari gambar input. Ini hebat karena kinerjanya lebih baik daripada algoritma rekayasa manusia, dan Anda juga tidak perlu membayar gaji. Namun, sisi sebaliknya adalah kita tidak benar-benar tahu mengapa jaringan saraf dalam bekerja dengan sangat baik. Sekadar gambaran, perilaku model klasifikasi gambar rata-rata Anda diatur oleh sekitar 50 juta angka yang diubah dan dioptimalkan selama pelatihan. Bahkan para peneliti dan insinyur terpintar pun sulit memahami kompleksitas ini.

Dalam komunitas ML, model DL secara efektif dianggap sebagai kotak hitam: model ini kuat dan akurat, tetapi kita kesulitan menjelaskan mengapa model ini membuat keputusan tertentu. Dapat dimengerti, para peneliti awalnya berfokus untuk membuat teknologi ini berfungsi, dan baru sekarang mereka mulai mengintip ke dalam kotak hitam tersebut. Kita biasanya mengatakan bahwa model mendalam kurang memiliki kemampuan menjelaskan, dan ini menjadi masalah bagi penerapannya dalam aplikasi yang sangat penting bagi keselamatan.

Misalnya, mobil tanpa pengemudi mungkin menggunakan model dengan akurasi 99% untuk mendeteksi pejalan kaki yang menyeberang jalan. Ini tampak bagus di atas kertas, tetapi mungkin tidak cukup untuk menciptakan produk yang aman: apa yang terjadi dalam 1% kasus sisanya? Dalam pembelajaran mesin konvensional, para insinyur setidaknya memiliki beberapa tingkat wawasan tentang bagaimana model beroperasi dan membuat keputusan. Hal ini seringkali tidak berlaku untuk pembelajaran mendalam, yang mendorong bidang penelitian baru yang disebut AI yang dapat dijelaskan.

Penelitian tentang AI yang dapat dijelaskan mengikuti dua arah utama: mengembangkan keluarga model yang keputusannya intuitif untuk dipahami manusia, dan menciptakan alat dan teknik baru untuk menjelaskan keputusan model yang sudah kita miliki. Dalam kedua kasus tersebut, idenya adalah kita tidak ingin mengorbankan akurasi demi kemampuan menjelaskan. Meskipun penelitian tentang model-model baru masih dalam tahap awal, telah ada banyak hasil menarik dalam alat visualisasi dan debugging model. Salah satu contohnya adalah makalah tahun 2016 oleh tiga peneliti dari University of Washington: “‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier” oleh Marco Ribeiro dkk. (<https://arxiv.org/abs/1602.04938>) yang memberikan banyak bahan renungan (bahkan dilengkapi dengan judul yang menarik).

Salah satu contoh makalah tersebut adalah aplikasi visi komputer yang menggunakan jaringan saraf tiruan dalam yang dilatih untuk mengklasifikasikan husky dan serigala. Ini merupakan tugas yang sulit, karena kedua hewan tersebut tampak serupa. Para peneliti melakukan sesuatu yang licik: mereka mengumpulkan gambar husky peliharaan saat mereka berada di rumah, dan gambar serigala yang hidup di alam liar di tempat yang dingin. Sebuah pengklasifikasi yang dilatih berdasarkan data ini berkinerja sangat baik, dan seorang insinyur pembelajaran mesin yang naif mungkin tergoda untuk berhenti dan mengirimkan produknya.

Namun, jika Anda menggali lebih dalam dan mencoba memahami apa yang diamati oleh algoritma tersebut fitur apa yang dibangunnya untuk memahami jenis hewan tersebut Anda akan menemukan sesuatu yang aneh. Algoritme sama sekali tidak peduli dengan hewan

tersebut. Satu-satunya fitur yang diidentifikasi adalah apakah ada salju di gambar. Karena semua gambar serigala memiliki latar belakang bersalju, cukup mudah baginya untuk memahami "Jika saya melihat latar belakang putih, maka itu serigala." Jika Anda memutuskan untuk menguji algoritma dengan husky di atas salju, algoritma akan selalu mengatakan "serigala". Bahkan, bahkan jika Anda meletakkan pug di atas salju, algoritma akan terus mengatakan "serigala"! Lucunya: kami mencoba membangun algoritma cerdas yang dapat membedakan serigala dari husky, dan yang kami dapatkan adalah algoritma yang menghitung piksel putih.

Lalu, apa inti sarinya? Seorang filsuf besar pernah berkata, "Semakin besar kekuatan, semakin besar pula tanggung jawabnya" (sebenarnya itu adalah Paman Ben dari Spiderman). Hal ini terutama berlaku untuk jaringan saraf dalam. Kemampuannya untuk mempelajari fitur secara mandiri sangat kuat tetapi harus ditangani dengan hati-hati: model tersebut mungkin telah mempelajari sesuatu yang tidak Anda duga. Memeriksa semua yang telah dipelajari oleh algoritma yang dilatih pada jutaan gambar adalah hal yang mustahil, jadi Anda harus bertindak cerdas. Cara bermain cerdas adalah dengan mengingat apa yang telah kami sampaikan di awal buku ini: dalam AI, satu hal terpenting yang perlu diperhatikan adalah data. Jika model menunjukkan bias, Anda harus melihat datanya untuk menemukan penyebab sekaligus solusinya.

4.7 AI UNTUK AUDIO

Data audio menyajikan banyak kompleksitas yang sama dengan gambar, karena datanya tampak serupa: segudang angka yang mewakili bentuk gelombang dan puncak audio. Oleh karena itu, kami akan membatasi diri untuk menyebutkan tiga aplikasi utama AI untuk audio:

- Ucapan ke teks, atau pengenalan ucapan
- Teks ke ucapan, atau pembangkitan ucapan
- Deteksi kata pemicu

Mari kita bahas masing-masing. Pengenalan ucapan adalah tentang memahami kata-kata dari rekaman audio dialog lisan. Inputnya adalah audio mentah, dan outputnya adalah transkripnya (gambar 4.16). Anda telah menggunakan teknologi ini setiap kali Anda bertanya kepada Siri, Asisten Google, atau Alexa.

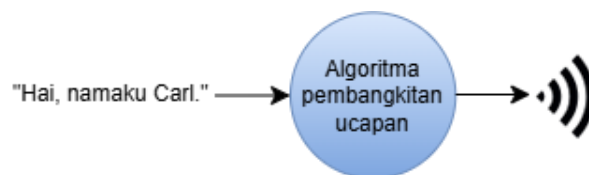
Ini adalah teknologi lain yang tidak terlalu akurat hingga pembelajaran mendalam berkembang dan kita mampu mengumpulkan dan memproses kumpulan data besar. Jika dipikir-pikir, masalahnya sama dengan gambar: jika Anda mendengarkan berkas audio berisi rekaman seseorang yang mengucapkan "Halo", Anda seharusnya dapat mengenali pengucapan yang membentuk bunyi huruf-huruf yang berbeda, lalu mengenali kata tersebut. Apakah ini mengingatkan Anda pada struktur berlapis jaringan saraf dalam (DNN)?

Meskipun jaringan saraf dalam telah membawa kemajuan besar dalam pengenalan suara, solusi saat ini masih memiliki kekurangan yang signifikan. Algoritmenya memiliki kinerja yang cukup baik, tetapi salah satu hambatan terbesar adalah data. Untuk melatih algoritme DL agar dapat melakukan pengenalan suara, Anda memerlukan berkas audio berlabel: rekaman

suara seseorang beserta transkrip ucapannya. Kecocokan alami untuk kebutuhan data ini berasal dari film: Anda dapat menggunakan audio sebagai input, dan subtitle sebagai label. Namun, tidak semua bahasa memiliki sejarah sinematografi yang kuat, dan beberapa bahasa hanya memiliki kumpulan data kecil yang tersedia.



Gambar 4.16 Algoritma pengenalan suara menerima rekaman audio dan menghasilkan transkripnya.



Gambar 4.17 Algoritma text-to-speech mengucapkan kalimat dengan lantang menggunakan suara yang terdengar alami.

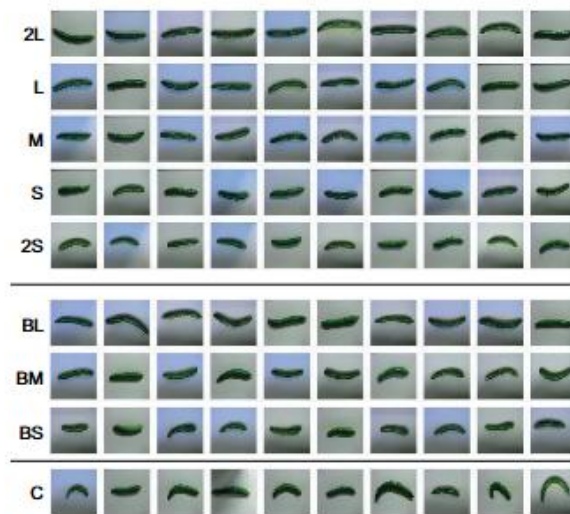
Selain itu, sebagian besar aktor memiliki pengucapan yang sempurna, sehingga sulit bagi algoritme untuk menggeneralisasikannya ke Bahasa Inggris yang kurang lancar. Text-to-speech adalah proses menghasilkan audio dari teks. Sebagaimana pengenalan suara adalah teknologi yang digunakan Alexa untuk memahami suara Anda, pembangkitan suara adalah teknologi yang digunakannya untuk berbicara kembali kepada Anda (gambar 4.17). Teknologi ini juga telah berkembang pesat berkat pembelajaran mendalam, dan suara yang dihasilkan oleh algoritma ini jauh lebih alami daripada suara robotik metalik yang biasa kita dengar di kereta bawah tanah. Namun, salah satu masalah utama algoritma baru ini adalah masih boros daya, dan sulit untuk menggunakan teknologi mutakhir pada perangkat seluler.

Tugas terakhir yang ingin kami sampaikan adalah deteksi kata pemicu. Teknologi ini membuka kunci ponsel Anda saat Anda mengucapkan, "Hai, Siri" atau "OK, Google." Dibandingkan dengan pengenalan ucapan umum, teknologi ini dioptimalkan untuk hanya mengenali satu ekspresi tersebut, sehingga dapat berjalan di latar belakang ponsel Anda tanpa menguras baterai.

Selain deteksi kata pemicu, pengenalan ucapan dan pembangkitan ucapan merupakan teknologi canggih yang sulit dikembangkan untuk organisasi kecil; bahkan raksasa teknologi seperti Google dan Facebook masih melakukan penelitian tentang teknologi tersebut. Untungnya, Anda akan melihat bahwa sebagian besar perusahaan ini menawarkan teknologi mereka sebagai layanan, dan itu hampir semua yang Anda butuhkan karena aplikasi tersebut cukup standar dan biasanya tidak memerlukan kustomisasi khusus.

4.8 OPTIMASI PERTANIAN DENGAN DEEP LEARNING: STUDI KASUS

Makoto Koike adalah seorang insinyur perangkat lunak yang berbasis di Jepang. Saat mulai membantu di kebun mentimun keluarganya, ia tak bisa tidak memperhatikan bahwa orang tuanya menghabiskan banyak waktu untuk memilah hasil panen berdasarkan kualitasnya. Pertanian Jepang sangat menjunjung tinggi kualitas produksi, yang berarti ibu Koike menghabiskan hingga delapan jam sehari untuk memilah mentimun berdasarkan bentuk, warna, dan jumlah durinya. Karena tidak ada klasifikasi standar untuk kualitas mentimun di Jepang, kebun Koike menerapkan pendekatan sembilan kelas, dari yang terbaik hingga yang kualitasnya paling rendah. Anda dapat melihat kelas-kelas ini pada Gambar 4.18.



Gambar 4.18 Contoh kelas mentimun. Mentimun dapat bervariasi bentuk dan ukurannya.

Koike memutuskan untuk mencoba mengotomatiskan proses ini agar keluarganya dapat fokus mengoptimalkan budidaya. Berkat latar belakangnya di bidang teknik otomotif, ia mampu merancang dan membangun komponen mekanis untuk mesin sortir otomatis berdasarkan desain sabuk konveyor konvensional.

Lengan otomatis (dikendalikan oleh komputer) mendorong setiap mentimun ke dalam salah satu wadah berdasarkan klasifikasi kualitasnya. Setelah mesin dibangun dan diuji, Koike harus mengimplementasikan algoritma kontrol untuk lengan-lengan tersebut. Dari sudut pandang implementasi, tugas tersulit yang harus diselesaikan adalah sebagai berikut:

1. Mengidentifikasi item yang berbeda saat bergerak melalui sabuk konveyor
2. Mengklasifikasikan setiap item berdasarkan bentuk, warna, dan tampilan permukaan
3. Mengkoordinasikan pergerakan lengan untuk mendorong setiap item ke wadah yang tepat

Koike dapat memanfaatkan keahliannya untuk menangani tugas 1 dan 3, yang termasuk dalam ranah rekayasa otomasi konvensional. Bahasa Indonesia: Dia kemudian menyadari bahwa tugas 2 adalah tugas klasifikasi objek, yang mana dia dapat menggunakan sejumlah besar penelitian dan kode sumber terbuka. Ini mengurangi upaya rekayasa perangkat lunak konvensional menjadi beberapa jam yang dibutuhkan untuk menyiapkan salah satu kerangka

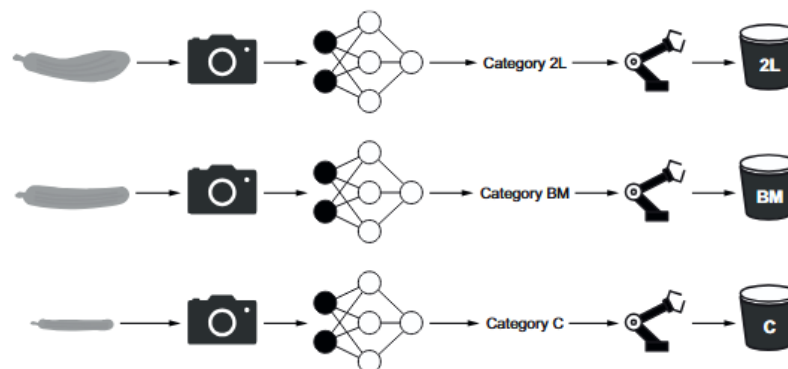
kerja yang tersedia secara bebas ini, dan dia berangkat ke tantangan nyata berikutnya, mengumpulkan data pelatihan.

Seperti yang sering terjadi dalam situasi khusus domain seperti ini, satu-satunya jalan untuk pengumpulan data adalah dengan mengandalkan manusia yang memiliki keahlian yang diperlukan. Koike mengumpulkan contoh (gambar) dan label saat ibunya dengan sabar mengklasifikasikan sekitar 8.000 mentimun menurut kriteria kualitas pertanian yang ketat. Dataset tersedia secara bebas di <https://github.com/workpiles/CUCUM-BER-9>.

Setelah data pelatihan yang cukup untuk prototipe dikumpulkan, yang tersisa hanyalah rekayasa perangkat lunak rutin: memasukkan prediksi model ke dalam sistem kontrol untuk ban berjalan. Gambar 4.19 mengilustrasikan prosesnya: kamera mengambil gambar mentimun yang kemudian diproses oleh jaringan saraf; Berdasarkan prediksi jaringan, lengan robot menempatkan mentimun di dalam ember yang sesuai dengan kelasnya. Seluruh proses berlangsung di ban berjalan kecil. Anda dapat melihat produk akhir pada Gambar 4.20.

Koike mengklaim akurasi 70% untuk keseluruhan sistem, meskipun akurasi yang diukur pada set uji seringkali mencapai lebih dari 95%. Hal ini menunjukkan bahwa data uji tidak secara akurat menangkap variabilitas sampel dalam situasi kehidupan nyata. Misalnya, beberapa bentuk atau warna mentimun mungkin kurang terwakili dalam set pelatihan. Biasanya, hal ini terjadi jika semua contoh dan label pelatihan dikumpulkan pada awal musim panen, sementara mentimun tumbuh lebih besar seiring bergantinya musim.

Untuk meningkatkan kinerja sistemnya, Koike memutuskan untuk mengadopsi pendekatan multikamera: kamera yang berbeda mengambil gambar mentimun dari berbagai sudut tetap, yang kemudian dimasukkan ke dalam jaringan saraf untuk mengklasifikasikan mentimun. Anda dapat melihat versi terbaru dari set data Koike pada Gambar 4.21.



Gambar 4.19 Langkah inti sistem Koike, dari mengambil gambar mentimun hingga penempatan akhirnya ke dalam ember yang tepat



Gambar 4.20 Bagian dari mesin sortir mentimun akhir, tempat kamera mengambil gambar mentimun dan komputer menguraikan prediksi, sebelum menggerakkan mentimun ke depan, dan meletakkannya di ember yang tepat.

Kisah Koike telah memukau banyak media berita, yang melaporkan kisah David versus Goliath ini, yang menunjukkan bahwa bahkan seorang insinyur pun dapat membangun aplikasi AI yang menyaingi aplikasi perusahaan besar. Banyak reporter ini tidak menyebutkan bahwa Koike membutuhkan waktu tiga bulan untuk mengumpulkan 8.000 sampel yang dibutuhkan untuk prototipe tersebut. Gambar yang ia ambil juga memiliki resolusi yang sangat rendah sehingga secara fisik mustahil bagi model untuk mengenali duri dan goresan. Di sisi lain, blog Google merilis versi ceritanya sendiri, yang menekankan peran pustaka sumber terbuka dan layanan cloud yang ditawarkannya untuk mempercepat pelatihan.



Gambar 4.21 Versi terbaru dari kumpulan data Makoto, yang dihasilkan dari setiap mentimun yang difoto dari tiga sudut berbeda.

1. Mengapa masalah ini cocok untuk pembelajaran mesin?
2. Pilihan dan kompromi utama apa yang memungkinkan Koike mengembangkan prototipe dengan cepat dengan investasi minimal?
3. Bagaimana Anda akan merancang pedoman untuk mengukur keberhasilan proyek ini?

DISKUSI KASUS

Kasus bisnis untuk mengotomatisasi proses produksi telah jelas sejak Revolusi Industri. Proyek ini merupakan contoh bagaimana AI membawa gelombang baru perbaikan dengan berfokus pada persepsi tingkat lanjut untuk robotika. Banyak pabrik dirancang agar robot industri dapat menyelesaikan tugas dengan kebutuhan fleksibilitas minimum: bayangkan ban berjalan yang membawa kaleng terbuka saat sedang diisi dengan produk. Dalam kasus seperti itu, algoritma yang mengendalikan pabrik tidak terlalu membutuhkan kamera atau sensor canggih lainnya, karena prosesnya ketat. Situasi dengan mentimun sangat berbeda, karena sulit untuk mengontrol bentuk mentimun (alam tidak menyukai standardisasi), dan Anda perlu memilahnya sendiri. Membuat aturan yang dikodekan secara keras itu sulit, dan oleh karena itu pembelajaran mesin (ML) sangat cocok untuk kemajuan pengambilan keputusan semacam ini.

Proyek Koike juga merupakan contoh relevan yang menjembatani teknologi abad ke-20 (otomatisasi industri) dengan AI dan perangkat lunak abad ke-21. Setelah mesin mekanik dibangun, algoritmanya menjadi aplikasi klasifikasi terawasi yang lazim di buku teks, yang telah ditulis oleh ribuan peneliti dalam ribuan makalah. Hal ini memungkinkan Koike untuk memanfaatkan teknik-teknik yang sudah dikenal dan bahkan menggunakan kembali kode sumber terbuka.

Meskipun kecil, proyek ini mencakup semua langkah dalam alur kerja untuk membangun aplikasi AI: mengenali nilai bisnis (menghemat waktu), mengumpulkan data pelatihan, dan memasukkan model ke dalam proses bisnis. Secara teknis, ini adalah contoh yang baik tentang apa yang terjadi ketika Anda menggunakan model mental yang tepat untuk suatu masalah: semuanya menjadi lebih mudah karena Anda dapat berkomunikasi dengan para insinyur secara lebih efektif, dan dapat memanfaatkan banyak teknologi sebelumnya. Setelah Koike menyadari bahwa inti dari alatnya adalah model klasifikasi objek, ia dapat langsung terjun dan mulai mengumpulkan data pelatihan berlabel, lalu menggunakan kode sumber terbuka dari internet.

Perhatikan bahwa membangun teknologi ini tidak sulit bagi Koike, karena ia dapat menggunakan penyedia cloud dan teknologi sumber terbuka. Yang membutuhkan waktu adalah pembuatan set data. Ini adalah contoh lain yang perlu digarisbawahi bahwa data adalah aset paling berharga saat membangun aplikasi AI. Hal ini khususnya berlaku untuk kasus ini; data yang dikumpulkan Koike memiliki beberapa properti yang berbeda dibandingkan dengan gambar yang ia kumpulkan dalam aplikasi sebenarnya, sebagaimana ditunjukkan oleh

perbedaan antara performa pada set pengujian dan di dunia nyata. Ini bisa menjadi aspek penting untuk dikerjakan jika proyek harus ditingkatkan lebih lanjut.

Terakhir, pertimbangkan masalah awal yang dihadapi Koike: ia kesulitan mengukur performa modelnya karena tampaknya performanya lebih baik pada contoh yang belum pernah dilihat sebelumnya (set pengujian) dibandingkan pada contoh yang telah dipelajarinya (set pelatihan). Hal ini, tentu saja, mencurigakan, dan mungkin karena mentimun dalam set pelatihannya diproduksi pada musim yang berbeda dari mentimun yang ada di set pelatihannya. Tantangan ini menyoroti pentingnya memilih data dengan cermat: set data besar saja tidak cukup; Anda juga membutuhkannya agar representatif terhadap kasus potensial yang akan Anda alami dalam aplikasi di dunia nyata.

Dalam hal akurasi minimum yang dapat diterima, Koike bisa saja menggunakan keputusan ibunya sebagai dasar kebenaran dan menghasilkan ambang batas yang masuk akal baginya. Namun, penting juga untuk berpikir kritis tentang arti sebuah kesalahan: berapa kerugian jika menempatkan mentimun di kotak yang salah? Karena kesalahan dalam aplikasi ini tidak akan membawa konsekuensi yang merusak (tidak seperti dalam kedokteran, misalnya), Koike bisa saja memutuskan untuk menerima hasil yang tidak terlalu menarik demi sepenuhnya melepaskan tugas berat ini dari ibunya yang malang.

- Pembelajaran mendalam adalah keluarga algoritma pembelajaran mesin yang mendukung sebagian besar aplikasi AI yang menangani gambar dan audio.
- Rahasia pembelajaran mendalam adalah kemampuannya untuk secara otomatis mengekstraksi fitur-fitur relevan dari data, sehingga menghemat waktu para insinyur dari pekerjaan yang panjang (atau bahkan mustahil).
- Pembelajaran transfer adalah teknik yang membantu mengurangi kebutuhan pembelajaran mendalam yang tak terpuaskan akan data pelatihan.
- Pengenalan wajah dan pembuatan konten otomatis adalah aplikasi cerdas yang memanfaatkan fleksibilitas model pembelajaran mendalam.
- Pembelajaran mendalam berfungsi dengan baik pada audio maupun gambar, dan merupakan metode tercanggih saat ini untuk mengenali dan menghasilkan ucapan.

BAB 5

AI UNTUK BAHASA ALAMI

Bab ini membahas

- Memahami tantangan utama pemrosesan bahasa alami
- Mengukur opini dari teks dengan analisis sentimen
- Mencari konten tekstual dengan kueri alami
- Membangun antarmuka percakapan (chatbot)
- Studi kasus: Penerjemahan bahasa berbantuan ML

Membuat komputer memahami bahasa telah menjadi impian para ilmuwan komputer sejak tahun 1950an. Hal ini mungkin karena sebagian besar pengetahuan dan budaya manusia dikodekan dalam bentuk tulisan dan penggunaan bahasa merupakan salah satu kemampuan paling kuat yang membedakan kita dari hewan. Istilah umum untuk teknik-teknik ini, yang mencakup pemahaman dan produksi bahasa, adalah pemrosesan bahasa alami (NLP).

Sayangnya, memberi mesin kemampuan untuk memahami dan memproduksi bahasa sama sulitnya dengan yang diinginkan: teknologi untuk mendukung interaksi komputer-manusia yang mulus seperti itu belum ada. Hal ini tidak menghentikan perusahaan, agensi, dan media untuk melebih-lebihkan potensi solusi menyeluruh. *Chatbot* adalah contoh utama dari fenomena ini, yang digembar-gemborkan sebagai agen otonom yang mampu berkiriman pesan dengan pelanggan, dengan mudah melengkapi atau menggantikan pekerja manusia untuk urusan kompleks seperti penjualan atau layanan purnajual. Hasilnya mengecewakan: pada tahun 2017, blog Silicon Valley, *The Information*, menerbitkan penelitiannya tentang kinerja asisten pribadi Facebook Messenger, M. Blog tersebut melaporkan bahwa M gagal menangani 70% permintaan pengguna. Seperti yang dapat Anda bayangkan, janji yang diingkari tidak menghasilkan apa-apa selain uang yang terbuang, waktu yang terbuang, dan frustrasi.

Jika pengantar ini tampak kurang menggembirakan, ada kabar baik. Pertama, teknologi berkembang pesat, dan cakrawala kemungkinan pun meluas. Kedua, Anda akan belajar cara menemukan dan mengembangkan proyek tersembunyi yang dapat memanfaatkan perangkat NLP yang lebih sederhana dan lebih matang. Peluang-peluang ini terlalu sering diremehkan di tengah gembar-gembor seputar NLP, tetapi dapat menjadi kisah sukses yang dahsyat bagi upaya AI Anda. Bab ini memiliki dua tujuan umum:

- Memberi Anda pengetahuan untuk menilai apakah suatu ide layak dengan teknologi terkini (dan sumber daya Anda)
- Mengajari Anda cara merencanakan upaya bahasa alami dengan tujuan memaksimalkan laba atas investasi. Ini akan memungkinkan Anda menjadi orang yang membuka wawasan dan menemukan peluang tersembunyi di organisasi Anda.

Kita akan mulai dengan menjelaskan mengapa bahasa sangat sulit bagi komputer, dan bagaimana pembelajaran mesin, sekali lagi, menjadi penyelamat selama beberapa tahun terakhir, memungkinkan kita membangun produk dan solusi yang luar biasa.

5.1 DAYA TARIK PEMAHAMAN BAHASA ALAMI

Mari kita akui. Cara kita mengonsumsi informasi di internet telah terjebak dalam kompromi yang tidak alami untuk sementara waktu: kita masih mengklik tombol dan memilih

menu tarik-turun, sementara di dunia nyata kita hanya menggunakan bahasa yang sederhana. Pemrosesan bahasa alami adalah teknologi yang memungkinkan kita menghilangkan hambatan ini dan memungkinkan orang untuk menggunakan dunia digital dengan cara yang sama seperti mereka mengalami dunia fisik.

Mari kita coba bayangkan bagaimana NLP dapat mengubah dunia di dunia situs web pialang properti yang kini sudah tidak asing lagi. Kita akan menempatkan diri kita pada posisi salah satu pengguna: seorang investor startup kaya dari San Francisco yang ingin membeli rumah. Ia menginginkan vila besar dengan setidaknya empat kamar, dekat dengan Jalan Raya 101, dengan pemandangan laut yang wajib. Saat ini, ketika ia mengunjungi situs web tersebut, ia dihadapkan dengan tampilan portal internet klasik yang biasa kita lihat: menu, beberapa kotak berisi harga, lokasi, kamar tidur, dan segudang informasi untuk setiap listing (gambar, deskripsi, dan sebagainya).

Bermain-main dengan semua tombol, pilihan, dan menu tarik-turun ini bisa terasa seperti berada di kokpit pesawat jet ketika yang Anda inginkan hanyalah jawaban sederhana untuk pertanyaan ini:

Apa saja vila di San Francisco yang dekat dengan jalan raya, memiliki pemandangan laut yang menakjubkan, dan setidaknya empat kamar tidur? Saya ingin mengunjungi vila-vila terbaik akhir pekan ini.

Setiap broker sungguhan dapat menjawab pertanyaan ini, tetapi kita tidak dapat memiliki manusia khusus untuk setiap orang yang mencari rumah di internet. Tetapi bagaimana jika kita bisa, berkat AI? Inilah janji pemahaman bahasa mesin: pengalaman superior bagi setiap orang yang berinteraksi dengan komputer, meningkatkan kepuasan dan efisiensi mereka, dalam skala besar. Anda sudah menyadari bahwa menyamai kemampuan manusia sepenuhnya dalam menggunakan bahasa berada di luar jangkauan teknologi saat ini. Jalan keluar dari tantangan ini adalah merancang solusi yang mengurangi kompleksitas agar sesuai dengan teknologi yang tersedia. Langkah pertama adalah memahami dan mengukur kompleksitas. Mari kita mulai.

5.2 MENGURAIKAN NLP: MENGUKUR KOMPLEKSITAS

Bahasa adalah salah satu kemampuan terpenting yang membedakan manusia dari hewan. Meskipun anjing peliharaan dapat menggunakan suara untuk berkomunikasi, kemampuan tersebut terbatas pada suasana hati yang terbatas (misalnya, rasa sakit, kesedihan, dan kemarahan). Di sisi lain, manusia dapat menggunakan kata-kata untuk mengomunikasikan (hampir) apa saja: mulai dari mengungkapkan perasaan, cinta dalam hidup kita, hingga memesan burger keju ganda. Jelas, spektrum kompleksitas berkisar dari kemampuan sederhana seekor anjing hingga nuansa yang dapat diungkapkan seorang penyair dengan bahasa.

Dalam teknologi, kompleksitas adalah musuh: semakin kompleks suatu tugas, semakin besar kemungkinan kita gagal dan menghasilkan produk yang kurang memuaskan, terutama

dengan teknologi yang belum matang seperti NLP. Bagaimana kita mengukur kompleksitas dalam ranah tugas NLP? Berdasarkan pengalaman kami, terdapat dua metrik kunci kualitatif:

- ✚ **Lebar** — Keragaman topik atau domain bahasa yang dibutuhkan untuk melakukan suatu tugas. Seberapa besar kosakata yang harus dikuasai algoritma kita?
- ✚ **Kedalaman** — Seberapa canggih hasil algoritma kita? Berapa banyak hal yang harus dipahami atau dilakukan?

Tabel 5.1 mencantumkan contoh tugas NLP dan menunjukkan lebar dan kedalamannya.

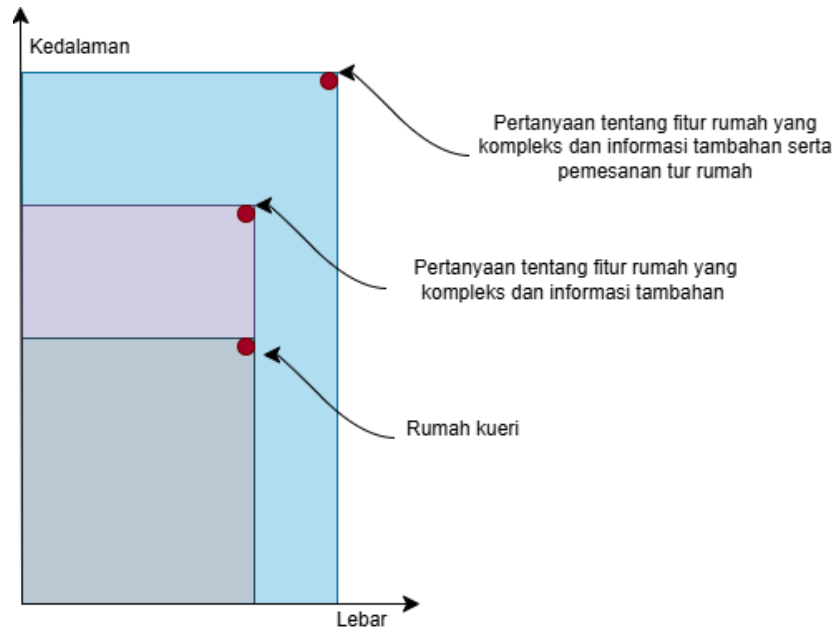
Tabel 5.1 Tugas NLP dan evaluasi lebar dan kedalamannya

Tugas	Lebar	Kedalaman
Pahami apakah ulasan produk Amazon positif atau negatif.	Tinggi: Ulasan dapat mengenai berbagai macam produk dan atribut.	Rendah: Kami hanya memiliki dua kemungkinan keluaran (sentimen positif atau negatif).
Bantu wisatawan memesan dan mengelola tiket pesawat mereka.	Rendah: Teks yang akan kita berurusan dengan relatif terbatas (penerbangan, pemesanan).	Tinggi: Agen harus memahami berbagai macam pilihan, mulai dari memeriksa harga hingga mengubah tiket, menangani keluhan, menyediakan bantuan bandara, dan seterusnya.
Bantu manajer merencanakan waktu dan janji temu mereka (bot sekretaris).	Tinggi: Seorang sekretaris perlu memahami pertanyaan tentang banyak domain-penjadwalan, janji temu, penagihan, perjalanan, dan seterusnya.	Tinggi: Dalam setiap topik ini, seorang bot sekretaris harus memiliki pemahaman yang kuat dan melakukan tindakan yang kompleks, seperti mengatur rapat pada waktu yang nyaman sambil mengatur transportasi.

Perhatikan bahwa karena lebar dan kedalaman berkontribusi pada kompleksitas suatu tugas, kita dapat mengatakan bahwa kompleksitas keseluruhan tugas sebanding dengan hasil perkaliannya:

$$Complexity = Width \times Depth$$

Jika kita menggambar grafik dengan kedua metrik ini pada sumbunya, kita mendapatkan cara yang efektif untuk memvisualisasikan kompleksitas tugas NLP. Tugas yang membentuk persegi panjang dengan area yang luas lebih kompleks daripada tugas dengan area yang lebih kecil, seperti yang ditunjukkan pada Gambar 5.1.



Gambar 5.1 Lebar dan kedalaman berbagai tugas NLP yang dapat diberikan situs web real estat melalui chatbot

Mari kita kembali ke contoh chatbot untuk situs web real estat kita. Tindakan sederhana yang dapat dilakukan chatbot ini adalah memahami kueri dasar pengguna, yang pada dasarnya menggantikan antarmuka pencarian konvensional. Berikut adalah contoh kueri:

Tunjukkan rumah dengan lebih dari empat kamar tidur di pusat kota San Francisco yang harganya kurang dari Rp. 20 miliar.

Dalam kasus ini, chatbot perlu melakukan satu tindakan membuat kueri ke basis data yang terkait dengan satu domain: properti rumah. Kita dapat meminta lebih banyak dari chatbot kita, dan memperluas fungsinya untuk mengekstrak informasi meta dari basis data ini atau menjawab kueri yang lebih kompleks, misalnya:

Berapa harga rata-rata rumah di lingkungan Presidio, dengan pemandangan laut, empat kamar tidur, dan berjarak kurang dari 10 menit dari Jalan Raya 101?

Ini tidak akan menambah lebarnya, karena domain bahasanya masih terbatas pada rumah, tetapi kedalamannya jauh lebih tinggi, karena jumlah kemungkinan tindakan yang harus dapat ditangani bot telah meningkat. Langkah terakhir bisa jadi adalah mengelola pemesanan tur rumah juga. Ini akan menambah kedalaman lagi seiring bertambahnya jumlah tindakan. Sekarang, lebarnya juga bertambah, karena bot perlu memahami ekspresi bahasa baru yang berkaitan dengan pemesanan.

Perlu diingat bahwa dua tugas yang sangat berbeda dapat memiliki kompleksitas yang serupa. Mari kita beralih sejenak dan melihat dua contoh umum NLP yang mungkin sudah Anda kenal:

- ❖ *Analisis sentimen pada tweet*—Untuk tugas ini, model membaca teks pendek dan harus memutuskan apakah teks tersebut mengekspresikan sentimen positif atau negatif. Dengan kata lain, analisis sentimen bertujuan untuk mencari tahu apa yang dipikirkan penulis tentang subjek utama teks tersebut. Contoh: "Saya suka film baru Spielberg" bersifat positif. "Kamera ponsel baru ini jelek" bersifat negatif.
- ❖ *Klasifikasi tiket dukungan pelanggan* — Untuk tugas ini, model membaca keluhan dari pelanggan dan mengelompokkannya ke salah satu dari beberapa kelas penyebab potensial. Contoh: "Kata sandi salah" adalah masalah autentikasi, dan "Situs tidak dapat dimuat" adalah masalah konektivitas.

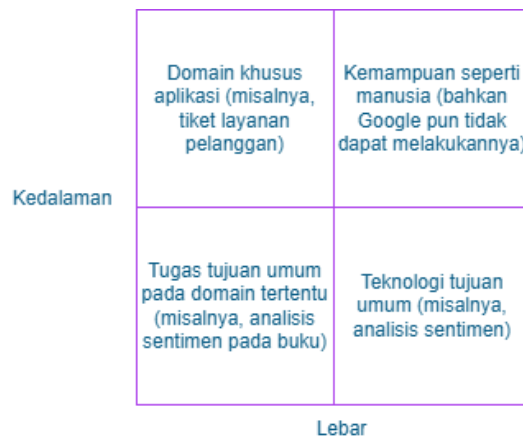
Tugas pertama memiliki lebar yang tinggi, karena dapat menangani banyak subjek (orang menggunakan Twitter untuk membicarakan segala hal), tetapi kedalamannya rendah karena hanya memiliki dua kemungkinan hasil: bendera ulasan positif atau negatif.

Tugas kedua, di sisi lain, memiliki lebar yang rendah karena perlu menangani tiket yang semuanya entah bagaimana terkait dengan domain yang sama (hal-hal TI tidak berfungsi), tetapi memiliki kedalaman yang lebih tinggi karena perlu menempatkan tiket ke dalam beberapa kategori. Pada grafik lebar/kedalaman kita, kedua tugas tersebut akan terlihat seperti Gambar 5.2.

Bagaimana kita menerjemahkan wawasan ini menjadi informasi yang dapat ditindaklanjuti? Melihat tugas NLP dalam hal lebar dan kedalaman membantu menilai kelayakannya. Secara umum, Anda ingin menjaga keduanya seminimal mungkin dengan mulai merancang aplikasi NLP Anda dari kombinasi lebar/kedalaman terendah. Anda akan mengembangkannya sesuai kebutuhan Anda, dan bergantung pada performa yang dapat Anda capai. Pada Gambar 5.3, Anda dapat melihat representasi kelayakan suatu aplikasi bergantung pada kedalaman dan lebarnya. Di bab selanjutnya dari bab ini, kita akan menggunakan model mental ini sebagai panduan kita dalam menambahkan kapabilitas NLP ke situs listing rumah kita.



Gambar 5.2 Dua tugas pada diagram lebar/kedalaman: mengkategorikan tiket dukungan pelanggan (kedalaman tinggi, lebar rendah), dan analisis sentimen dari tweet (lebar tinggi, kedalaman rendah)



Gambar 5.3 Aplikasi NLP berdasarkan lebar dan kedalamannya. Semakin dekat suatu aplikasi ke pojok kanan atas, semakin kecil kemungkinannya untuk diterapkan dengan teknologi saat ini.

5.3 MENAMBAHKAN KAPABILITAS NLP KE ORGANISASI ANDA

Mari kita kembalikan situs broker real estat kita dari bagian yang kita tinggalkan di bab sebelumnya. Pengklasifikasi gambar berbasis DL kami telah sangat meningkatkan pengalaman pengguna, dan jumlah pengunjung bulanan berkembang pesat. Pelanggan mengunjungi situs web kami, dapat dengan cepat menemukan detail yang mereka minati melalui gambar setiap listing, dan melanjutkan proses dengan meminta informasi lebih lanjut tentang rumah favorit mereka. Saat ini, seseorang yang tertarik dengan listing tertentu memiliki dua pilihan untuk mencari tahu lebih lanjut:

1. Luangkan waktu yang lama untuk membaca deskripsi yang ditinggalkan oleh pemilik rumah dan semua ulasan dari orang lain yang pernah mengunjunginya (anggap saja platform real estat kita memiliki ulasan tentang rumah).
2. Hubungi layanan pelanggan, yang akan menghubungkan pelanggan dengan broker yang menangani rumah tersebut.

Pilihan pertama bukanlah pengalaman pengguna yang baik, dan yang kedua sangat mahal untuk bisnis Anda. Kami telah membayangkan sebuah program komputer mirip manusia (sering disebut bot, chatbot, atau agen) yang dapat sepenuhnya mengambil alih peran broker rumah, tetapi kami juga memahami bahwa penerapan NLP semacam itu kemungkinan besar tidak akan berfungsi dengan teknologi saat ini. Kami dapat menyebut aplikasi canggih ini sebagai brokerbot: ia memiliki lebar dan kedalaman yang tinggi.

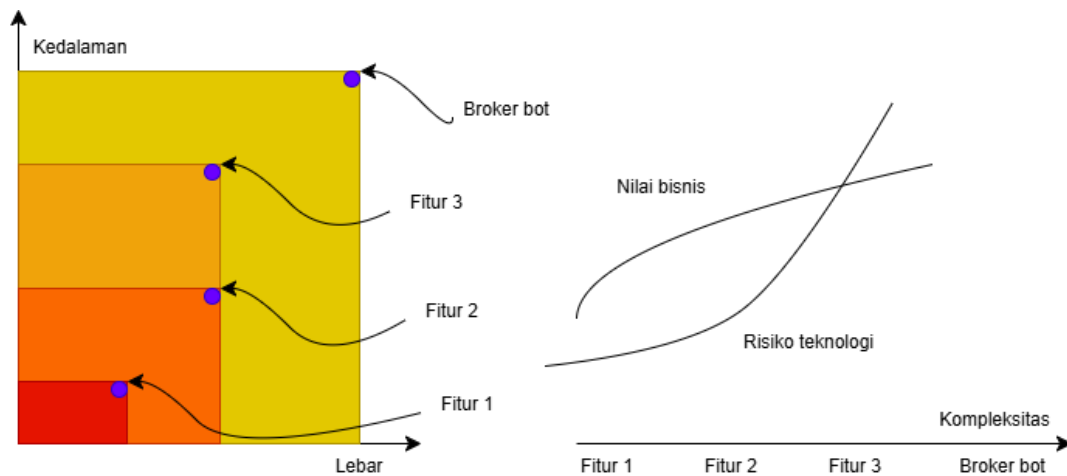
Mari kita mundur selangkah dan melihat tujuan bisnis kita. Tujuan bisnisnya adalah memungkinkan pelanggan menemukan informasi yang mereka butuhkan tentang setiap listing dengan cepat. Jika Anda melihatnya dari sudut pandang ini, brokerbot agak berlebihan: kita masih bisa sedikit demi sedikit menuju pengalaman pengguna yang lebih baik dan pencarian yang lebih cepat dengan menggunakan teknologi yang lebih sederhana.

Kita tahu bahwa pengguna menghabiskan banyak waktu untuk membaca ulasan yang ditulis oleh calon pembeli rumah lainnya. Ulasan-ulasan ini sarat dengan informasi yang

seringkali tidak disertakan dalam deskripsi rumah, seperti opini langsung tentang perlengkapan pencahayaan, lingkungan sekitar, dan sebagainya. Kami ingin menggunakan informasi ini untuk membantu pengguna lain membuat keputusan terbaik, tetapi hanya sedikit yang mau meluangkan waktu untuk membaca semua materi ini. Yang akan kita lakukan di bab-bab selanjutnya adalah secara bertahap menghadirkan fitur-fitur NLP yang lebih kompleks untuk situs web kita, dengan memanfaatkan konten bahasa alami utama yang kita miliki: ulasan pengguna.

Anda akan menemukan bahwa bahkan fungsionalitas NLP dasar pun dapat bermanfaat bagi pengguna, tanpa harus merancang fitur-fitur super kompleks yang mirip manusia yang kemungkinan besar akan mengecewakan dan membuat pengguna frustrasi. Seperti yang Anda lihat pada Gambar 5.4, dalam bidang yang kompleks dan belum matang seperti pemrosesan bahasa alami, meningkatkan kompleksitas produk Anda secara eksponensial meningkatkan risiko teknologi, sementara nilai bisnis akan stagnan setelah beberapa saat.

Langkah pertama untuk menambahkan kapabilitas NLP ke situs web kita seharusnya menjadi hal termudah yang dapat kita bangun yang tetap memberikan nilai bagi pengguna. Dalam kasus kita, hal tersebut dapat berupa pengembangan sistem yang mengklasifikasikan setiap ulasan rumah sebagai positif atau negatif, sehingga memudahkan pengguna baru untuk memahami seberapa bagus rumah yang terdaftar tersebut sebenarnya.



Gambar 5.4 Berbagai fitur dengan kompleksitas yang meningkat membawa risiko teknologi yang jauh lebih tinggi secara eksponensial, sementara nilai bisnis menurun dengan cepat.

Pada dasarnya, kita ingin beralih dari bagian ulasan yang tidak terorganisir seperti pada Tabel 5.2 ke tata letak yang jauh lebih mudah dipahami dan informatif, seperti pada Tabel 5.3.

Tabel 5.2 Daftar ulasan positif dan negatif yang digabungkan

Ulasan
Lokasi bagus!
Rumah itu sangat tua dan bobrok.
Pemandangan dari atap sungguh menakjubkan.

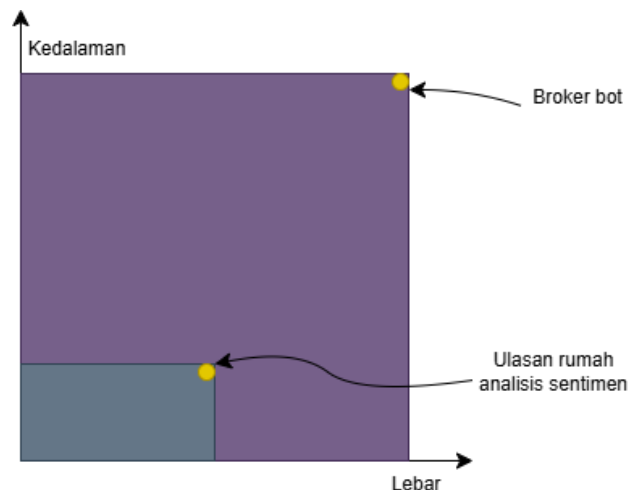
Saya tidak akan pernah tinggal di lantai dasar di lingkungan yang begitu berbahaya.

Tabel 5.3 Tabel terorganisir ulasan positif dan negatif

Apa yang disukai orang	Apa yang tidak disukai orang
Lokasi bagus!	Rumah itu sangat tua dan bobrok.
Pemandangan dari atap sungguh menakjubkan.	Saya tidak akan pernah tinggal di lantai dasar di lingkungan yang begitu berbahaya.

Mari kita lihat bagaimana analisis sentimen berjalan dalam hal kedalaman dan lebar. Ingat bahwa kedalaman suatu aplikasi berkaitan dengan seberapa kompleks kapabilitas model yang dibutuhkan. Dalam kasus ini, satu-satunya hal yang kita minta chatbot kita klasifikasikan adalah ulasan ke dalam salah satu dari dua kelas (positif atau negatif), sehingga kedalamannya rendah. Lebar mengukur jumlah kosakata yang harus "diketahui" oleh model (jumlah kata yang harus dipahami untuk mengklasifikasikan ulasan dengan benar).

Dalam hal ini, jenis kata yang dibutuhkan terbatas pada kata-kata positif seperti "indah", "aman", "menakjubkan", dan kata-kata negatif seperti "usang", "mengerikan", "berbahaya". Kata-kata ini merupakan gabungan dari kata-kata positif dan negatif umum seperti "indah" dan "jelek", bersama dengan beberapa istilah khusus domain seperti "luas" dan "usang" yang spesifik untuk domain rumah. Oleh karena itu, lebarnya sedikit lebih tinggi daripada tugas analisis sentimen standar, tetapi tidak terlalu signifikan. Gambar 5.5 membandingkan analisis sentimen dengan broker-bot pada diagram lebar/kedalaman yang diperkenalkan pada Gambar 5.1.



Gambar 5.5 Analisis sentimen versus brokerbot pada diagram lebar/kedalaman. Analisis sentimen memiliki kedalaman dan lebar yang lebih rendah karena hasil akhirnya sempit dan hanya membahas kata-kata yang mengungkapkan opini.

Seperti yang Anda lihat, area ini jauh lebih kecil dibandingkan dengan broker-bot, sehingga jauh lebih mudah dibangun sekaligus memberikan peningkatan signifikan pada pengalaman pengguna. Analisis sentimen merupakan aplikasi NLP yang klasik dan penting sehingga perlu

dibahas lebih detail. Kami juga akan menggunakannya sebagai alat untuk menjelaskan cara kerja internal algoritma NLP di bagian selanjutnya.

Analisis Sentimen

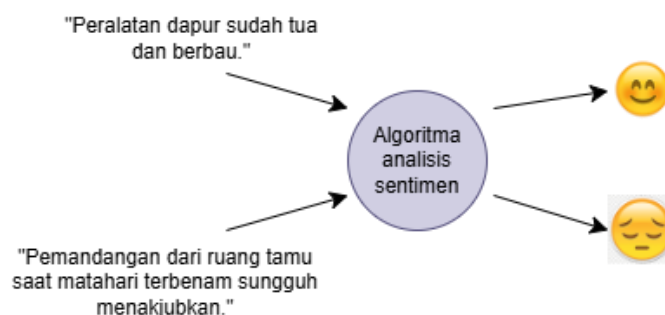
Sejumlah besar konten tekstual yang dihasilkan oleh pengguna di internet berkaitan dengan opini dan perasaan. Misalnya, pikirkan ulasan pengguna di Amazon, balasan Twitter, atau bahkan artikel blog. Algoritma analisis sentimen mencerna semua teks bebas ini, dan mencari tahu apakah penulis memiliki opini positif, negatif, atau netral tentang subjek tersebut.

Analisis sentimen adalah tugas penting, dan dapat digunakan untuk lebih dari sekadar menyortir ulasan rumah. Anda dapat memantau hasil kampanye pemasaran Anda secara real-time dengan mengikuti umpan dari media sosial, atau bahkan mengukur reaksi terhadap presentasi individu selama konferensi. Perusahaan perdagangan Wall Street secara rutin menggunakan analisis sentimen untuk mengukur apa yang dipikirkan publik tentang suatu perusahaan dan untuk menginformasikan keputusan perdagangan mereka.

Tapi mari kita kembali ke situs web daftar real estat kita. Berikut kutipan dari beberapa ulasan yang diposting oleh pengguna:

- Peralatan dapurnya tua dan berbau.
- Pemandangan dari ruang tamu saat matahari terbenam sungguh menakjubkan.

Yang ingin kami lakukan adalah membangun algoritma yang dapat memberi label positif atau negatif pada setiap kalimat, seperti pada gambar 5.6.



Gambar 5.6 Konsep dasar analisis sentimen

Anggaplah kita tidak memahami pembelajaran mesin, dan kita perlu menjelaskan kepada komputer cara menilai kalimat-kalimat ini dengan mengembangkan sebuah algoritma. Aturan praktis yang baik adalah dengan melihat kata-kata tertentu yang menunjukkan makna, seperti "mengerikan" dan "menakjubkan". Faktanya, beginilah cara kerja algoritma paling awal untuk analisis sentimen: para peneliti dengan susah payah membangun kamus kata-kata penting, dan memberi label masing-masing sebagai positif, negatif, atau netral. Misalnya, kamus kata-sentimen tersebut mungkin terlihat seperti ini:

- *Senang* — Positif
- *Dibunuh* — Negatif
- *Berteriak* — Negatif
- *Meja* — Netral

Setelah Anda memiliki glosarium emosi seperti itu, Anda dapat mengklasifikasikan setiap kalimat dengan menghitung jumlah kata positif dan negatif dan mendapatkan skor akhir untuk kalimat tersebut.

Pendekatan sederhana ini memiliki banyak masalah. Bahasa sangatlah kompleks, dan kita menggunakannya dengan cara yang sangat berbeda dari orang ke orang; Kata yang sama dapat digunakan dengan cara yang berbeda untuk menyampaikan pesan yang sepenuhnya berlawanan. Katakanlah Anda mencantumkan "bagus" sebagai kata bernada positif, dan salah satu pengulas menulis ini:

Akan lebih baik jika mereka menghapus permadani itu sepenuhnya.

Meskipun kalimat tersebut secara keseluruhan memiliki opini negatif tentang rumah tersebut, sistem naif kita akan menganggapnya positif, karena konotasi positif dari "bagus". Mungkin Anda dapat mencoba memperbaiki sistem ini dengan menambahkan aturan yang lebih kompleks, seperti ini:

'Akan menjadi [KATA POSITIF] jika ..' => negatif

Meskipun aturan ini akan berfungsi pada cuplikan sebelumnya, aturan ini masih mudah untuk dikelabui. Misalnya, kalimat berikut sebenarnya sangat positif, tetapi akan digolongkan sebagai opini negatif:

Akan lebih baik jika saya bisa segera pindah ke tempat ini!

Haruskah kita terus menambahkan aturan yang sudah dikodekan? Permainan ini sudah menjadi rumit (dan membosankan), namun masih mudah untuk mengelabui sistem kita. Perhatikan juga bahwa kita masih bermain hanya dengan beberapa kata; Kita bahkan belum mulai memetakan lautan kosakata bahasa Inggris yang luas. Kalaupun kita berhasil sampai ke akar-akarnya, kita hampir perlu memulai dari awal lagi untuk bahasa-bahasa lain.

Sekarang, Anda mungkin mulai berpikir bahwa pendekatan ini pasti akan gagal. Jika Anda familier dengan tema utama buku ini, Anda mungkin berpikir bahwa kami memiliki sesuatu yang jauh lebih baik untuk Anda, dan Anda benar. Mungkin sebuah mesin dapat mengetahui cara melakukan tugas itu sendiri, tanpa kita menjelaskannya.

Kita dapat mencoba melihat analisis sentimen dengan cara yang sama seperti kita melihat masalah ML lainnya di bab-bab sebelumnya. Diberikan sebuah teks, kita ingin memutuskan apakah teks tersebut termasuk dalam kelas "positif" atau "negatif": sebuah masalah klasifikasi. Kelas-kelas (label) akan berupa "positif" atau "negatif", dan fiturnya akan berupa kata-kata dalam kalimat. Alih-alih membuat aturan secara manual untuk mengklasifikasikan setiap kata sebagai positif atau negatif, kita dapat melatih model ML untuk melakukan pekerjaan itu untuk kita. Ternyata, ada atau tidaknya kata-kata tertentu dalam teks sudah cukup bagi model untuk mengetahui apakah opini tersebut negatif atau positif.

Pendekatan ini sederhana namun efektif. Salah satu orang pertama yang mencetuskan ide ini adalah Paul Graham, pendiri legendaris akselerator startup paling sukses di Silicon Valley, Y Combinator. Ia memikirkan pendekatan ini saat mencoba memperbaiki apa yang ia anggap sebagai salah satu ancaman terbesar bagi internet awal: email spam. Saat itu, kotak masuk email Anda akan dipenuhi iklan berbagai macam penipuan, mulai dari hadiah lotre palsu hingga Viagra palsu. Pada tahun 2002, Graham menemukan metode baru yang ia jelaskan dalam esainya "A Plan for Spam" (www.paulgraham.com/spam.html).

Rencana Graham dimulai dengan kesadaran bahwa sebagian besar pengembang akan tertarik untuk memprogram aturan eksplisit guna mendeteksi email spam. Di masa-masa awal internet, dengan menggunakan aturan eksplisit, tidak sulit untuk mengidentifikasi hampir 80% spam dengan tepat. Namun, menangkap persen terakhir spammer ternyata sangat berbeda. Di sisi lain, pendekatan ML yang sederhana lebih efektif sejak awal. Konsepnya cukup sederhana: biarkan algoritma pembelajaran mesin (ML) melihat sekumpulan email spam dan bukan spam, lalu biarkan ia mencari tahu sendiri kata-kata mana yang paling mungkin mengindikasikan spam.

Dengan upaya terbatas, algoritma Graham mampu mengalahkan aturan kode tangan paling rumit yang dikodekan dengan susah payah oleh para programmer ahli. Algoritma ini juga menemukan korelasi mengejutkan antara kata-kata aneh dan email spam yang tak terpikirkan oleh siapa pun: ternyata ff0000, kode warna untuk "merah" yang digunakan di web, sama efektifnya sebagai indikator spam seperti halnya kata-kata pornografi.

Pengalaman Graham memberikan dua hal menarik. Pertama, pada tahun 2002, Graham mempertimbangkan penyaringan spam otonom sebagai AI. Saat ini, kita menganggap fitur ini biasa saja dan menganggapnya kurang canggih dibandingkan mesin pembuat kopi. Selain itu, kami tidak akan membahas detail pengklasifikasi spesifik yang digunakan Graham (sebagai informasi, namanya Naive Bayes), tetapi mungkin ini salah satu algoritma paling sederhana yang dapat dipikirkan oleh seorang matematikawan. Namun, algoritma ini tetap mampu mengungguli Graham dan beberapa insinyur terpintar di dunia. Belajar dari pengalaman mengalahkan aturan yang dikodekan secara manual, sekali lagi.

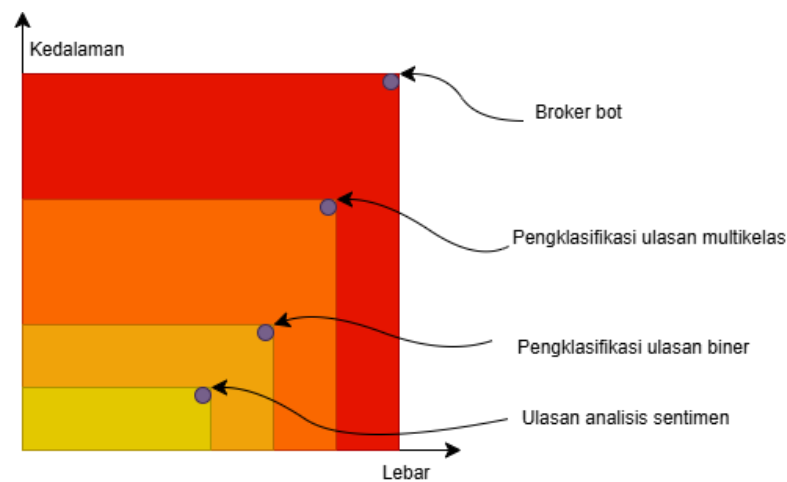
Dari Analisis Sentimen ke Klasifikasi Teks

Analisis sentimen telah meningkatkan pengalaman pengguna kami di situs, tetapi kami ingin mengambil beberapa langkah lebih lanjut menuju brokerbot. Kami memutuskan untuk mengatasi masalah lain: ulasan bertele-tele dan membahas berbagai aspek tentang rumah, mulai dari pencahayaan di ruang tamu hingga lingkungan sekitar. Calon pembeli yang sangat tertarik dengan topik tertentu harus menelusuri semua ulasan untuk menemukan ulasan yang membahas topik tersebut. Menandai setiap ulasan berdasarkan topiknya akan menjadi fitur luar biasa yang memungkinkan calon pembeli menemukan informasi yang mereka inginkan dengan cepat.

Alih-alih hanya memiliki dua kelas (positif dan negatif) seperti dalam analisis sentimen, kini kami memiliki lebih banyak kelas, karena ulasan dapat mencakup furnitur, pemandangan, prestise lingkungan sekitar, dan sebagainya. Tugas ini biasanya disebut klasifikasi teks atau dokumen.

Apakah klasifikasi dokumen lebih kompleks atau kurang kompleks daripada analisis sentimen? Sekali lagi, mari kita gunakan diagram lebar/kedalaman untuk memvisualisasikan perbedaannya. Lebar tugas lebih tinggi, karena model klasifikasi dokumen perlu memahami kosakata yang lebih luas. Istilah seperti "rak", "mesin pencuci piring", atau "kebun buah" dapat diabaikan dalam analisis sentimen, tetapi jelas penting untuk klasifikasi. Kedalaman penerapan NLP mengukur seberapa kompleks tugas tersebut.

Dalam kasus analisis sentimen, kami hanya ingin menempatkan ulasan kami dalam salah satu dari dua kategori. Untuk klasifikasi teks, kami dapat memiliki kategori sebanyak kategori yang ingin kami bagi ulasan kami; misalnya, ulasan tentang ruang tamu, kerusakan rumah, lingkungan sekitar, dan sebagainya. Semakin banyak kelas yang kami putuskan untuk digunakan, semakin mendalam modelnya, seperti yang ditunjukkan pada Gambar 5.7.



Gambar 5.7 Membandingkan brokerbot dengan aplikasi NLP lain yang lebih sederhana yang telah kita kembangkan dalam bab ini

Model ML dasar biasanya cukup baik untuk analisis sentimen, tetapi mulai mengalami masalah dengan tugas yang lebih kompleks seperti klasifikasi, terutama karena model tersebut mengabaikan arti kata, dan hanya menghitung seberapa sering kata tersebut muncul dalam dataset berlabel. Ini berarti pengklasifikasi tidak memiliki pemahaman tentang fakta bahwa "mengagumkan" dan "menakjubkan" memiliki arti yang serupa. Mengapa ini menjadi masalah? Misalkan Anda memiliki 20 contoh pelatihan dengan kata "mengagumkan" tetapi tidak ada yang menggunakan kata "menakjubkan".

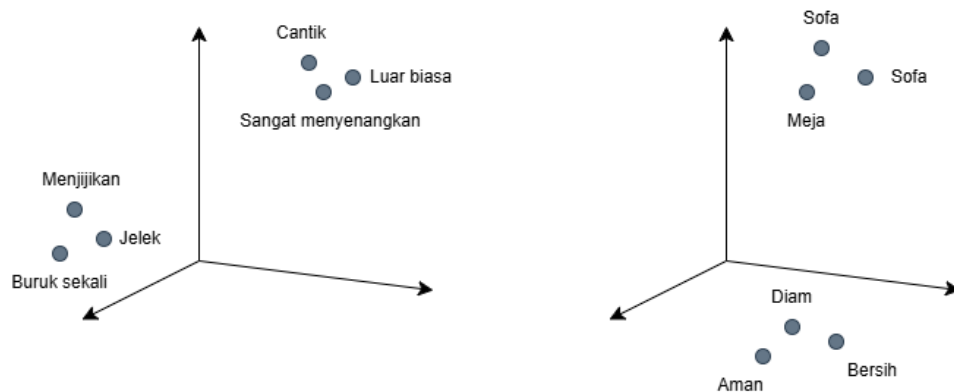
Model yang tidak memiliki pengetahuan tentang arti kata akan belajar mengklasifikasikan setiap ulasan dengan kata "mengagumkan" sebagai contoh positif, tetapi tidak akan tahu apa yang harus dilakukan dengan kata "menakjubkan". Jika sebuah model dapat memahami arti kata-kata, cukup baginya untuk melihat beberapa contoh dengan beberapa kata positif untuk memperluas daya klasifikasinya ke semua kata positif (artinya, model akan menafsirkan "menakjubkan", "luar biasa", dan "cantik" dengan benar meskipun contoh pelatihan hanya berisi "luar biasa").

Variabilitas kata yang tinggi menciptakan masalah yang serupa dengan yang kita bahas di Bab 4 dengan gambar. Ketika kita mulai membahas visi komputer, masalah utama kita adalah gambar terbuat dari jutaan piksel, dan sulit bagi kita untuk menemukan cara memecah kompleksitas data ini menjadi elemen yang lebih sederhana, seperti keberadaan bentuk dasar. Solusi kami untuk masalah ini adalah menggunakan pembelajaran mendalam, kelas algoritma khusus yang mampu mengekstraksi fitur kompleks secara otonom (dalam kasus gambar, dari informasi dasar seperti sudut dan tepi, hingga fitur kompleks seperti ciri wajah untuk pengenalan wajah). Anda juga melihat cara menggunakan algoritma ini untuk membangun embedding, ringkasan konten gambar yang ramah komputer.

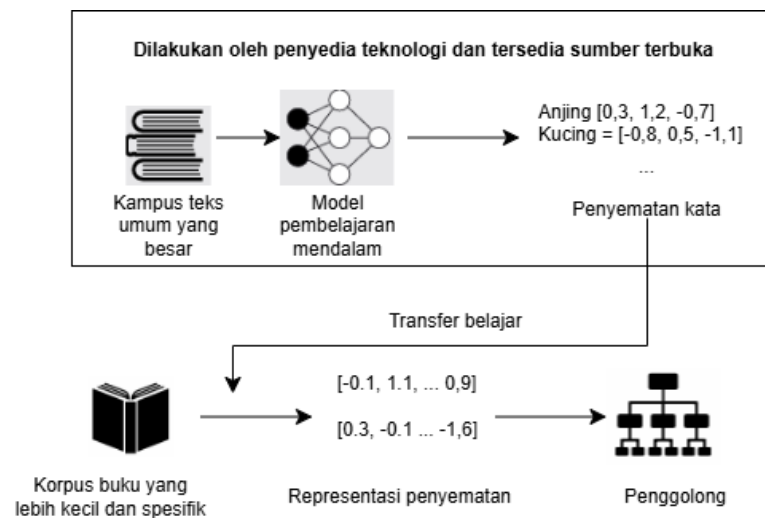
Beruntung bagi kita, pembelajaran mendalam juga berfungsi pada teks dan dapat membantu kita membangun fitur yang lebih canggih daripada sekadar keberadaan kata-kata tertentu dalam sebuah kalimat. Dengan cara yang sama seperti kita menggunakan embedding untuk mengubah piksel gambar mentah menjadi informasi yang bermakna (misalnya, hidung mancung atau telinga besar), kita dapat menggunakannya untuk mengubah kata menjadi serangkaian angka yang merepresentasikan fitur kata abstrak, seperti "kepositifan" dan "kekuatan". Sebagaimana dua gambar dengan objek serupa memiliki embedding yang serupa, kata-kata serupa juga akan memiliki embedding yang serupa (artinya, kata-kata serupa seperti "mengagumkan" dan "luar biasa" akan ditransformasikan menjadi serangkaian angka yang serupa).

Jika kita menginterpretasikan angka-angka dari embedding kata sebagai koordinat dalam ruang, kita dapat memahami secara visual kekuatan representasi ini. Misalnya, pada gambar 5.8, Anda dapat melihat kata-kata yang direpresentasikan pada bidang 3-D. Pada bidang kiri, Anda dapat melihat bahwa kata-kata positif berdekatan satu sama lain, dan kata-kata tersebut berjauhan dari kata-kata negatif (yang, di sisi lain, berdekatan satu sama lain). Hal yang sama berlaku untuk diagram di sebelah kanan: objek yang berkaitan dengan furnitur di ruang tamu cenderung mengelompok bersama, dan terletak jauh dari kata sifat yang berkaitan dengan lingkungan sekitar.

Dalam pengaturan ini, kedekatan berarti bahwa penempatan kata dari dua kata yang berdekatan serupa, yang memungkinkan pengklasifikasi dokumen teks kita untuk sangat menyederhanakan pekerjaannya: sekarang dapat mengabaikan perbedaan antara kata-kata yang serupa seperti "menakjubkan," "luar biasa," atau "cantik," dan menganggap semuanya sebagai angka serupa yang mewakili perasaan "positif".



Gambar 5.8 Pada penempatan kata yang direpresentasikan pada bidang 3-D, kata positif dan negatif berada di sebelah kiri, sedangkan kata yang berkaitan dengan lingkungan sekitar dan ruang tamu berada di sebelah kanan.



Gambar 5.9 Alur kerja NLP yang umum: Untuk tiga langkah pertama (dalam kotak abu-abu), Anda dapat memanfaatkan model dan embedding sumber terbuka. Tiga langkah terakhir melibatkan kumpulan data spesifik domain.

Sama seperti visi komputer, aplikasi NLP juga dapat memanfaatkan pembelajaran transfer untuk mengurangi jumlah data pelatihan yang dibutuhkan. Seperti yang Anda lihat di bab 4, pembelajaran transfer adalah cara untuk "memanaskan" model dengan sejumlah besar data generik yang mudah didapat sebelum melatihnya dengan menggunakan data pelatihan khusus domain, yang seringkali lebih sulit didapat. Layaknya aplikasi visi komputer, di mana model dapat mempelajari bentuk dan pola sederhana, model NLP akan mempelajari kosakata dan tata bahasa dasar, yang berguna untuk hampir semua hal ("indah" memiliki arti yang sama, baik untuk rumah maupun hal lainnya).

Komunitas riset NLP telah merilis beberapa model pembelajaran berbasis pembelajaran (DL) yang dilatih pada basis data teks yang sangat besar sehingga dapat memahami dan membuat embedding kosakata yang hampir lengkap. Gambar 5.9

menggambarkan ide dasar di balik adopsi embedding yang tersedia secara bebas untuk tugas klasifikasi spesifik Anda.

Rencananya adalah menggunakan model pra-latih yang tersedia secara gratis untuk menghitung embedding teks Anda sendiri (misalnya, kumpulan ulasan rumah di situs web). Model-model ini melakukan semua kerja keras dalam memahami tata bahasa dan kosakata karena telah dilatih pada korpus teks yang besar yang memungkinkan mereka mempelajari semua aturan kompleks tersebut. Ketika Anda memiliki vektor embedding untuk teks yang Anda minati (misalnya, ulasan rumah), akan jauh lebih mudah untuk membangun pengklasifikasi untuk analisis sentimen atau tugas lainnya, karena embedding mengekspresikan makna yang jauh lebih dalam daripada teks mentah.

Cakupan Proyek Klasifikasi NLP

Klasifikasi teks adalah aplikasi umum dan bermanfaat yang dapat membantu dalam berbagai konteks di seluruh organisasi. Tabel 5.4 memberikan beberapa inspirasi.

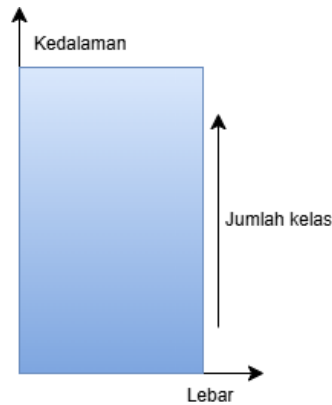
Tabel 5.4 Tiga contoh klasifikasi berbasis NLP untuk aplikasi bisnis

Aplikasi	Masukkan data	Kelas yang memungkinkan
Operator tiket otomatis	Tweet dan email dikirim ke layanan pelanggan	Akar penyebab masalah (kurangnya koneksi internet, kesalahan penagihan, dan sebagainya)
Pengkategorian klaim asuransi mobil otomatis	Laporan klaim asuransi	Jenis kerusakan (mesin rusak, tergores, dan sebagainya)
Pengklasifikasi klaim medis	Laporan dokter	Penyakit

Saat menerapkan klasifikasi dokumen pada permasalahan bisnis, penting untuk diingat bahwa risiko dan kinerja bergantung pada jumlah kelas yang ingin kita identifikasi. Pada grafik lebar/kedalaman pada Gambar 5.10, Anda dapat memvisualisasikan bagaimana kompleksitas meningkat dengan membayangkan bahwa semakin banyak kelas yang kita minta untuk diidentifikasi oleh algoritma AI kita, semakin tinggi kedalaman penerapannya, sehingga tugas menjadi lebih menantang.

Hal lain yang perlu diingat adalah jika Anda memiliki jumlah contoh yang terbatas untuk suatu kelas dalam data pelatihan Anda, akan sulit bagi algoritma untuk belajar mengenali elemen dari kelas tersebut, karena belum melihat cukup banyak kasus. Kembali ke pengklasifikasi tinjauan awal kita, katakanlah Anda memiliki situasi seperti pada Tabel 5.5. Dalam hal ini, kita memiliki banyak data pelatihan untuk tiga kelas pertama, tetapi kita kekurangan contoh untuk kelas terakhir. Kelas 4 hanya memiliki tiga contoh, dibandingkan dengan ratusan tinjauan di kelas-kelas lainnya. Solusi untuk situasi seperti ini adalah mendapatkan lebih banyak data pelatihan, atau cukup menggabungkan kelas yang kurang beruntung dengan kelas serupa lainnya. Misalnya, kelas 3 dan 4 keduanya terkait dengan

lingkungan tersebut, jadi kita bisa menggabungkannya ke dalam kelas "ulasan tentang lingkungan" yang lebih luas, yang totalnya akan menjadi $988 + 3 = 991$ ulasan. Dataset akhir memiliki satu kelas lebih sedikit, tetapi jauh lebih seimbang dalam hal contoh, seperti yang ditunjukkan pada Tabel 5.6.



Gambar 5.10 Kedalaman tugas NLP klasifikasi meningkat seiring bertambahnya jumlah kelas, sementara lebarnya tetap sama.

Tabel 5.5 Contoh awal dataset ulasan rumah

ID Kelas	Nama kelas	Jumlah contoh di kelas
1	Tinjauan keausan	1,541
2	Ulasan furnitur	769
3	Pelayanan publik dalam ulasan lingkungan	988
4	Tinjauan keamanan lingkungan	3

Tabel 5.6 Dataset yang sama setelah penggabungan kelas 3 dan 4

ID Kelas	Nama kelas	Jumlah contoh di kelas
1	Tinjauan keausan	1,541
2	Ulasan furnitur	769
3	Ulasan lingkungan sekitar	991

Mengurangi jumlah kelas juga membantu mengurangi kedalaman tugas, meningkatkan akurasi secara keseluruhan, dan dengan demikian mengurangi risiko sistem membuat kesalahan saat digunakan di dunia nyata.

Pencarian Dokumen

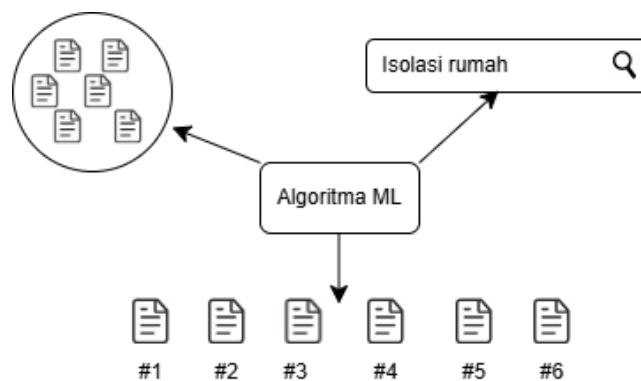
Di awal bab ini, kami memperkenalkan brokerbot: sebuah aplikasi AI yang mampu menginterpretasi teks layaknya manusia. Brokerbot terlalu rumit untuk teknologi AI modern, jadi kami mulai dari aplikasi paling sederhana yang dapat kami bangun: pengklasifikasi analisis sentimen. Langkah selanjutnya menuju pemahaman bahasa yang lebih mendalam adalah klasifikasi teks. Sayangnya, klasifikasi teks masih memiliki dua keterbatasan penting:

- Ini adalah tugas klasifikasi, yang berarti kami perlu mengumpulkan label.

- Kelas-kelas didefinisikan di awal: jika pengguna ingin mencari sesuatu yang tidak termasuk dalam kelas, kami tidak dapat membantu mereka.

Masalah pertama dapat diatasi dengan trik pengumpulan data yang cerdas atau hanya dengan membayar manusia untuk menambahkan label secara manual. Masalah kedua lebih menantang, dan inilah keahlian Google: bilah pencariannya yang kecil memungkinkan Anda mengetik apa pun yang Anda minati dan langsung mendapatkan daftar hasil yang diurutkan berdasarkan relevansi. Tugas ini merupakan cabang NLP yang sering disebut pencarian dan pengambilan dokumen.

Yang ingin kami lakukan adalah sebagai berikut: kami ingin memiliki bilah pencarian di bagian atas situs web kami, dan ketika pengguna mengetik sesuatu yang mereka minati (misalnya "isolasi rumah"), algoritme akan menelusuri semua ulasan yang tersedia dan memilih yang paling relevan dengan pencarian mereka, seperti pada gambar 5.11.

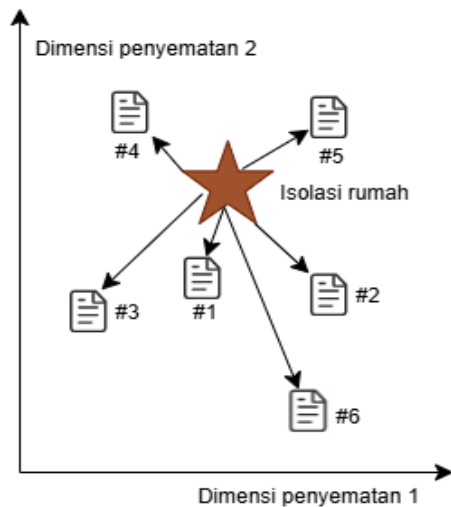


Gambar 5.11 Dalam pencarian dokumen, algoritma dilatih pada sekumpulan dokumen (kiri atas). Ketika menerima kueri dari pengguna (kanan atas), algoritma akan menghasilkan daftar konten yang paling relevan.

Untuk pencarian dokumen, konsep penyisipan kata (word embedding) sangat membantu. Ingatlah bahwa penyisipan kata adalah sekumpulan ratusan angka yang mengkodekan arti sebuah kata, dengan sifat bahwa kata-kata yang serupa memiliki penyisipan yang serupa. Penyematan memungkinkan kita untuk mentransfer makna "kemiripan" dari dunia sastra yang samar ke dunia angka yang dingin. Kita dapat memanfaatkan hal ini sehingga ketika pengguna bertanya tentang "isolasi rumah", kita dapat mengubah kedua kata ini menjadi penyisipannya, dan jika kita telah melakukan hal yang sama dengan semua ulasan, kita kemudian dapat mencari ulasan yang representasi numeriknya lebih dekat dengan yang kita dapatkan untuk "isolasi rumah".

Kami ingin membantu Anda memvisualisasikan konsep ini di kepala Anda, jadi bayangkan penyisipan dua elemen: setiap kata hanya diwakili oleh dua angka, seperti yang digambarkan pada gambar 5.12. Tentu saja, penyematan seperti itu tidak akan terlalu berguna dalam praktiknya, karena dua angka saja tidak cukup untuk menangkap kekayaan kosakata (penyematan yang digunakan dalam aplikasi dunia nyata terdiri dari ratusan atau ribuan angka).

Bayangkan menggambar sebuah titik untuk setiap dokumen yang kita miliki di basis data, lalu menggambar titik baru yang sesuai dengan penyematan kueri pengguna.



Gambar 5.12 Algoritma pencarian dokumen menjawab pertanyaan pengguna dengan mencari dokumen yang memiliki kesamaan penempatan dengan kalimat pertanyaan (ditunjukkan dengan bintang pada plot).

Berkat cara penyematan tersebut dibangun, dokumen yang paling mirip dengan kueri pengguna akan secara fisik dekat dengan titik kueri tersebut. Ini berarti mudah untuk menjawab kueri pengguna: cukup pilih beberapa titik yang paling dekat dengannya, dan titik-titik tersebut akan sesuai dengan dokumen yang paling mirip. Solusi ini juga berfungsi ketika vektor penyematan memiliki lebih dari dua angka; hanya saja lebih sulit untuk memvisualisasikannya.

Pencarian dokumen adalah aplikasi NLP yang cukup matang yang sering kali berperan sebagai blok penyusun dalam produk yang lebih kompleks. Tentu saja tidak secanggih teknik lainnya, tetapi sering kali menjadi salah satu langkah pertama yang dapat Anda ambil untuk mendapatkan nilai dari aset data berbasis teks Anda. Secara khusus, metode pencarian dokumen yang kami jelaskan di sini tidak dapat benar-benar menjawab pertanyaan yang bergantung pada logika atau interpretasi; satu-satunya hal yang dapat dilakukan adalah mengembalikan dokumen yang terdengar mirip dengan pertanyaan Anda, dengan harapan Anda akan menganggapnya bermanfaat.

Percakapan Alami

Dalam segala upaya kami untuk mendekati brokerbot, kami selalu mengandalkan teks yang dihasilkan manusia, dan dengan senang hati menciptakan algoritma yang memudahkan orang mengakses informasi tersebut. Sebuah peningkatan alami adalah model yang dapat menghasilkan bahasa secara mandiri, alih-alih merangkai kalimat-kalimat yang sudah jadi.

Kami ingin menyampaikan kabar buruknya segera: bidang ini masih berada di garis depan penelitian, dan teknologi saat ini belum siap untuk digunakan (sejauh ini). Tujuan kami dengan bagian ini adalah untuk menunjukkan kepada Anda batasan-batasannya, dan cara mengatasinya mengingat apa yang kita miliki saat ini.

Teknologi tercanggih di pasaran adalah model yang disebut GPT-2, yang dikembangkan oleh organisasi riset nirlaba OpenAI. Model ini mampu melakukan berbagai tugas, dan salah satu yang paling mengesankan adalah menghasilkan teks berdasarkan perintah manusia. Kami mengunduh implementasi sumber terbuka dari model ini dan mencobanya sendiri:

- Perintah kami —Mencapai kinerja seperti manusia dalam menghasilkan bahasa akan menjadi tonggak sejarah yang luar biasa bagi AI. Bisakah model GPT-2 OpenAI yang baru menjadi titik balik?
- Penyelesaian GPT-2—Dan apa dampak model ini terhadap proyek kecerdasan buatan lainnya seperti AIQ dan HCI?
- OpenAI lahir untuk mengatasi masalah seperti ini. Kami telah bekerja di bidang pengenalan suara selama lima tahun terakhir. Selama bekerja di bidang ini, kami mengembangkan pengenalan suara kami dalam kode sumber terbuka. Setelah pengumuman GPT-2 beberapa minggu yang lalu, saatnya untuk mengambil langkah selanjutnya dalam pengenalan suara dan kami bangga mempersembahkan versi 2.
- Tantangan terbesar yang dihadapi AI, masalah terpentingnya adalah
 - Informasi secara umum
 - Pengetahuan seringkali tidak dapat diakses
 - Informasi dapat diperoleh dari banyak sumber yang dapat saling bertentangan

Seperti yang Anda lihat, hasilnya cukup meyakinkan. Kontennya sendiri tidak selalu masuk akal, tetapi tetap sesuai topik dengan tata bahasa yang hampir sempurna. Hal ini dimungkinkan karena jumlah data yang sangat besar yang telah digunakan untuk melatih model ini: lebih dari 8 juta halaman web. Kembali ke real estat, berikut percobaan terakhir GPT-2 dengan kalimat lain yang berkaitan dengan rumah:

- 🚩 Permintaan kami—Rumah ini adalah penawaran yang luar biasa: pemandangan dari ruang tamunya luar biasa, dan lingkungannya sangat berkembang.
- 🚩 Penyelesaian GPT-2—Saya berharap untuk meningkatkan ke rumah dua kamar tidur karena saya sangat menyukai rumah dan lingkungan ini. Jika Anda tinggal di Florida tengah, Anda mungkin harus mencobanya.

Suami saya menyukai rumah saya di sisi utara kota dan kami dekat dengan Miami dan banyak tempat pantai. Lingkungan kami hanya satu blok jauhnya. Kami berdua sangat bahagia dan rumah kami sangat nyaman. Kami memang pindah tetapi tidak pernah berencana untuk pergi. Kami juga memiliki tempat lain untuk beristirahat dan belum pernah memiliki tempat yang nyaman ini untuk berlibur. Saya sangat merekomendasikannya! Kami menyukainya!!!

Model tersebut "memperhatikan" bahwa teks yang diberikannya menyerupai ulasan tentang sesuatu seperti daftar rumah Airbnb, dan terus menambahkan teks seolah-olah itu adalah ulasan rumah. Anda juga dapat menggunakan GPT-2 untuk menjawab pertanyaan tentang suatu teks, yang mungkin akan menjadi aplikasi yang lebih menarik daripada menghasilkan kalimat acak. Bahkan, ini akan sangat mirip dengan brokerbot, karena kita dapat menggunakan GPT-2 untuk menjawab pertanyaan pengguna tentang rumah dalam bahasa yang mudah dipahami. Berikut beberapa contoh kemampuan Tanya Jawab GPT-2:

- ❖ *Pertanyaan kami* — Model GPT-2 dikembangkan oleh OpenAI, melatih model pembelajaran mendalam tingkat lanjut pada lebih dari 40 GB teks internet dari 8 juta halaman web. Model ini memiliki performa yang luar biasa, tetapi masih jauh dari sempurna.

T: Siapa yang mengembangkan GPT-2?

- ❖ *Jawaban GPT-2* — GPT-2 oleh OpenAI dikembangkan untuk mengembangkan dan melatih model pembelajaran mendalam, yang kemudian dapat digunakan dalam aplikasi pembelajaran mesin. Tim telah menetapkan tujuan untuk membangun "mesin pembelajaran mesin untuk internet" untuk aplikasi generasi berikutnya.
- ❖ *Pertanyaan kami* — *From Zero to AI* adalah buku tentang kecerdasan buatan yang ditulis untuk orang-orang nonteknis. Buku ini akan membantu Anda memahami prinsip-prinsip inti AI dan cara menerapkannya pada bisnis Anda. Buku ini ditulis oleh Nicolò Valigi dan Gianluca Mauro, dan diterbitkan oleh Manning Publications.

T: Siapa yang menulis *From Zero to AI*?

- ❖ *Jawaban GPT-2*—*From Zero to AI* selalu ditulis oleh Nicolò Valigi. Saya selalu menjadi penggemar berat Valigi, dan buku-bukunya sebelumnya cukup berpengaruh dalam tulisan saya saat ini. Saya mengenalnya dari karyanya untuk Apple di Mac, iPod, dan iPhone mereka. Dia juga mengerjakan skenario untuk "The Lego Movie", film "Hobbit", dan "Serendity". Saya berkesempatan menulis buku ini untuk tahun pertama kuliah putri saya musim panas lalu, dan cukup menarik membaca bagaimana dia menggunakan pengalaman dan kesuksesannya selama bertahun-tahun di tingkat tertinggi industri untuk membantu menulis buku yang akan menarik bagi khalayak yang lebih luas. Karena itu, saya ingin berterima kasih kepada Nicolò Valigi, dan berharap dia akan terus menulis buku yang hebat untuk membantu kita semua menjadi lebih cerdas.

Meskipun kinerja ini tampak luar biasa, keterbatasan yang parah masih memengaruhi teknologi ini, dan para peneliti sedang bekerja keras untuk memperbaikinya. Contohnya adalah sebagai berikut:

- Sulit untuk mengontrol keluaran model. Misalnya, kami harus mencoba tiga hingga empat kali sebelum mendapatkan hasil yang disajikan di sini. Bahkan dengan topik-topik yang sangat populer dan banyak terdapat dalam data pelatihan, para peneliti OpenAI melaporkan bahwa model tersebut hanya dapat menghasilkan hasil yang kredibel sekitar 50% dari waktu.
- Ketika model tidak mengetahui suatu topik, ia mengarang hal-hal berdasarkan apa yang mungkin masuk akal, meskipun sepenuhnya salah. Contoh yang bagus adalah yang terakhir: model mengarang bahwa salah satu dari kami, para penulis, memiliki pengalaman di Apple dan menulis skenario, hanya karena pertanyaannya tentang teknologi dan buku.
- Model tersebut tidak benar-benar memahami arti dari apa yang dikatakannya: terkadang ia menyebutkan skenario yang mustahil seperti api yang menyala di bawah air.
- Model tersebut dapat macet dan mulai mengulang kata-kata secara berulang.
- Model tersebut dapat berganti topik secara tidak wajar.

Bisa dibilang, keterbatasan terbesar model-model ini adalah mereka mungkin tampak memahami konsep-konsep dalam prompt, tetapi sebenarnya menipu kita: yang mereka lakukan hanyalah menebak kata mana yang lebih mungkin muncul setelah sebuah kalimat,

berdasarkan miliaran kalimat yang telah mereka lihat. Dengan kata lain, mereka hanya mengulang-ulang data pelatihan mereka yang sangat besar, bukan memahami.

Anda dapat memvisualisasikan perbedaan antara pemahaman dan penalaran probabilistik dengan membandingkan seorang dokter dan algoritma ML. Seorang dokter dapat mendiagnosis pasien dengan membaca buku teks, memahami isinya, dan mungkin berlatih dengan sinar-X. Di sisi lain, tidak ada algoritma ML saat ini yang dapat melakukan hal yang sama. Untuk membuat diagnosis pada pemindaian, algoritma ML membutuhkan banyak contoh, dan hanya "membaca" satu buku sebelumnya tidak akan membantunya meningkatkan diagnosisnya dengan cara apa pun.

Lalu, bagaimana kita bisa lebih dekat dengan brokerbot? Apakah kita hanya perlu duduk menunggu pembuatan teks matang? Untungnya, tidak. Anda dapat mengadopsi strategi lain untuk menciptakan nilai hari ini, jadi bersabarlah.

Merancang Produk yang Mengatasi Keterbatasan Teknologi

Bayangkan kita benar-benar ingin membangun brokerbot dengan teknologi apa pun yang tersedia saat ini. Kita bisa kembali ke algoritma analisis sentimen dan pengklasifikasi topik kita, dengan daftar kalimat siap pakai yang rapi untuk setiap topik. Katakanlah pengguna menulis sesuatu seperti ini:

Apa apartemen dengan pemandangan terbaik?

Pengklasifikasi topik kita mungkin memahami bahwa topik pertanyaannya adalah panorama yang dapat Anda nikmati dari apartemen tersebut, dan analisis sentimen kita tahu bahwa kita tertarik pada sesuatu yang positif (kita juga tahu bahwa ini adalah pertanyaan karena tanda tanya di akhir, tidak perlu teknologi canggih). Sekarang, Anda mungkin memiliki serangkaian pertanyaan yang sesuai dengan persyaratan "topik = pemandangan" dan "sentimen = baik," seperti ini:

Apakah Anda lebih suka berada di tepi laut atau dekat taman?

Perhatikan bahwa pertanyaan ini tidak dihasilkan oleh algoritma AI yang cerdas; seorang manusia yang membuatnya dan menyimpannya dalam basis data bersama banyak pertanyaan lain. Pengguna dapat menjawab sesuatu seperti ini:

Saya suka pohon! Taman Golden Gate adalah favorit saya.

Sekali lagi, analisis sentimen kita akan memahami bahwa kalimat tersebut memiliki nuansa positif, pengklasifikasi topik kita akan memahami bahwa kita sedang membicarakan pohon (yang memiliki embedding yang mirip dengan taman yang disebutkan sebelumnya—lihat betapa hebatnya embedding?), dan oleh karena itu jawaban untuk pertanyaan tersebut adalah "taman". Jika Anda membuat pengklasifikasi untuk melakukannya, Anda juga dapat menangkap referensi ke Taman Golden Gate dan mencari rumah-rumah di dekatnya. Dari

sana, Anda mungkin telah membuat skrip yang memahami bahwa inilah waktu yang tepat untuk menampilkan properti, dan menyarankan rumah terbaik untuk pengguna tertentu tersebut.

Kelihatannya luar biasa dan mulus, bukan? Tetapi bagaimana jika jawaban yang diberikan pengguna untuk pertanyaan sebelumnya seperti ini?

Saya akan mengatakan bahwa saya suka laut, tetapi ketika musim dingin, ada banyak badai di sini, jadi saya akan memilih yang lain.

Ini mungkin hal terburuk yang bisa kita dapatkan. Jelas, pengguna lebih menyukai pemandangan daripada taman yang indah, tetapi kalimat tersebut menawarkan beberapa tantangan:

- ✓ Kalimat tersebut membahas tentang pemandangan laut.
- ✓ Kalimat tersebut tidak pernah benar-benar menyebutkan apa pun yang berhubungan dengan taman.
- ✓ Kalimat tersebut menyebutkan "cinta," yang menyampaikan makna positif.
- ✓ Tidak ada kata yang mengisyaratkan sentimen negatif tentang pemandangan laut.

Algoritma kemungkinan besar akan bingung dengan kalimat seperti ini, dan melanjutkan skrip yang telah ditentukan dengan pertanyaan lanjutan yang salah. Lebih buruk lagi, pengguna mungkin menjawab seperti ini:

Tahukah Anda? Saya sangat suka melihat gedung-gedung tinggi dengan arsitektur yang indah.

Bagaimana jika Anda belum memikirkan jawaban ini? Pengguna mengikuti alur skrip yang buntu, dan Anda tidak memiliki jawaban. Anda mungkin akan menjawab dengan sesuatu yang umum seperti ini: "Saya tidak mengerti maksud Anda, bisakah Anda mengulanginya?" Pengguna tidak akan pernah bisa menjawab pertanyaan dengan cara yang Anda kuasai (Anda sama sekali tidak mengodekan apa pun yang sesuai dengan minat tersebut!) dan akan frustrasi lalu meninggalkan produk tersebut.

Apakah kedengarannya familiar? Tentu saja familiar bagi kami, dan inilah pengalaman frustrasi yang dialami konsumen dari perusahaan yang memutuskan menggunakan chatbot tanpa memastikan fungsionalitasnya sesuai dengan kemampuan teknologi saat ini. Solusinya terletak pada cara Anda menentukan cakupan aplikasi. Jika Anda menyadari keterbatasan teknologi saat ini, berusaha sebaik mungkin untuk tetap berada dalam batasan ini, membangun aplikasi Anda di sekitarnya, dan menjaga kompleksitas tetap rendah, baik dari segi lebar maupun kedalaman. Dalam kasus brokerbot, kita mungkin secara eksplisit membatasi fungsionalitas brokerbot ke sebagian kecil tindakan, seperti hanya menanyakan rumah berdasarkan parameter dasar (lokasi, ukuran, harga). Selain itu, sebisa mungkin hindari bahasa alami, gunakan pertanyaan tertutup (ya/tidak), atau tombol yang memungkinkan pengguna memilih dari berbagai opsi. Misalnya, bot mungkin menjadi yang memulai

percakapan dengan pengguna sehingga dapat mengarah ke topik tertentu, sehingga membatasi luasnya jawaban. Bot mungkin memulai percakapan dengan menanyakan hal ini:

Halo! Saya brokerbot, saya di sini untuk membantu Anda menemukan rumah baru. Pertama-tama, berapa anggaran Anda?

Di sini, Anda dapat memutuskan untuk membiarkan pengguna memberikan jawaban dalam bahasa alami atau menyediakan slider untuk memilih harga. Pilihan yang terakhir adalah yang paling aman, tetapi karena kami tahu bahwa jawaban pengguna akan seputar harga, kami juga yakin bahwa kami akan dapat memahami jawaban dalam bahasa alami. Dari sana, bot dapat melanjutkan dengan pertanyaan tentang lokasi, ukuran, dan sebagainya, hingga akhirnya dapat menunjukkan pilihan rumah kepada pengguna, dan bahkan meminta umpan balik.

Skrip dan desain percakapan masih merupakan satu-satunya cara untuk membangun teknologi yang memberikan nilai bagi pengguna. Kami mendorong Anda untuk menggunakan pengetahuan yang Anda peroleh dalam bab ini untuk berpikir kritis tentang apa yang Anda inginkan dari chatbot Anda dan bagaimana Anda ingin chatbot tersebut berhubungan dengan pengguna.

5.4 STUDI KASUS: TRANSLATED

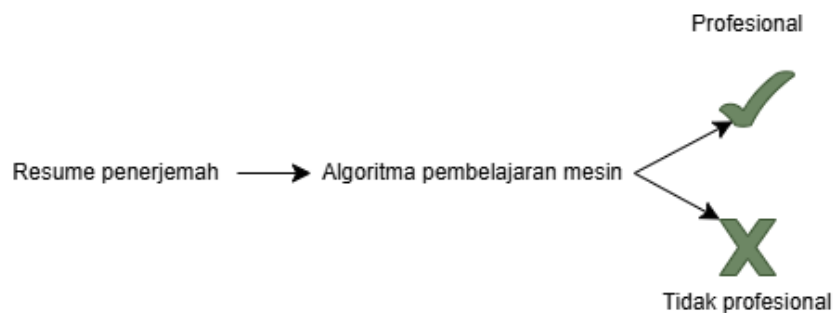
Pada tahun 1999, Marco Trombetti dan Isabelle Andrieu memutuskan untuk memulai bisnis bersama. Trombetti sedang belajar fisika, dan Andrieu adalah seorang ahli bahasa. Untuk memanfaatkan keahlian mereka, pasangan muda ini mencetuskan ide untuk membangun perusahaan penerjemahan berbasis internet. Mereka menamainya Translated, dan meluncurkannya dengan investasi awal sebesar Rp. 1.000.000 untuk domain internet Translated.net. Dalam 20 tahun berikutnya, Translated menjadi perusahaan penerjemahan profesional daring terbesar di dunia, tumbuh secara organik tanpa investasi eksternal.

Translated berfokus pada penerjemahan profesional, seperti jurnal medis dan buku panduan teknik. Penerjemahan semacam itu biasanya dilakukan oleh penerjemah profesional yang dikontrak oleh perusahaan yang bertindak sebagai perantara. Mencocokkan teks dengan penerjemah bukanlah hal yang mudah: banyak dokumen membutuhkan keterampilan bahasa yang mendalam dan keahlian subjek. Misalnya, menerjemahkan makalah tentang implantasi katup aorta transkateter dari bahasa Inggris ke bahasa Jerman membutuhkan seorang ahli jantung yang ahli dalam kedua bahasa tersebut. Broker penerjemahan yang memiliki basis data penerjemah dapat memanfaatkan hal ini dan mengenakan premi yang besar untuk menyediakan penerjemah yang tepat bagi klien.

Semakin besar perusahaan penerjemahan, semakin besar pula tantangan dan inefisiensi dalam mencocokkan penerjemah dengan pekerjaan. Ini adalah peluang yang coba dimanfaatkan Trombetti dan Andrieu. Tantangan pertama yang harus dihadapi Trombetti dan Andrieu adalah membangun basis data penerjemah yang besar. Sejak awal, keduanya berinvestasi dalam optimasi mesin pencari (SEO) agar mereka dapat berada di peringkat teratas pencarian Google untuk layanan penerjemahan profesional. Karena saat itu internet

masih baru, investasi mereka segera membuahkan hasil, dan Translated mulai menerima ratusan resume per bulan. Keberhasilan awal ini menimbulkan tantangan baru: banyak penerjemah yang mengirimkan resume mereka tidak memenuhi standar kualitas perusahaan, dan memilah penerjemah profesional dari yang non-profesional membutuhkan banyak upaya manusia.

Translated berisiko mengalami inefisiensi yang sama seperti broker penerjemahan lainnya. Inspirasi untuk mengatasi tantangan ini datang kepada Trombetti pada tahun 2002, ketika ia membaca esai Paul Graham yang sama, "Rencana untuk Spam", yang telah kita bahas sebelumnya di bab ini. Trombetti berpikir bahwa metode yang sama yang digunakan untuk membedakan email spam dan bukan spam dapat digunakan untuk membuat penilaian awal apakah resume yang mereka terima berasal dari penerjemah profesional atau nonprofesional, seperti yang digambarkan pada gambar 5.13.



Gambar 5.13 Algoritma ML Translated membantu perusahaan mengidentifikasi penerjemah profesional dan nonprofesional secara otomatis.

Algoritma ini pertama kali diperkenalkan pada tahun 2002, ketika perusahaan mulai menerima ratusan permintaan terjemahan per hari, dan terus ditingkatkan sejak saat itu. Evolusi algoritma selanjutnya adalah mengubahnya menjadi AI pencocokan penerjemah-dokumen otomatis, melatihnya menggunakan pencocokan yang dibuat oleh manajer proyek (PM) Translated selama bertahun-tahun.

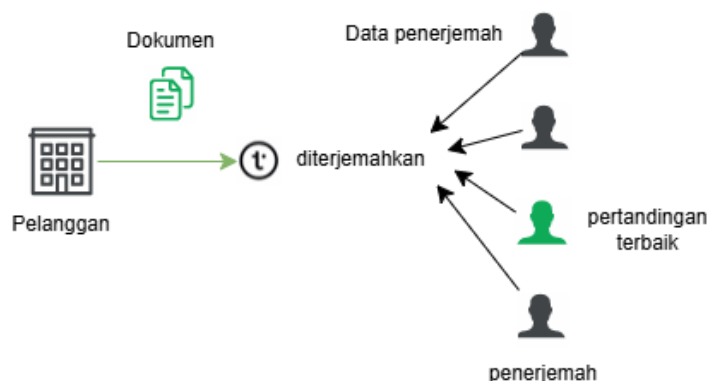
Setelah diterapkan, pada setiap permintaan terjemahan baru, algoritma akan memeriksa dokumen dan kumpulan penerjemah perusahaan dan menghitung probabilitas kecocokan untuk setiap penerjemah. PM Translated kemudian akan memvalidasi kecocokan dan menugaskan tugas. Proses ini memungkinkan startup muda ini untuk menangani banyak klien dan penerjemah tanpa mengorbankan kecepatan pengiriman (faktor yang seringkali krusial bagi klien) dan menjaga biaya overhead mereka tetap rendah dengan memiliki tim PM yang lebih kecil.

Saat ini, Translated bertindak sebagai platform dua sisi: di satu sisi, penerjemah mengajukan penawaran jasa mereka; dan dari sisi lain, pelanggan mengirimkan dokumen yang perlu mereka terjemahkan. Di sisi belakang, Translated mencocokkan penerjemah yang tepat untuk setiap pekerjaan dan memberikan hasil akhir kepada klien, tanpa perlu berinteraksi satu sama lain. Versi algoritma mereka saat ini diberi nama T-Rank.

Algoritme ini memperhitungkan hingga 30 faktor dari dokumen sumber, termasuk resume penerjemah (bidang keahlian) dan proyek sebelumnya (kualitas dan ketepatan waktu pekerjaan sebelumnya). Dengan menggunakan data ini, T-Rank dapat memeringkat penerjemah dalam portofolio perusahaan berdasarkan kesesuaiannya dengan pekerjaan dan memberikan profesional terbaik untuk setiap tugas, seperti pada gambar 5.14. Seperti pada versi pertama algoritma ini, rekomendasi akhir diperiksa oleh PM ahli. Jika PM menolak rekomendasi, umpan balik dicatat dan digunakan untuk menyempurnakan algoritma agar terus meningkatkan kinerjanya.

Kesuksesan di pasar ini membawa Translated baik uang tunai maupun data. Arus kas yang konsisten memungkinkan Trombetti dan Andrieu menjalankan perusahaan tanpa mencari investasi eksternal dan mempertahankan kepemilikan penuh. Di sisi lain, Translated mulai mengumpulkan terjemahan yang dibuat oleh para profesional melalui sistem terjemahan berbantuan komputer yang disebut Matecat, menyempurnakan algoritma pencocokannya, dan mendorong perusahaan menuju misi baru: menjadi pemimpin dalam penerjemahan berbantuan mesin.

Untuk mengevaluasi kinerja terjemahan yang dihasilkan oleh mesin, Translated selalu melacak persentase kata yang disarankan yang diubah oleh penerjemah profesional. Pada awalnya, penerjemah harus mengubah 45% dari kata yang disarankan.



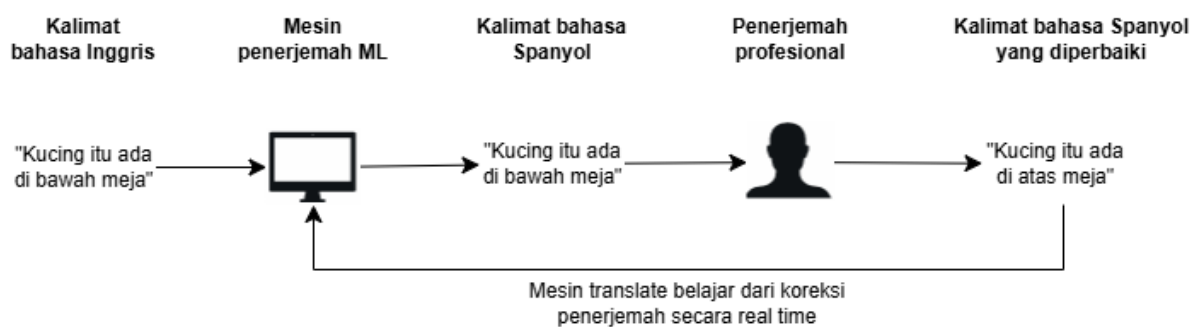
Gambar 5.14 Translated adalah platform dua sisi: pelanggan mengirimkan dokumen, penerjemah mengirimkan data mereka, dan perusahaan menemukan penerjemah terbaik dalam portofolionya untuk setiap dokumen yang akan diterjemahkan.

Ditambah dengan pencocokan penerjemah yang dibantu AI, hal ini cukup untuk menghasilkan terjemahan lebih cepat daripada pesaing. Pada tahun 2018, perusahaan dapat mengandalkan lebih dari 1,2 juta terjemahan profesional, dan persentase kata yang diubah oleh penerjemah profesional turun menjadi 24% pada delapan bahasa yang paling banyak digunakan, 30% lebih rendah daripada mesin penerjemah lainnya.

Untuk lebih meningkatkan pengalaman dan produktivitas penerjemah, perusahaan telah berinvestasi dalam teknologi yang diberi nama ModernMT, terjemahan mesin saraf adaptif pertama. Secara tradisional, algoritma ML dilatih pada sejumlah besar data (fase pelatihan) dan bersifat statis saat digunakan (fase inferensi). Jika seorang penerjemah

diberikan saran yang buruk dan memperbaikinya, perbaikan ini tidak akan dipertimbangkan oleh algoritma terjemahan mesin (MT) hingga pelatihan diulang dengan data baru. Ini berarti bahwa saat seorang penerjemah mengerjakan sebuah dokumen, perbaikan yang sama mungkin harus diterapkan beberapa kali. Hasilnya adalah pengalaman yang membuat frustrasi dan tidak efisien bagi penerjemah.

Seperti yang diilustrasikan pada Gambar 5.15, ModernMT mempelajari dan menyesuaikan secara langsung (real-time) saat penerjemah mengerjakan terjemahan. Jika frasa yang sama muncul lebih dari sekali dalam sebuah dokumen dan penerjemah memperbaikinya sekali, ModernMT akan mempelajari koreksi tersebut untuk pertama kalinya dan menghindari pengulangan kesalahan.



Gambar 5.15 Mesin penerjemah ModernMT dari Translated mengusulkan terjemahan yang pada akhirnya dapat dikoreksi oleh penerjemah. Koreksi penerjemah tersebut kemudian dipelajari oleh ModernMT.

Pada tahun 2019, Translated menerima 2.000 resume per bulan dari para penerjemah. Perusahaan ini telah menjadi perantara 1,2 juta terjemahan profesional untuk lebih dari 130.000 pelanggan, dan terjemahan tersebut dilakukan dalam 151 bahasa oleh lebih dari 180.000 penerjemah profesional. Algoritma yang dibangun berdasarkan kumpulan data ini memungkinkan kinerja penerjemahan mesin yang canggih pada 27 bahasa yang paling umum. Algoritma Translated memfasilitasi, tetapi tidak menggantikan, pekerjaan para profesional. Algoritma ini menyediakan terjemahan otomatis pertama dari kalimat-kalimat yang hanya perlu disempurnakan dan disetel dengan baik: saat ini, penerjemah hanya perlu mengubah 24% kata yang disarankan oleh platform perusahaan. Ini 30% lebih baik daripada solusi lain yang saat ini ada di pasaran. Translated melaporkan pendapatan sebesar Rp. 210 miliar pada tahun 2018, dengan 51 karyawan: 12 pengembang, 35 di bagian operasional, dengan perkiraan pertumbuhan sebesar 40% untuk tahun 2019.

Translated masih berinvestasi dalam meningkatkan mesin penerjemahannya. Tujuannya bukanlah untuk membuat AI menghasilkan terjemahan yang sempurna. Bahkan, jika terjemahan yang dilakukan oleh penerjemah profesional ditinjau oleh profesional lain, orang tersebut masih akan mengubah sekitar 11% kata. Hal ini disebabkan oleh ekspresivitas bahasa yang ekstrem: konsep yang sama dapat diungkapkan dengan kata-kata yang berbeda untuk mengekspresikan makna dan nuansa yang sedikit berbeda. Menurut Trombetti, ini adalah tugas terpenting dan memuaskan yang dibutuhkan seorang penerjemah: memastikan

bahwa kekayaan dokumen terekspresikan dalam terjemahannya. Oleh karena itu, tujuannya adalah meningkatkan algoritma AI perusahaan agar penerjemah dapat berfokus pada 11% kata yang perlu mereka ubah untuk menyampaikan pesan asli dokumen dengan cara yang paling ekspresif. Trombetti dan Andrieu ingin memungkinkan para penerjemah melakukan pekerjaan yang lebih memuaskan dan berkualitas tinggi, serta mendemokratisasi proses penerjemahan. Berikut pernyataan mereka saat wawancara:

Teknologi kami menciptakan banyak peluang, tidak hanya bagi kami, tetapi juga bagi para penerjemah, yang akhirnya tidak lagi bekerja mengoreksi hal-hal yang sama dan membosankan, tetapi meluangkan waktu untuk menjadi lebih manusiawi, mengomunikasikan makna teks yang sebenarnya, dan menjadi lebih persuasif dan kreatif.

PERTANYAAN KASUS

1. Bagaimana Anda dapat mengembangkan strategi untuk AI dalam tugas-tugas NLP?
2. Apakah AI merupakan alat untuk menggantikan manusia?

DISKUSI KASUS

Translated adalah contoh yang bagus dari strategi inkremental yang sukses untuk AI. Pada awalnya, perusahaan menggunakan teknologi AI untuk menangani kasus penggunaan sederhana: mengklasifikasikan resume penerjemah menjadi profesional atau nonprofesional. Perhatikan bahwa kasus penggunaan pertama ini merupakan salah satu aplikasi paling sederhana yang dapat dibangun perusahaan, tetapi risiko teknologinya sangat rendah karena Trombetti dan Andrieu tahu bahwa pendekatan yang sama telah dicoba di domain lain (klasifikasi spam email). Meskipun sederhana, kasus penggunaan ini memiliki beberapa karakteristik yang membuatnya cocok untuk AI pada saat itu, seperti berikut:

- Memainkan peran yang sangat strategis dalam mengungguli pesaing.
- Memiliki waktu yang singkat untuk mencapai ROI (hasilnya dapat langsung terlihat).
- Memiliki risiko implementasi yang rendah.

Keberhasilan aplikasi pertama ini mendatangkan lebih banyak pelanggan dan data. Aset data Translated yang diperkaya memungkinkan siklus positif, menjadikan teknologi pencocokan penerjemah lebih canggih seiring bertambahnya volume data, dan memungkinkan aplikasi baru yang sebelumnya tidak mungkin dilakukan, seperti terjemahan berbantuan AI.

Meskipun terjemahan berbantuan AI akan menjadi fitur unggulan 20 tahun yang lalu, perusahaan tidak langsung berinvestasi di dalamnya. Kurangnya data adalah alasan pertama untuk tidak melakukannya, tetapi alasan terbesar mengapa Trombetti menunggu adalah karena teknologinya belum cukup matang. Ini adalah contoh utama dari memanfaatkan gelombang teknologi: membiarkan penelitian mengambil langkah pertama yang sulit menuju kematangan suatu teknologi, dan menggunakan temuan tersebut bersama dengan aset data strategis untuk mengembangkannya, mencapai kinerja mutakhir.

Waktu yang tepat dapat menjadi faktor kunci kesuksesan bagi perusahaan seperti Translated yang bekerja di ujung tombak teknologi. Perhatikan juga bahwa Translated mengetahui bahwa penerjemahan mesin adalah tugas yang sangat kompleks, yang akan cocok dengan baik di bagian atas dan di sebelah kanan diagram lebar/kedalaman, tetapi membingkai teknologinya sedemikian rupa sehingga tetap sangat berguna. Perusahaan melakukannya dengan memberikan keputusan akhir kepada penerjemah profesional: teknologinya tidak menghasilkan dokumen akhir, melainkan perkiraan awal yang dapat dikerjakan oleh manusia ahli.

Fokus pada pengalaman pengguna ini merupakan elemen kunci dalam strategi perusahaan. Translated memfokuskan upaya penelitian ML-nya untuk memecahkan masalah yang paling sering dialami oleh penerjemah mereka, dan hal ini mengarah pada pengembangan ModernMT: jaringan saraf pertama yang beradaptasi dengan alur kerja penerjemahan pada umumnya. Inovasi ini memungkinkan penerjemah untuk menghindari perbaikan terjemahan mereka beberapa kali pada dokumen yang sama, sehingga menghemat waktu dan frustrasi mereka. Mencapai kinerja yang sama menggunakan jaringan saraf statis standar akan membutuhkan upaya besar dalam hal penelitian dan data.

Sekarang mari kita bahas tentang hubungan antara AI dan penggantian manusia. Seperti yang kini Anda ketahui, AI modern didasarkan pada pembelajaran mesin: sebuah teknologi yang mampu mempelajari cara melakukan tugas dari data. Ini berarti bahwa aplikasi AI saat ini dapat menangani tugas-tugas tunggal yang terdefinisi dengan baik, dengan keluaran spesifik dan metrik evaluasi yang terdefinisi dengan jelas.

Sebagian besar pekerjaan manusia lebih kompleks dari ini. Suatu pekerjaan mungkin memerlukan serangkaian tugas yang, jika dilihat secara terpisah, mungkin cocok untuk algoritma ML tetapi secara keseluruhan memerlukan keterampilan tambahan agar dapat dilakukan dengan benar. Contoh keterampilan yang dibutuhkan adalah sebagai berikut:

- Penanganan pengecualian
- Empati
- Kreativitas
- Membuat keputusan berdasarkan data yang tidak lengkap

Di sisi lain, algoritma AI biasanya lebih baik daripada manusia dalam tugas-tugas berikut:

- Membuat keputusan berdasarkan data dalam jumlah besar
- Kecepatan
- Objektivitas (asalkan dilatih dengan data yang tidak bias)

Jika kita melihat tugas-tugas yang dilakukan manusia untuk memenuhi pekerjaan mereka, kita menemukan subset yang tidak memerlukan keterampilan khusus manusia dan cocok untuk agen AI. Bagi Translated, ini seperti memindai resume. Jika perusahaan mempekerjakan orang untuk memindai ribuan resume dan menemukan kecocokan terbaik untuk sebuah terjemahan, mereka akan membutuhkan tim yang besar untuk membuat keputusan yang lugas (misalnya, membuang resume dokter untuk menerjemahkan manual pesawat). Seleksi tingkat tinggi ini sebagian besar membutuhkan kecepatan dan kemampuan untuk memproses informasi dalam jumlah besar, dua karakteristik yang menjadi keunggulan algoritma ML.

Saat mengidentifikasi proyek ML berpotensi tinggi, Trombetti memiliki aturan sederhana:

Perhatikan tugas-tugas yang dapat dilakukan perusahaan dengan menggunakan 1.000 pekerja magang yang tidak terlatih. Tugas-tugas ini dapat diotomatisasi oleh AI.

Perusahaan tahu bahwa memilih penerjemah untuk pekerjaan yang rumit membutuhkan kepekaan dan pengalaman. Karena alasan ini, keputusan akhir tidak diserahkan kepada algoritma: hasilnya tetap diperiksa oleh manajer proyek yang menggunakan keterampilan lunak untuk memvalidasi atau menolak pilihan algoritmik.

Bagaimana dengan layanan terjemahan berbantuan AI dari Translated? Melakukan terjemahan dokumen berkualitas tinggi membutuhkan empati dan kreativitas. Kedua karakteristik ini merupakan ranah manusia, dan algoritma pembelajaran mesin (ML) jelas bukan solusi yang layak untuk masalah tersebut. Namun, Translated juga telah berhasil menggunakan AI untuk aplikasi ini. Alasan utama di balik kesuksesan Translated dalam menawarkan terjemahan berbantuan AI adalah karena mereka tidak pernah memasarkan AI sebagai pengganti manusia. Ini berarti melampaui kemampuan AI dan menghasilkan produk yang buruk. Yang dilakukan perusahaan ini adalah menggunakan AI sambil dengan tegas menyatakan bahwa hasilnya tidak sempurna, tetapi cukup baik untuk mempermudah pekerjaan penerjemah. Dengan cara ini, perhatian mereka dapat diarahkan pada bagian-bagian dokumen yang membutuhkan lebih banyak empati dan kreativitas, sambil menyerahkan tugas menerjemahkan kalimat-kalimat yang paling mudah kepada AI Translated.

- Model pemrosesan bahasa alami dapat memahami dan mengklasifikasikan makna teks, dan bahkan menulis sesuatu sendiri.
- Teknologi NLP saat ini tidak selalu sesuai dengan harapan. Membingklai kompleksitas dalam hal lebar dan kedalaman membantu tim Anda memahami apa yang mungkin dan apa yang tidak.
- Dengan memecah dan menentukan ruang lingkup proyek NLP, Anda dapat mengatasi keterbatasan dalam algoritma mutakhir. Analisis sentimen, klasifikasi teks, dan pencarian dokumen adalah contoh dari blok bangunan yang lebih matang ini.
- Peluncuran fitur AI secara bertahap adalah cara yang bagus untuk membangun kepercayaan pada model NLP dan membangun pusat kekuatan berbasis AI, seperti yang Anda lihat dalam studi kasus Terjemahan.

BAB 6

AI UNTUK VALIDASI KONTEN DAN PENGEMBANGAN KOMUNITAS

Bab ini membahas

- Menggunakan sistem rekomendasi untuk menyarankan konten dan produk yang menarik
- Memahami dua pendekatan sistem rekomendasi: berbasis konten dan berbasis komunitas
- Memahami kekurangan rekomendasi algoritmik
- Studi kasus: menggunakan sistem rekomendasi untuk menghemat Rp. 10 triliun dalam churn

Sistem rekomendasi adalah kunci di balik pengalaman "personalisasi" masa kini, dan merupakan alat fundamental untuk membantu konsumen menavigasi katalog media, produk, dan pakaian yang sangat besar. Setiap kali Anda mengklik penawaran terkait dari Amazon atau menonton film yang disarankan dari Netflix, perusahaan-perusahaan ini memanfaatkan sistem rekomendasi untuk mendorong keterlibatan pengguna dan pertumbuhan penjualan. Tanpa sistem rekomendasi, menavigasi beragam produk dan konten digital yang tersedia di internet akan mustahil. Dalam studi kasus di akhir bab ini, Anda akan melihat bagaimana Netflix yakin bahwa sistem rekomendasinya telah menghemat lebih dari Rp. 10 triliun setiap tahun sejak 2015.

6.1 KUTUKAN PILIHAN

Pernahkah Anda memasuki mal besar untuk berbelanja pakaian dan merasa bingung? Sulit menemukan kaus impian Anda ketika ada ratusan pilihan, dengan warna, bahan, merek, dan harga yang berbeda. Internet tidak memiliki batasan ruang, sehingga Anda berpotensi harus memilih bukan dari ratusan, tetapi puluhan ribu kaus. Atau film. Atau lagu. Atau artikel berita. Atau tablet pencuci piring. Frustrasi sudah di depan mata, dan pelanggan yang frustrasi tidak baik untuk bisnis.

Internet begitu penuh dengan pilihan sehingga kita perlu menemukan cara untuk membantu mewujudkannya. Di bab 3, kami mencoba menemukan cara untuk mengelompokkan pengguna berdasarkan selera mereka dan menargetkan mereka dengan penawaran yang sesuai dengan selera tersebut, menggunakan serangkaian teknik yang disebut pembelajaran tanpa pengawasan. Namun, itu masih belum cukup untuk mencapai singularitas pemasaran: tujuan utama komunikasi satu lawan satu, di mana setiap pelanggan mendapatkan penawaran dan rekomendasi yang sepenuhnya dipersonalisasi untuk kepuasan maksimal. Inilah topik bab ini.

Bagaimana Anda dapat membantu orang membuat pilihan? Jika kita melihat asisten toko yang hebat, mereka melakukan salah satu atau kedua hal ini:

- Mereka mengenal katalog mereka. Jika Anda menyukai celana jin, mereka akan dapat menyarankan celana jin lain yang serupa dan dengan cepat mengarahkan Anda ke pilihan yang tepat.
- Mereka mengenal Anda. Jika Anda adalah pelanggan tetap, mereka mempelajari selera Anda dan dapat menyarankan barang yang mungkin Anda sukai: "Anda harus mencoba celana jin ini; celana jin ini sangat trendi saat ini dan sangat cocok dengan gaya Anda."

Di sepanjang bab ini, Anda akan melihat cara membangun algoritma yang dapat melakukan kedua hal ini, bahkan lebih baik daripada manusia. Algoritma ini, yang disebut sistem rekomendasi, menghasilkan rekomendasi yang disesuaikan untuk pelanggan dalam skala besar.

6.2 MENINGKATKAN KETERLIBATAN DENGAN SISTEM REKOMENDASI

Berkat fitur berbasis AI yang telah kami bangun di bab-bab sebelumnya, FutureHouse semakin menarik pengguna. Estimasi harga jual gratis dan daftar otomatis juga telah menarik banyak penjual baru. Semua kesuksesan ini membawa tantangan baru: begitu banyak properti yang terdaftar untuk dijual sehingga sulit bagi pembeli untuk menemukan rumah yang sempurna. Kita perlu menemukan cara untuk menyoroti properti yang sesuai dengan selera mereka, dan memudahkan mereka menemukan rumah impian. Solusinya jelas: kita perlu membangun fitur Rumah Rekomendasi untuk Anda yang mempelajari selera setiap pengguna dan memilihkan rumah untuk mereka.

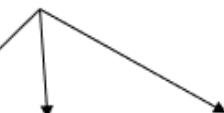
Mari kita coba pikirkan bagaimana FutureHouse dapat memutuskan rumah mana yang akan ditampilkan kepada setiap pengguna. Bagaimana seorang manusia yang berpengalaman akan mengatasi masalah ini? Seorang agen real estat dapat melihat satu atau lebih rumah yang disukai pelanggan dan mengusulkan rumah lain yang serupa (misalnya, di lingkungan yang sama atau dengan jumlah kamar yang sama).

Bagaimana kita menerjemahkan pendekatan ini ke dunia internet? Ide yang jelas adalah melacak rumah-rumah yang dikunjungi setiap pelanggan di situs, dan menyarankan mereka untuk memeriksa properti yang serupa dengan yang telah mereka temukan. Misalnya, jika calon pembeli telah melihat beberapa rumah tiga kamar tidur di dekat stasiun kereta, kemungkinan besar mereka akan tertarik pada properti lain yang memiliki atribut yang sama. Namun, bagaimana kita mendefinisikan arti "mirip"? Jika kita memiliki ide yang jelas tentang cara mengukur kesamaan, kita bisa menjelaskannya kepada komputer. Namun, ide kita tentang "mirip" bersifat intuitif dan alami, sehingga kita kesulitan untuk melakukannya. Sekarang, Anda sudah familiar dengan fakta bahwa ini adalah titik awal yang sempurna untuk pembelajaran mesin.

Di bab 2, kami memperkenalkan konsep fitur, aspek terukur dari setiap rumah yang dapat digunakan sebagai input untuk algoritma ML. Lebih spesifiknya, dasar yang kami lakukan untuk membangun model prediksi harga menunjukkan bahwa jumlah kamar, luas bangunan, dan jarak ke transportasi umum merupakan beberapa faktor terpenting yang perlu diperhitungkan. Serupa dengan yang kami lakukan untuk prediksi harga, kami dapat

memasukkan nilai fitur untuk setiap rumah dalam lembar kerja Excel, seperti yang kami lakukan pada Gambar 6.1.

Fitur



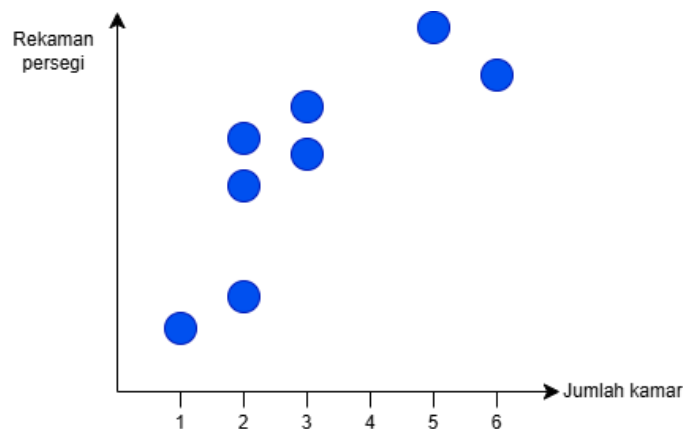
	A	B	C
1	Kamar	Meter persegi	Jarak dari pusat kota
2	2	80	5,4 km
3	1	42	7 km
4	3	120	23 km
5	2	65	2 km

Contoh



Gambar 6.1 Lembar kerja Excel sangat bagus untuk memvisualisasikan fitur-fitur rumah. Setiap baris adalah properti yang berbeda, dan setiap kolom adalah fitur.

Mari kita mulai dengan sederhana dan pertimbangkan satu fitur rumah: luas persegi. Jelas bahwa rumah seluas 1.000 kaki persegi dan rumah seluas 1.100 kaki persegi lebih mirip satu sama lain daripada rumah seluas 12.000 kaki persegi. Oleh karena itu, algoritma kami harus merekomendasikan rumah seluas 1.000 kaki persegi kepada pengguna yang telah mengunjungi rumah seluas 1.100 kaki persegi, dan menghindari menampilkan vila seluas 12.000 kaki persegi.

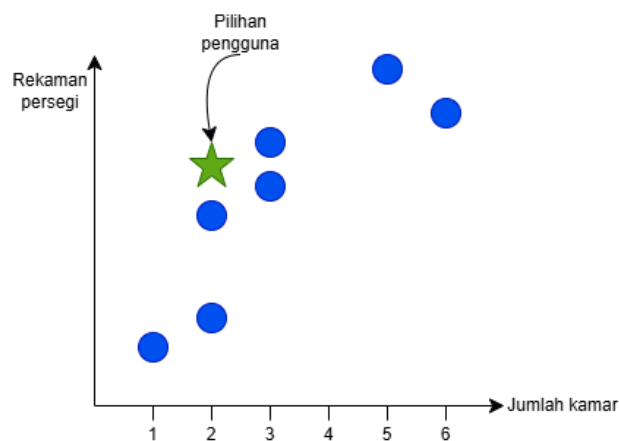


Gambar 6.2 Representasi rumah berdasarkan dua fitur: jumlah kamar dan luas persegi

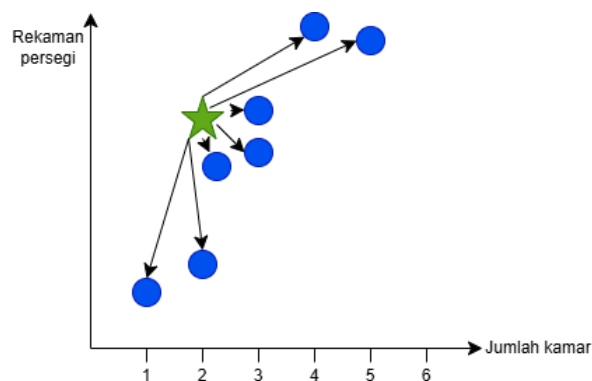
Dalam istilah ML, hal ini diformalkan oleh konsep jarak antara nilai fitur: 1.100–1.000 jauh lebih kecil daripada 12.000–1.000, sehingga algoritma dapat mengetahui bahwa keduanya lebih mirip satu sama lain. Seperti yang telah Anda lihat di bab 2, sebagian besar aplikasi pembelajaran mesin di dunia nyata tidak menggunakan satu fitur saja, melainkan puluhan atau bahkan ratusan. Sekarang mari kita tingkatkan kemampuan kita dan pertimbangkan fitur kedua: jumlah kamar. Jika kita menggambar lingkaran untuk setiap rumah dalam kumpulan data rumah (yang sangat kecil), diagramnya akan terlihat seperti Gambar 6.2.

Bayangkan pengalaman seorang pengguna baru, sebut saja Lucie. Lucie mendaftar di Future-House dan mulai menjelajah. Ia langsung mengklik salah satu rumah dan menghabiskan banyak waktu di halamannya, melihat-lihat gambar, membaca deskripsi, mempertimbangkan ulasan, dan sebagainya. Kita dapat berasumsi bahwa ini merupakan indikasi yang jelas bahwa Lucie menyukai rumah tersebut. Jika kita mengambil plot pada Gambar 6.2 dan mengubah titik yang sesuai dengan rumah favorit Lucie menjadi bintang, kita mendapatkan Gambar 6.3.

Tugas kita adalah membantunya menemukan rumah lain yang sesuai dengan seleranya: rumah-rumah yang mirip dengan rumah yang baru saja ia lihat. Bahkan hanya dengan melihat Gambar 6.4, tampak intuitif bahwa ada tiga rumah lain yang cukup mirip dengan rumah favorit Lucie. Keempat rumah lainnya tampak jauh lebih berbeda: dua rumah di kiri bawah terlalu kecil, dan dua di kanan atas mungkin terlalu besar.



Gambar 6.3 Lucie sangat menyukai salah satu rumah yang ditandai dengan bintang.



Gambar 6.4 Komputer mengambil setiap pasang rumah dan mengukur jarak garis lurus di antara keduanya.

Sekarang kita perlu menerjemahkan penalaran ini ke dalam istilah komputasi agar kita dapat mengotomatiskannya dan membantu ribuan orang seperti Lucie secara bersamaan. Kita akan kembali menggunakan konsep jarak yang telah kita perkenalkan sebelumnya. Bayangkan menggunakan penggaris untuk mengukur jarak antara titik-titik (yaitu, rumah) yang tercetak pada Gambar 6.3: inilah yang dilakukan komputer, seperti yang ditunjukkan pada Gambar 6.4.

Setelah kita mengukur jarak dari rumah berbintang ke semua rumah lain dalam kumpulan data, kita dapat meminta komputer untuk mengurutkannya dari terpendek hingga terpanjang, dan memilih dua atau tiga teratas untuk ditunjukkan kepada Lucie.

Tentu saja, kita masih hanya mempertimbangkan luas persegi dan jumlah kamar, jadi konsep kesamaan kita masih kasar dan tidak lengkap: kita mengabaikan banyak fitur rumah penting lainnya termasuk lokasi, tahun pembangunan, dan banyak lagi. Untungnya, ternyata algoritmanya tidak terlalu berubah ketika kita menambahkan semua fitur lainnya kembali; hanya saja menjadi lebih sulit untuk menggambarinya di atas kertas dan mengonseptualisasikannya bagi kita manusia yang malang dan terbatas. Para insinyur dapat dengan mudah membangun algoritma jarak yang bekerja dengan semua fitur yang kita gunakan di bab 2 untuk prediksi harga, dan komputer dapat menghitungnya dengan mudah.

Apa yang baru saja kita jelaskan adalah sistem rekomendasi paling sederhana yang pernah ada, tetapi tetap efektif, dan konsep dasar ini mudah diterapkan dalam berbagai situasi bisnis. Ambil halaman yang menampilkan detail dan gambar setiap rumah. Para insinyur dapat dengan mudah menambahkan bilah sisi dengan tautan ke tiga atau empat rumah yang paling mirip dengan yang sedang dijelajahi pengguna, dan menikmati lalu lintas tambahan.

Model yang lebih canggih juga melacak bagaimana selera pengguna berkembang seiring waktu, dengan mempertimbangkan semua rumah yang pernah mereka kunjungi. Lebih baik lagi, kita dapat mengelompokkan pilihan berdasarkan kategori barang dan waktu dalam setahun, alih-alih memperlakukan semua preferensi secara sama. Misalnya, sebuah situs e-commerce akan memilih untuk merekomendasikan merek pensil favorit anak-anak Anda selama musim kembali ke sekolah, tetapi menyadari bahwa rekomendasi yang sama tidak berguna selama musim panas.

Perhatikan bahwa sistem rekomendasi Lucie hanya melihat fitur-fitur rumah favorit Lucie dan sama sekali mengabaikan siapa dia dan apa yang telah dilakukan pengguna lain di komunitas. Hal ini juga berlaku untuk contoh-contoh yang lebih canggih yang kami buat: untuk saat ini kami masih berfokus hanya pada konten. Inilah mengapa rangkaian sistem rekomendasi ini disebut berbasis konten.

Sistem Berbasis Konten di Luar Fitur Sederhana

Meskipun contoh FutureHouse fiktif kami sejauh ini telah membantu, Anda mungkin bertanya-tanya bagaimana sistem rekomendasi dapat diadaptasi untuk jenis barang lain, seperti pakaian di toko daring favorit Anda. Berbicara tentang rumah, kami beruntung memiliki serangkaian fitur deskriptif yang juga cukup mudah dikumpulkan. Sekali lagi, fitur-fitur tersebut sama dengan yang kami gunakan di bab 2 untuk memprediksi harga jual rumah. Namun, banyak organisasi menerapkan sistem rekomendasi pada katalog barang yang fiturnya jauh lebih sulit untuk dipilih. Bayangkan contoh di awal bab ini: menemukan kaos terbaik untuk Anda. Fitur apa saja yang bisa kami gunakan?

- Warna dominan (hitam, putih, dan sebagainya)
- Bahan (katun, sintetis)
- Leher berbentuk V atau C
- Ukuran pas (ramping, reguler)

- Lengan (panjang, pendek, tiga perempat).

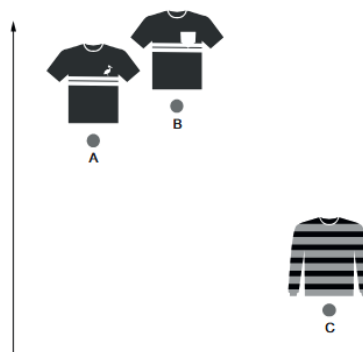
Jika membeli kaus semudah itu, banyak perancang busana akan kehilangan pekerjaan. Kenyataannya, sulit untuk menemukan kriteria yang menggambarkan gaya kaus, seperti halnya pakaian apa pun, atau secara umum media visual apa pun.

Untungnya, kami telah memecahkan tantangan yang sama di bab 4, ketika kami menyadari kekurangan pembelajaran mesin konvensional dan memperkenalkan model canggih berdasarkan pembelajaran mendalam, serangkaian algoritma yang secara otomatis mengekstraksi fitur dari sumber data yang kompleks. Model-model ini mampu mengubah gambar menjadi serangkaian angka kecil yang disebut *embedding*, yang merepresentasikan karakteristik tingkat tinggi gambar dalam bentuk yang ringkas.

Mari kita ingat kembali apa yang telah kita bahas tentang jaringan saraf dalam ketika dilatih untuk mengenali wajah. Ketika Anda melatih algoritma pengenalan wajah, algoritma tersebut belajar mengenali karakteristik tingkat tinggi wajah dan mengubah setiap gambar yang dimasukkan ke dalam serangkaian angka yang disebut *embedding*. *Embedding* ini merepresentasikan keberadaan (atau ketiadaan) karakteristik wajah penting tertentu dalam gambar; misalnya, apakah orang dalam gambar memiliki hidung mancung, telinga besar, dan sebagainya.

Apa yang terjadi jika Anda melakukan hal yang sama dengan kaus? Nah, Anda akan secara ajaib memiliki cara untuk mengubah gambar kaus tersebut dari jutaan piksel menjadi beberapa ratus angka yang mengekspresikan beberapa karakteristik gaya tingkat tingginya. Ini bisa berupa keberadaan garis-garis vertikal, garis-garis horizontal, cetakan berbagai objek, kalimat, logo, dan banyak lagi. Meskipun kita secara intuitif tahu bahwa semua fitur ini relevan untuk mencirikan sebuah kaus, kita tidak memiliki cara untuk melakukannya secara manual. Algoritma pembelajaran mendalam dapat melakukannya secara otomatis, sehingga menghemat waktu kita semua.

Setelah kita mengubah gambar menjadi *embedding*-nya menggunakan jaringan saraf dalam, kita tiba-tiba menemukan diri kita dalam situasi yang sama seperti yang kita jelaskan di bagian sebelumnya: kita dapat mengubah semua gambar menjadi sekumpulan titik di ruang angkasa, dan menggunakan kriteria kesamaan untuk menemukan kaus yang mirip dengan favorit pengguna. Tentu saja, *embedding* akan dibuat dalam ratusan atau ribuan dimensi untuk menangkap nuansa gaya sebuah kaos. Untuk memvisualisasikan konsep ini, kami menggambar representasi 2D *embedding* dari tiga kaos, seperti pada Gambar 6.5.

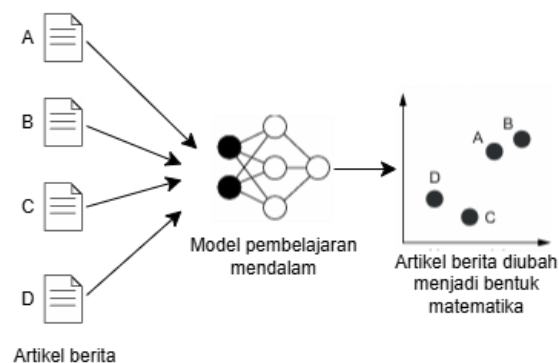


Gambar 6.5 Embedding tiga kaos yang direpresentasikan pada bidang 2-D. Kaos A dan B serupa, sehingga embedding-nya berdekatan. Kaos C berbeda dari A dan B, sehingga jauh dari keduanya.

Bayangkan seorang pengguna mengklik halaman produk kaos A dan menghabiskan waktu melihatnya, memeriksa ukuran, warna, dan bahan yang tersedia. Kita dapat menginterpretasikan hal ini sebagai manifestasi minat terhadap gaya kaos tersebut, dan merekomendasikan kaos lain yang memiliki embedding serupa; misalnya, kaos B. Kita akan menghindari menyarankan kaos C, karena kita tahu embedding-nya jauh dari kaos yang disukai pengguna, dan oleh karena itu mungkin tidak sesuai dengan selera mereka.

Sekali lagi, inilah keajaiban pembelajaran mendalam dan embedding: jaringan saraf tiruan menemukan cara untuk "memahami" gambar dan mengubahnya menjadi sekumpulan angka kecil yang benar-benar masuk akal (setidaknya untuk komputer). Jika dipikir-pikir, gambar bukanlah satu-satunya aplikasi yang menunjukkan kekuatan embedding. Di bab 5, Anda telah melihat betapa hebatnya embedding ketika menangani jenis data rumit lainnya: teks.

Pendekatan yang sama yang kami jelaskan juga berlaku untuk teks. Misalkan kita mencoba menyarankan artikel berita kepada pengguna di situs web, dan kita ingin menemukan berita yang mirip dengan berita yang paling sering digunakan pengguna. Kita dapat mengubah kata-kata di setiap artikel dalam embedding-nya dan menempatkan artikel vektor baru kita di ruang angkasa, seperti yang ditunjukkan pada Gambar 6.6.



Gambar 6.6 Pembelajaran mendalam dapat menerjemahkan setiap artikel berita ke dalam embedding-nya, yaitu deskripsi matematis ringkas yang dapat digunakan untuk membandingkannya.

Setelah kita memiliki representasi spasial dari artikel berita, kita dapat membiarkan konsep kesamaan memandu kita dalam merekomendasikan berita kepada pengguna. Jika pengguna membaca artikel A, kita tahu bahwa mereka kemungkinan besar akan tertarik pada artikel B, karena embedding mereka serupa dan oleh karena itu konten mereka pun serupa. Inilah keajaiban pembelajaran mendalam dan embedding: keduanya memungkinkan kita untuk mengubah bahkan sumber data yang kompleks seperti artikel dan gambar menjadi

representasi numerik yang bermakna, menerapkan teknik yang sama yang kita gunakan untuk rumah di awal bab ini.

Keterbatasan Fitur dan Kesamaan

Dari apa yang telah kita bahas sejauh ini, Anda mungkin berpikir bahwa konsep kesamaan (bersama dengan beberapa keajaiban pembelajaran mendalam) dapat memecahkan setiap masalah rekomendasi, mulai dari artikel berita hingga produk rumah tangga. Bagian ini akan membantu Anda membangun intuisi tentang mengapa, dalam beberapa kasus, kesamaan bukanlah strategi terbaik. Mari kita mulai dengan contoh klasik lain yang sudah tidak asing lagi bagi kita: merekomendasikan film untuk ditonton.

Mari kita atasi masalah baru ini dengan mengikuti pendekatan yang sama seperti yang kita gunakan untuk merekomendasikan rumah kepada calon pembeli. Langkah pertama adalah menemukan fitur-fitur yang dapat digunakan algoritma untuk mendeskripsikan film. Kebanyakan orang membuat daftar seperti ini:

- Sutradara
- Tahun rilis
- Genre (Komedi, Drama, Animasi, Romantis)
- Tiga aktor utama teratas

Jika kita lanjutkan dengan daftar fitur ini, kita akan segera menyadari bahwa rekomendasi yang dihasilkan oleh algoritma tersebut paling banter tidak menarik, dan paling buruk benar-benar konyol. Sepanjang karier mereka, baik aktor maupun sutradara telah mengerjakan berbagai macam film: bahkan jika Anda menyukai *Titanic*, bukan berarti Anda akan menikmati *The Wolf of Wall Street* hanya karena dibintangi Leonardo DiCaprio.

Kita dapat mencoba menyempurnakan model kita dengan memanfaatkan kekuatan pembelajaran mendalam. Kita bisa memasukkan semua dialog film ke model bahasa alami yang dapat menghasilkan fitur-fitur tambahan yang menangkap topik dan suasana film. Ini kemungkinan akan bekerja dengan baik: dialog bertema cinta dan diskusi sang kapten yang mengkhawatirkan akan menambah nuansa pada rekomendasi, yang berpotensi menghasilkan kisah cinta terlarang dan bencana lainnya. Namun, rekomendasi Anda di masa mendatang kemungkinan akan bertahan dalam kategori cinta dan bencana untuk sementara waktu, yang mungkin Anda anggap kurang menarik.

Sayangnya, secanggih apa pun kita dalam membangun fitur, kita masih memiliki batasan mendasar: mesin rekomendasi kita masih berbasis kesamaan. Ini berarti bahwa model tersebut hanya dapat menyarankan item yang sesuai dengan pilihan pelanggan sebelumnya. Ini pasti akan membuat mereka bosan setengah mati: berapa banyak film fantasi Angelina Jolie atau dialog murahan yang dapat Anda tahan sebelum membatalkan langganan Netflix Anda? Dengan demikian, kelemahan rekomendasi berbasis konten adalah bahwa keluarannya bisa menjadi hambar dan dapat diprediksi. Meskipun ini tidak bagus untuk hiburan atau belanja mode, ini sangat cocok untuk beberapa industri lain; Misalnya, rekomendasi obat untuk dokter.

Tidak mengherankan, cara terbaik untuk menambahkan sentuhan manusia pada rekomendasi adalah dengan melibatkan manusia. Dengan menggunakan komunitas pengguna

yang mengonsumsi dan menilai katalog, kita dapat membangun sistem rekomendasi yang lebih bernuansa. Inilah yang akan kita bahas di bagian selanjutnya.

6.3 KEBIJAKSANAAN KERUMUNAN: PENYARINGAN KOLABORATIF

Sebagai recap, jenis sistem rekomendasi yang telah kita bahas sejauh ini umumnya disebut berbasis konten, karena menggunakan atribut deskriptif dari item yang disukai pengguna sebelumnya untuk menemukan dan merekomendasikan item serupa dalam katalog. Namun, pendekatan ini terbagi dalam dua hal: mungkin sulit untuk mengungkapkan fitur yang bermakna, dan rekomendasi yang dihasilkan dapat diprediksi dan membosankan.

Mari kita berhenti sejenak dan pikirkan bagaimana kita mengatasi masalah ini dalam kehidupan nyata. Orang-orang selalu meminta rekomendasi dari teman mereka: buku, restoran, film, dan sebagainya. Seiring waktu, kebanyakan dari kita mengetahui bahwa beberapa teman memiliki selera musik atau makanan yang sama, sehingga kita dengan antusias mendengarkan rekomendasi mereka karena kita mengantisipasi bahwa preferensi serupa yang kita miliki di masa lalu juga akan berlanjut di masa mendatang.

AI dapat memperluas tradisi manusia ini ke kelompok yang jauh lebih besar dengan memanfaatkan komunitas pengguna yang semuanya mengekspresikan preferensi tentang item dalam katalog bersama. Jenis algoritma AI yang dibangun berdasarkan konsep ini disebut algoritma penyaringan kolaboratif: sebuah pendekatan sosial terhadap sistem rekomendasi yang melakukan hal yang hampir sama dengan yang kita lakukan dalam kehidupan nyata dengan teman-teman kita.

Jika kita hanya bisa mengandalkan riwayat selera dan preferensi pengguna yang serupa, peran model kolaboratif adalah mencocokkan pengguna yang memiliki selera yang sama, sehingga mereka dapat saling merekomendasikan. Dalam kehidupan nyata, kita akan mulai saling berbagi riwayat preferensi dan menyadari bahwa keduanya sangat cocok: "Oh, kamu juga suka *Toy Story* dan *Titanic*! Sudahkah kamu menonton..." Model kolaboratif dapat melakukan hal yang sama, dan mencocokkan pengguna dengan selera yang sama, meskipun dalam skala yang jauh lebih besar.

Sama seperti model berbasis konten, Anda harus mengumpulkan preferensi pengguna. Mari kita kembali ke contoh platform real estat kita dan melihat data contoh untuk memahami intuisi di balik model kolaboratif. Tabel 6.1 menunjukkan preferensi tiga pengguna imajiner platform: Alice, Bob, dan Jane.

Tabel 6.1 Preferensi rumah Alice, Bob, dan Jane

Rumah	Alice	Bob	Jane
Rumah A: Satu kamar tidur di lingkungan Dogpatch	?	Menyukai	Tidak seperti
Rumah B: Tiga kamar tidur di Distrik Keuangan	Menyukai	Tidak seperti	Menyukai
Rumah C: Studio di distrik Mission	Tidak seperti	?	Tidak seperti
Rumah D: Satu kamar tidur di distrik Mission	Tidak seperti	Menyukai	Tidak seperti

Rumah E: Tiga kamar tidur di lingkungan Marina	Menyukai	?	Menyukai
--	----------	---	----------

Setiap baris tabel menunjukkan pendapat Alice, Bob, dan Jane tentang masing-masing dari lima rumah di katalog kami (yang memang kecil). Layaknya model berbasis konten, ada banyak cara untuk mengumpulkan data preferensi ini. Cara yang paling mudah adalah dengan bertanya kepada mereka; misalnya, dengan widget bintang lima di halaman deskripsi rumah. Banyak organisasi yang cerdas melangkah lebih jauh dan mencatat informasi yang lebih detail, seperti berapa banyak waktu yang dihabiskan pengguna untuk membaca deskripsi atau menggulir gambar. Jika Bob telah melihat daftar rumah 10 kali selama dua hari terakhir dan menghabiskan lima menit untuk setiap sesi, itu merupakan indikator yang cukup baik bahwa ia benar-benar tertarik.

Dataset contoh ini mencakup rumah-rumah yang tidak sempat dilihat Alice dan Bob (diwakili dengan tanda tanya pada tabel 6.1). Alice tidak pernah menemukan Rumah A, dan Bob tidak melihat Rumah C dan E. Seperti yang telah Anda pelajari, tujuan sistem rekomendasi adalah untuk memilih rumah "yang hilang" mana yang kemungkinan besar mereka minati, sehingga kami dapat menempatkan mereka di bagian atas hasil pencarian dan membantu mereka menemukan rumah baru yang berpotensi mereka beli.

Dengan melihat tabel tersebut, kita sudah dapat mengetahui bahwa Alice dan Jane memiliki minat yang sama: keduanya menyukai Rumah E dan tidak menyukai Rumah C dan D. Di sisi lain, Bob tampaknya memiliki selera yang sangat berbeda dibandingkan dengan Jane dan Alice, karena ia tidak sependapat dengan rumah mana pun yang mereka beri peringkat.

Saat menghasilkan rekomendasi, sistem penyaringan kolaboratif akan memasang Alice dan Jane sebagai dua orang dengan selera yang sama, dan tidak akan menampilkan Rumah A kepada Alice, karena tidak sesuai dengan preferensi Jane. Mengenai Bob, kami tahu bahwa ia mencari sesuatu yang sama sekali berbeda dari Alice dan Jane, sehingga kami menyarankan agar ia memeriksa Rumah C yang tidak disukai Alice dan Jane, dan tidak akan menyarankan Rumah E yang disukai keduanya.

Dengan kata lain, ide di balik penyaringan kolaboratif adalah mengabaikan kesamaan antar item, dan berfokus pada kesamaan antar pengguna, berdasarkan interaksi mereka dengan item dan preferensi mereka. Perhatikan bahwa model ini mengabaikan fitur pengguna, dan hanya mengandalkan peringkat yang diberikan kepada rumah yang dijelajahi. Ini berarti bahwa dalam contoh mainan kami, Alice dan Jane tidak dipasangkan berdasarkan fitur sosial umum seperti jenis kelamin atau usia, tetapi dicocokkan karena selera mereka terhadap rumah.

Keajaiban sistem kolaboratif yang sesungguhnya adalah mereka tidak perlu tahu apa pun tentang item atau pengguna dalam katalog: selama ada pengguna yang menilai sesuatu, kami dapat mengetahui pengguna mana yang memiliki selera serupa dan saling merekomendasikan sesuatu. Model yang sama berfungsi dengan baik untuk rumah seperti halnya film, selama komunitas memberikan peringkat yang cukup. Hal ini juga membuat sistem penyaringan kolaboratif lebih mudah diintegrasikan ke dalam platform yang sudah ada, karena Anda tidak perlu mencari-cari data tambahan tentang setiap item (misalnya, sutradara film),

dan dapat bersantai sementara pengguna sibuk memberi tahu Anda pendapat mereka. Lebih baik lagi, karena rekomendasi ini didasarkan pada preferensi orang lain di dunia nyata, rekomendasi ini bisa terasa baru dan mengejutkan, seperti yang Anda dapatkan dari teman.

Sistem penyaringan kolaboratif hanya berfungsi dengan baik ketika jumlah pengguna jauh lebih banyak daripada item dalam katalog. Jika tidak, sulit untuk memilih kelompok pengguna dengan selera yang sama: katalog besar dengan banyak item yang belum pernah dinilai (atau hanya sekali) tidak akan efektif. Misalnya, mari kita asumsikan kasus hiperbolik di mana situs web real estat Anda hanya memiliki Bob, Jane, dan Alice sebagai pengguna, tetapi 500 rumah terdaftar untuk dijual. Sangat kecil kemungkinannya bahwa ketiga pengguna kita yang kesepian telah melihat rumah yang sama, yang berarti kita tidak dapat menggunakan trik cerdas pencocokan selera yang telah kita jelaskan.

Contoh perusahaan yang sangat tepat memanfaatkan penyaringan kolaboratif adalah Netflix, yang memiliki puluhan juta pengguna yang berinteraksi hanya dengan ratusan film. Di sisi lain, bayangkan diri Anda di posisi Pinterest: meskipun memiliki ratusan juta pengguna, konten yang tersedia bagi para pengguna untuk berinteraksi pada dasarnya adalah seluruh web: ratusan miliar gambar. Dalam kasus Pinterest, pendekatan terbaik adalah menggunakan rekomendasi berbasis konten. Inilah yang telah dilakukan perusahaan tersebut: pada Januari 2017, perusahaan mengumumkan bahwa sistem pembelajaran mendalam barunya yang menyediakan rekomendasi pin berbasis konten meningkatkan keterlibatan pengguna sebesar 30% dalam semalam.

Poin penting lainnya adalah penyaringan kolaboratif berfungsi ketika Anda memiliki interaksi pengguna-item yang cukup untuk membangun model Anda. Jika Anda adalah perusahaan baru dengan riwayat interaksi yang singkat, Anda mungkin harus meninggalkan pendekatan ini sama sekali. Dalam hal ini, yang dapat Anda lakukan hanyalah memulai dengan sistem berbasis konten sehingga Anda dapat mulai memberikan rekomendasi tanpa basis data besar berisi peringkat yang ada. Seiring pertumbuhan basis pengguna, dan peringkat mulai berdatangan, Anda dapat memperkenalkan komponen kolaboratif untuk memberikan rekomendasi yang lebih unik. Seperti yang akan Anda lihat dalam contoh, organisasi terbaik di kelasnya sebenarnya menggabungkan kedua pendekatan tersebut, memberikan beragam rekomendasi yang dapat menarik bagi sebagian besar basis pengguna mereka.

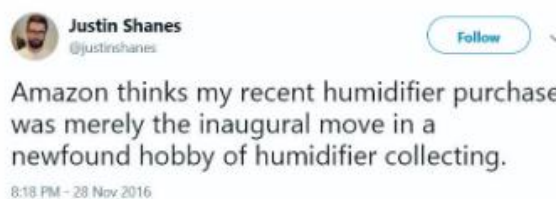
6.4 REKOMENDASI YANG SALAH

Banyak tugas berbasis AI yang telah kita bahas sejauh ini memiliki definisi kinerja yang lugas yang biasanya kita terjemahkan ke dalam bentuk akurasi untuk membandingkan model dalam situasi yang berbeda. Misalnya, kita telah membahas tentang akurasi dalam memprediksi harga jual rumah, atau dalam mengklasifikasikan anjing dan kucing dalam gambar. Kita dapat dengan mudah menerapkan pendekatan yang sama untuk sistem rekomendasi misalnya, dengan menyisihkan serangkaian uji peringkat dan mengukur akurasi prediksinya. Namun, pendekatan sederhana ini mengabaikan realitas kompleks selera dan preferensi manusia, serta pengaruhnya terhadap kinerja bisnis.

Masalah pertama berkaitan dengan pengalaman pengguna. Rekomendasi otomatis umumnya berfungsi dengan baik, tetapi terkadang kita bertanya-tanya bagaimana saran yang konyol bisa muncul. Ruang di layar terbatas, begitu pula perhatian pengguna; Anda tentu tidak ingin menyia-nyiakannya dengan menampilkan produk yang tidak berguna.

Seperti yang telah Anda lihat, sebagian besar sistem rekomendasi mutakhir beroperasi berdasarkan berbagai sinyal, baik berdasarkan fitur produk maupun peringkat komunitas. Karena selera manusia cukup sulit diprediksi, satu saran yang tidak tepat sasaran saja dapat berdampak kuat pada kepercayaan pelanggan, sehingga secara drastis kurang efektif dalam mendorong nilai bisnis. Bergantung pada spesifikasi komunitas, sistem rekomendasi mungkin salah mengira fitur sepele (seperti tahun rilis film) sebagai fitur penting, yang mengarah pada rekomendasi yang mengejutkan (dan mengecewakan).

Selain itu, suatu strategi dapat berfungsi dengan baik untuk suatu kategori produk tetapi sama sekali tidak berguna untuk kategori lain: mungkin menyenangkan melihat kaos yang mirip dengan yang saya beli, tetapi jangan penuhi layar saya dengan AC yang mirip dengan pembelian terakhir saya (Gambar 6.7 menunjukkan laporan dari laporan).



Gambar 6.7 Seorang pengguna Amazon tidak terlalu terkesan dengan rekomendasi yang diterimanya.

Meskipun pengalaman pengguna merupakan pertimbangan penting, akan sangat bodoh jika mengabaikan dampak besar sistem rekomendasi dalam membentuk opini dan wacana publik. Sebagian besar, jika tidak semua, media yang kita konsumsi (baik berita, video, maupun unggahan media sosial) pada suatu saat telah disaring oleh sistem rekomendasi. Ini berarti bahwa teknologi ini diposisikan secara unik untuk memengaruhi pandangan dunia kolektif kita. Kita akan membatasi pembahasan hanya pada implikasi bisnis dari teknologi ini untuk saat ini, dan mengeksplorasi dampak yang lebih luas terhadap masyarakat di bab 10.

Impian Sistem Rekomendasi

Mari kita akhiri bagian ini dengan beberapa hal untuk direnungkan. Mari kita asumsikan bahwa saat ini, 20% rekomendasi Amazon dikonversi menjadi pembelian. Misalkan salah satu ilmuwan data Amazon menemukan cara cerdas untuk meningkatkan algoritma rekomendasinya sedemikian rupa sehingga sekarang 100% rekomendasi berubah menjadi pembelian. Mengapa Amazon harus menunggu Anda berbelanja online padahal mereka sudah tahu persis apa yang ingin Anda beli? Solusi termudah adalah langsung mengirimkan barang ke rumah Anda, dengan kartu pos cantik bertuliskan "Sama-sama."

Tentu saja, tingkat konversi 100% terasa terlalu berlebihan. Namun, konsep ini masih valid: secara teoritis, terdapat ambang batas di mana rekomendasi produk menjadi begitu

efektif sehingga pengguna akan lebih mudah mengembalikan produk yang tidak mereka sukai atau butuhkan daripada berbelanja secara proaktif. Dalam skenario seperti itu, Amazon bisa saja mengirimkan 10 produk kepada kita, dan kita akan mengirimkan kembali produk yang salah dari model tersebut.

Ini mungkin terdengar futuristik, tetapi perusahaan-perusahaan fesyen sudah bereksperimen dengan model ini. Bayangkan: meningkatkan satu teknologi saja dapat menyebabkan disrupsi total terhadap model bisnis dan kebiasaan pembelian pelanggan. Inilah kekuatan sistem rekomendasi dan banyak teknologi AI lainnya.

6.5 STUDI KASUS: NETFLIX MENGHEMAT RP. 10 TRILIUN PER TAHUN

Ketika Netflix didirikan pada tahun 1997, perusahaan ini mulai mengoperasikan model berlangganan untuk mengirimkan DVD fisik ke rumah tangga, yang kemudian mengembalikan media tersebut melalui pos setelah menonton filmnya. Selama dekade berikutnya, ketersediaan akses internet cepat dan perangkat pribadi dalam skala besar memungkinkan perusahaan untuk beralih ke distribusi konten digital. Dengan biaya berlangganan bulanan tunggal, pelanggan dapat menonton film dan serial TV sebanyak yang mereka inginkan, menggunakan ponsel pintar, tablet, laptop, atau TV pintar.

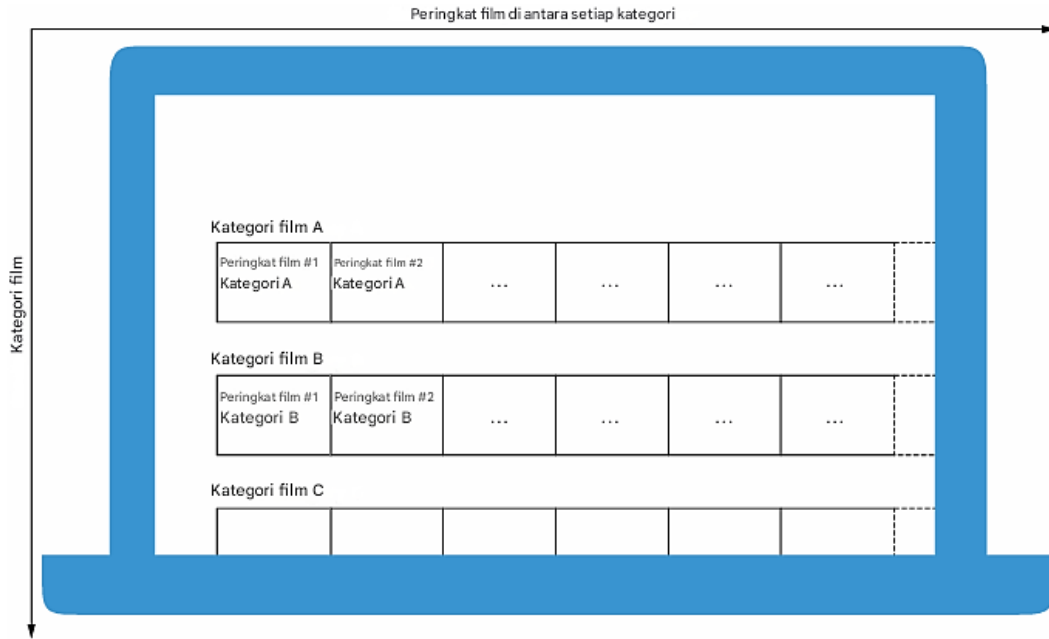
Kenyamanan dan harga yang lebih rendah dibandingkan dengan media fisik atau TV kabel menjadikan Netflix salah satu perusahaan dengan pertumbuhan tercepat dalam beberapa tahun terakhir. Dari tahun 2005 hingga 2018, pendapatan tumbuh dari Rp. 6,82 triliun menjadi hampir Rp. 160 triliun, dan harga saham naik hampir 150 kali lipat, dari Rp. 20.000 per saham pada awal tahun 2005 menjadi hampir Rp. 3.000.000 pada akhir tahun 2018. Pada tahun 2018, perusahaan ini memiliki 150 juta pelanggan dari lebih dari 190 negara.

Penggerak utama pertumbuhan eksplosif tersebut tentu saja adalah konsep baru TV berbasis internet dengan pilihan yang tampaknya tak terbatas. Namun, peluang baru membawa serta masalah baru yang belum pernah terjadi sebelumnya. Manusia pada umumnya buruk dalam membuat keputusan ketika dihadapkan dengan terlalu banyak pilihan, dan memilih acara TV untuk ditonton tidak terkecuali. Riset konsumen melaporkan bahwa pengguna Netflix pada umumnya akan berhenti menonton setelah hanya 60 detik dihabiskan untuk memilih judul baru untuk ditonton.

Untuk menghadapi tantangan baru ini, perusahaan berinvestasi besar-besaran dalam sistem rekomendasi untuk membantu pengguna menemukan konten yang menarik dalam katalog Netflix yang luas.

Sistem Rekomendasi Netflix

Meskipun konsep TV internet yang dipopulerkan Netflix menghadirkan tantangan baru, konsep ini juga menghadirkan peluang baru untuk mengumpulkan data. Berkat infrastruktur sesuai permintaan, Netflix memiliki akses ke data dalam jumlah besar tentang aktivitas setiap pengguna: apa yang mereka tonton, di mana dan kapan mereka masuk, dan sebagainya.



Gambar 6.8 Komposisi halaman beranda Netflix

Cara terpenting sistem rekomendasi Netflix berinteraksi dengan pelanggan adalah halaman beranda yang mereka kunjungi tepat setelah masuk (lihat gambar 6.8). Meskipun antarmukanya berubah tergantung jenis perangkat, titik awalnya adalah kumpulan sekitar 40 baris yang mewakili berbagai kategori film, dan hingga 75 item per baris. Untuk menentukan cara mengisi ruang layar ini, Netflix memasukkan semua data yang dikumpulkannya ke dalam sekelompok algoritma, yang masing-masing memiliki spesialisasi dalam tugas rekomendasi yang berbeda. Bersama-sama, algoritma ini membentuk sistem rekomendasi Netflix.

Inti dari sistem ini adalah algoritma yang disebut *Personalized Video Ranker* (PVR). PVR, berdasarkan sejumlah besar data yang dikumpulkan Netflix tentang kebiasaan menonton pengguna, digunakan untuk memperkirakan kemungkinan pengguna akan menonton film di semua kategori film yang berbeda (misalnya, Thriller atau Drama). Halaman beranda Netflix juga memiliki baris khusus lain yang memiliki algoritme rekomendasi spesifiknya sendiri. Salah satu contohnya adalah baris Pilihan Teratas, yang menyarankan konten terbaik yang dipilih di semua kategori. Algoritme ini disebut Top N Video Ranker. Sementara PVR digunakan untuk menemukan film terbaik dalam subset katalog tertentu, Top N ranker memeriksa seluruh katalog.

Baris khusus lainnya adalah untuk konten Trending Now. Netflix menggabungkan rekomendasi yang dipersonalisasi dengan sinyal yang berasal dari tren temporal. Secara khusus, perusahaan telah menemukan dua jenis faktor yang memengaruhi perilaku pengguna. Faktor pertama terkait dengan tren yang berulang, seperti Natal atau Hari Valentine. Jenis tren lainnya terkait dengan peristiwa jangka pendek, seperti pemilu atau bencana alam yang menarik minat. Bahkan dalam kasus ini, Netflix tidak menampilkan baris yang sama kepada setiap pengguna, melainkan menggabungkan sinyal dari tren keseluruhan dengan item yang dipersonalisasi.

Baris khas lainnya adalah Lanjutkan Menonton. Sementara semua baris lainnya berfokus pada konten yang belum pernah ditonton pengguna, baris ini berfokus pada konten yang sudah mulai dikonsumsi pengguna tetapi tidak pernah selesai. Peringkat Lanjutkan Menonton memilih judul dengan probabilitas tertinggi untuk diselesaikan, menggunakan fitur-fitur khusus aplikasi berikut:

- ❖ Waktu yang telah berlalu sejak penayangan terakhir
- ❖ Perangkat
- ❖ Titik penghentian (tengah program versus awal atau akhir)
- ❖ Judul lain yang ditonton sejak penayangan terakhir

Tabel 6.2 Algoritma yang menyusun sistem rekomendasi Netflix

Algoritma	Menggunakan	Kriteria
Pemeringkat Video yang Dipersonalisasi	Jika diberikan kategori film, pilih film yang paling mungkin ditonton pengguna dalam kategori tersebut.	Pemfilteran kolaboratif
Peringkat Video N Teratas	Di antara semua film dalam katalog, pilih yang terbaik untuk pengguna tertentu.	Pemfilteran kolaboratif
Sedang Tren Sekarang	Berdasarkan berbagai tren temporal, pilih film yang sejalan dengan tren dan yang paling mungkin ditonton oleh pengguna.	Pemfilteran kolaboratif
Lanjutkan Menonton	Jika ada koleksi film yang telah dimulai pengguna tetapi tidak selesai, pilih film yang paling mungkin untuk dilanjutkan.	Pemfilteran kolaboratif
Kemiripan Video-Video	Diberikan sebuah film, temukan film yang paling mirip.	Berbasis konten (kesamaan video)
Pembuatan Halaman	Pilih baris mana yang akan ditampilkan untuk pengguna tertentu dan dalam urutan apa.	Pemfilteran kolaboratif

Semua algoritma yang telah kita lihat sejauh ini mengandalkan teknik penyaringan kolaboratif untuk memilih saran mereka menyimpulkan selera pengguna dengan melihat pilihan pengguna lain yang memiliki pola menonton serupa (dan karenanya memiliki selera serupa). Baris Karena Anda Menonton merupakan pengecualian, karena baris tersebut menawarkan konten yang serupa dengan yang telah dinikmati pengguna sebelumnya. Tugas ini dilakukan oleh algoritma Video-Video Similarity, sebuah sistem rekomendasi berbasis konten. Algoritma ini tidak mempertimbangkan selera pengguna: hasilnya berupa peringkat kemiripan antara pilihan pengguna dan katalog Netflix lainnya. Namun, pilihan film yang akan menggunakan algoritma Video-Video Similarity dipersonalisasi berdasarkan selera pengguna.

Selain rangkaian algoritma ini, pada tahun 2015 Netflix memperkenalkan algoritma Page Generation untuk menentukan baris mana yang akan ditampilkan kepada pengguna tertentu dalam situasi yang berbeda. Tabel 6.2 merangkum algoritma Netflix yang membentuk sistem rekomendasinya.

Rekomendasi dan Pengalaman Pengguna

Sistem rekomendasi Netflix merupakan cara utama pelanggan berinteraksi dengan layanan, dan karenanya telah bergeser untuk mengakomodasi evolusi platform itu sendiri.

Selama bertahun-tahun Netflix beroperasi dengan mengirimkan DVD fisik kepada pelanggan, siklus umpan balik tonton/beri peringkat/tonton lagi yang mendorong keterlibatan pelanggan jauh lebih lambat daripada saat ini. Pelanggan hanya akan menerima konten baru seminggu sekali, dan pengiriman mingguan sebaiknya mencakup sesuatu yang menghibur mereka pada Jumat malam.

Dengan streaming daring, aturan mainnya berubah: karena pengguna dapat memulai dan berhenti mengonsumsi konten apa pun kapan saja, rekomendasi dapat lebih fleksibel. Pada saat yang sama, riset pengguna yang ekstensif telah mengungkapkan bahwa pendorong retensi terpenting di era streaming adalah jumlah waktu yang dihabiskan pelanggan untuk menonton konten. Tujuan sistem rekomendasi Netflix pun berkembang. Kritik terhadap perubahan (bertahap) ini menunjukkan bahwa sistem ini terlalu condong ke arah konten yang "cukup baik" dengan tujuan tunggal untuk membuat pelanggan tetap terhubung selama beberapa jam sehari. Saran yang lebih berisiko, dengan risiko kekecewaan yang lebih tinggi, justru kurang tertimbang.

Algoritma apa pun yang digunakan oleh ratusan juta orang di seluruh dunia tidak luput dari kritik, dan sistem rekomendasi Netflix pun demikian. Meskipun rekomendasi yang sangat salah sering beredar di media sosial, kesalahan prediksi yang lebih halus juga mengungkapkan masalah yang belum terpecahkan di ranah tersebut. Pengguna sering mengeluh tentang model yang tidak dapat melacak beberapa "suasana hati" dan kecenderungan konsumsi media yang independen. Contoh tipikal adalah pesta nonton TV sampah Jumat malam yang memengaruhi rekomendasi untuk dua bulan berikutnya.

Nilai Bisnis dari Rekomendasi

Pada tahun 2015, Chief Product Officer Netflix, Neil Hunt, dan Vice President of Product Innovation, Carlos Gomez-Uribe, menulis "The Netflix Recommender System: Algorithms, Business Value, and Innovation" (<https://dl.acm.org/citation.cfm?id=2843948>). Mereka melaporkan bahwa sistem rekomendasi Netflix bertanggung jawab atas 80% jam streaming di platform tersebut, dengan 20% sisanya berasal dari fungsi pencarian. Namun, karena pengguna sering mencari judul yang tidak ada dalam katalog, sebagian dari 20% itu menjadi masalah rekomendasi.

Saat mengevaluasi efektivitas algoritma, Netflix bergantung pada metrik tertentu. Dua yang paling penting adalah ukuran katalog efektif (ECS) dan tingkat pengambilan. ECS adalah metrik yang menggambarkan seberapa tersebar penayangan di antara item dalam katalog. Metrik ini memiliki formulasi matematis, tetapi ide utamanya cukup sederhana: jika sebagian besar penayangan berasal dari satu video, nilainya akan mendekati 1. Jika semua video menghasilkan jumlah penayangan yang sama, ECS mendekati jumlah video dalam katalog. Jika katalog memiliki beberapa film yang jarang ditonton dan beberapa yang lebih populer, ECS berada di antara keduanya. Semakin tinggi nilai ini, semakin banyak orang yang menonton semua film dalam katalog, yang menghasilkan kepuasan pelanggan yang lebih tinggi dan peningkatan laba atas uang yang dibelanjakan Netflix untuk memperoleh hak distribusi media atau memproduksi.

Oleh karena itu, tujuan Netflix adalah memaksimalkan nilai ini. Untuk menguji efektivitas sistem rekomendasi, Netflix mencoba menyajikan konten kepada pengguna dengan mengikuti dua pendekatan berikut:

- *Menggunakan metrik popularitas* —Dimulai dengan film terpopuler dalam katalog dan secara bertahap menambahkan film populer lainnya (garis hitam)
- *Menggunakan sistem personalisasi* —Dimulai dengan film yang menempati peringkat pertama dalam skor PVR dan menambahkan film lain mengikuti urutan yang ditentukan oleh algoritma PVR

Penggunaan sistem personalisasi menghasilkan peningkatan ECS sebesar 400%.

Metrik keterlibatan utama lainnya yang digunakan oleh Netflix adalah tingkat penerimaan, yang didefinisikan sebagai fraksi rekomendasi yang ditawarkan yang menghasilkan pemutaran. Nilai yang mendekati 1 berarti rekomendasi tersebut akan dipilih 100% dari waktu, sementara nilai yang mendekati 0 berarti rekomendasi tersebut tidak akan dipilih. Netflix mencoba menawarkan serangkaian rekomendasi kepada pengguna dengan mengikuti dua pendekatan:

- Menampilkan N film terpopuler pertama
- Menampilkan N film yang, menurut sistem rekomendasi, merupakan pilihan terbaik bagi pengguna

Dengan membandingkan kedua pendekatan tersebut, Netflix memperhatikan bahwa sistem personalisasi memberikan peningkatan substansial dibandingkan pendekatan "terpopuler", dengan peningkatan tingkat penerimaan hampir empat kali lipat untuk film "paling sesuai" dibandingkan film "terpopuler", dan penurunan kinerja seiring dengan beralih ke film yang dinilai kurang sesuai.

Peningkatan rekomendasi dan metrik ini berkaitan erat dengan peningkatan kinerja bisnis. Untuk layanan berbasis langganan murni seperti Netflix, tiga metrik utama yang perlu diperhatikan adalah tingkat perolehan anggota baru, tingkat pembatalan anggota (tingkat churn), dan tingkat kembalinya mantan pelanggan. Hunt dan Uribe melaporkan bahwa upaya Netflix dalam personalisasi dan rekomendasi telah mengurangi tingkat churn hingga beberapa poin persentase. Secara keseluruhan, mereka memperkirakan bahwa upaya gabungan dari semua algoritma rekomendasi mereka telah menghemat Netflix lebih dari Rp. 10 triliun per tahun, pada saat pendapatan keseluruhan berada di bawah Rp. 70 triliun.

PERTANYAAN KASUS

- Apakah katalog produk yang lebih besar jelas lebih baik bagi konsumen? Dan bagi bisnis?
- Bagaimana sistem rekomendasi berevolusi seiring peralihan dari media fisik, lalu streaming daring, dan akhirnya produksi orisinal?
- Selain rekomendasi, apa saja cara penting lain Netflix dapat memanfaatkan keunggulan datanya saat bersaing dengan platform distribusi mapan (misalnya, TV, bioskop, media fisik)?

DISKUSI KASUS

Netflix mungkin merupakan salah satu perusahaan yang paling mengguncang industri hiburan. Kesuksesannya merupakan salah satu alasan utama di balik kegagalan raksasa penyewaan DVD Blockbuster, dan peralihannya ke konten orisinal menimbulkan ancaman bagi produsen konten seperti jaringan TV dan studio film. Berawal dari layanan penyewaan DVD berbasis internet, perusahaan ini benar-benar mulai berkembang pesat ketika memelopori konsep TV internet. Alternatif utama untuk TV internet pada saat itu adalah siaran linear dan sistem kabel. Meskipun keduanya memiliki jadwal tetap, TV internet menempatkan pengguna di posisi yang menentukan, memungkinkan mereka memilih konten apa pun yang mereka inginkan, kapan pun mereka mau.

Kekuatan pilihan tampaknya menjadi daya tarik tersendiri bagi bisnis apa pun. Namun, seperti yang ditunjukkan Hunt dan Gomez-Uribe, orang-orang ternyata tidak terlalu pandai memilih, terutama di antara banyaknya pilihan. Apa gunanya memiliki banyak koleksi film jika pengguna tidak dapat menemukan apa pun yang mereka sukai? Hal ini tidak hanya berlaku untuk industri hiburan. Bayangkan e-commerce apa pun: apa nilai katalog merek besar jika pengguna tidak membeli apa pun?

Masalah utamanya di sini adalah meskipun penawaran suatu perusahaan berpotensi tak terbatas, orang-orang hanya memiliki sedikit waktu untuk menelusurinya. Dalam kasus Netflix, perusahaan tersebut melaporkan bahwa sebagian besar pengguna akan menyerah dalam pencarian film mereka jika mereka tidak menemukan apa pun dalam 60 hingga 90 detik. Artinya, tanpa membantu pengguna dengan cara apa pun, menambahkan elemen ke katalog justru dapat mempersulit pengguna menemukan sesuatu yang mereka sukai, dan justru kontraproduktif dari segi pengalaman pengguna dan bisnis. Inilah nilai sistem rekomendasi: memastikan pengguna dapat menemukan nilai dalam penawaran Anda dengan pendekatan berbasis data terhadap saran.

Salah satu aset utama internet yang telah digunakan Netflix adalah kesempatan untuk mengumpulkan data dan memantau preferensi pengguna. Bayangkan sejenak perbedaan antara versi sewa DVD Netflix dan versi TV internet. Dalam kasus pertama, setelah DVD dikirimkan, perusahaan hanya memiliki satu cara untuk mengetahui apakah pengguna menyukainya: skor yang mereka berikan (jika mereka mau memberikannya). Dengan layanan streaming, Netflix dapat mencatat informasi yang jauh lebih lengkap, seperti berikut:

- ✚ Kapan pengguna menonton film (waktu, hari)
- ✚ Di perangkat apa
- ✚ Apakah pengguna menonton semuanya dalam satu sesi atau menjedanya beberapa kali?

Selain itu, Netflix dapat menguji berbagai versi algoritma rekomendasi dan mengukur kinerjanya berdasarkan metrik tertentu. Melimpahnya data merupakan aset utama yang memungkinkan Netflix tidak hanya mendorong keterlibatan pengguna, tetapi juga mengukur kinerjanya dan menyesuaikan produknya.

- ❖ Sistem rekomendasi menawarkan rekomendasi yang dipersonalisasi kepada pelanggan, memungkinkan mereka menjelajahi katalog produk atau layanan yang lebih luas, dan meningkatkan keterlibatan.
- ❖ Sistem rekomendasi berbasis konten menggunakan riwayat pengguna (halaman yang mereka kunjungi, produk yang mereka beli, dan sebagainya) untuk merekomendasikan item serupa dari katalog.
- ❖ Penyaringan kolaboratif adalah pendekatan lain yang menemukan pengguna dalam komunitas yang memiliki selera serupa, dan berbagi rekomendasi di antara mereka.
- ❖ Sistem rekomendasi merupakan landasan platform e-commerce dan distribusi media modern, seperti yang Anda lihat dalam studi kasus Netflix.

BAB 7

SIAP MENEMUKAN PELUANG AI

Bab ini membahas

- Memahami perbedaan antara visi AI dan proyek AI
- Mencari peluang AI di organisasi Anda
- Memprioritaskan dan menguji ide proyek AI Anda
- Memecah produk AI kompleks di dunia nyata menjadi komponen-komponen independen bertenaga ML
- Menerjemahkan kebutuhan bisnis ke dalam tugas-tugas ML menggunakan Kanvas Framing

Dalam bab ini, Anda akan mengambil langkah pertama memasuki dunia inovasi AI yang kompleks. Bayangkan pembelajaran mesin sebagai kotak peralatan yang penuh dengan berbagai benda seperti palu, obeng, dan gergaji. Setelah membaca bab sebelumnya, Anda akan mengetahui bagaimana semua alat ini dapat digunakan untuk membangun produk. Langkah Anda selanjutnya adalah mencari tahu apa yang akan dibangun dengan alat-alat ini dan bagaimana caranya.

Bab ini dimulai dengan menyajikan kerangka kerja yang kami kembangkan untuk mengidentifikasi, memilih, dan memvalidasi peluang AI yang paling menjanjikan di organisasi Anda. Kami juga akan menunjukkan cara memecah produk yang kompleks menjadi komponen-komponen yang lebih ramah ML yang dapat dibangun secara independen. Terakhir, Kanvas Pembimbingan kami akan mengajarkan Anda cara menerjemahkan hasil dari semua pekerjaan awal ini menjadi deskripsi proyek Anda yang dapat digunakan oleh tim teknis.

7.1 WASPADA HYPE: INOVASI AI UNTUK BISNIS

Pada bab sebelumnya mengikuti kisah FutureHouse, sebuah platform real estat fiktif yang memungkinkan pengguna untuk membeli dan menjual rumah secara daring. Kami menciptakan fitur-fitur AI yang canggih dan membayangkan bagaimana fitur-fitur tersebut dapat membantu FutureHouse menarik lebih banyak pengguna. Sementara itu, kami mencoba menjelaskan konsep teknis di baliknya.

Kami tidak memfokuskan perhatian kami pada kisah di balik layar: bagaimana FutureHouse memunculkan ide-ide ini? Bagaimana mereka memilih yang terbaik untuk dikerjakan? Kecuali Anda bekerja untuk raksasa teknologi, kemungkinan besar organisasi Anda baru saja mengambil langkah pertama dalam perjalanan AI-nya, dan Anda mungkin juga menghadapi pertanyaan-pertanyaan ini.

Inilah sebabnya kami memutuskan untuk kembali ke awal perjalanan AI FutureHouse dan membayangkan percakapan yang terjadi di ruang rapatnya. Mereka akan memberi Anda skenario untuk dipegang sementara kami mengungkap praktik terbaik untuk memperkenalkan AI dalam sebuah organisasi. Mari kita mulai dengan rapat awal fiktif:

- CEO : AI sedang ramai diberitakan akhir-akhir ini; sepertinya akan memainkan peran besar dalam bisnis apa pun. Kita jelas perlu mulai mempelajarinya sebelum pesaing kita atau perusahaan rintisan teknologi lainnya melakukannya terlebih dahulu.
- CTO : Anda benar; perkembangannya sangat cepat, dan kita perlu bersiap. Saya sarankan kita mulai berinvestasi dalam infrastruktur agar kita dapat memproses semua data kita. Kemudian kita dapat melatih teknisi kita atau merekrut ilmuwan data baru dan mulai membangun algoritma AI.
- Manajer Pemasaran : Sejujurnya, teman-teman, saya pikir ini hanya tren sesaat. Sekarang terlihat keren untuk mengatakan "Kami menggunakan AI," tetapi pada akhirnya akan memudar.
- Pimpinan AI : Saya setuju bahwa beberapa orang terlalu cepat mengikuti tren ini, tetapi saya yakin mengabaikan AI sama sekali adalah tindakan yang picik. Tentu saja, ini bukan solusi instan; kita butuh strategi untuk memastikan kita bisa mendapatkan nilai dari hal-hal yang AI kuasai.
- CEO : Mmm, tapi saya tidak yakin saya sepenuhnya paham apa saja yang AI kuasai. Ada begitu banyak kebingungan di media sehingga saya kesulitan memahami mana yang nyata dan mana yang hanya sensasi.
- Pemimpin AI : Menurut saya, ini hal pertama yang perlu kita perbaiki. Bagaimana kalau kita menyelenggarakan pelatihan untuk tim eksekutif agar kita semua lebih memahami apa yang sebenarnya bisa dilakukan AI, lalu kita merencanakan langkah selanjutnya?
- CTO : Saya pikir itu ide yang bagus; ini juga akan membantu tim teknis dan pebisnis belajar bekerja sama.
- Manajer Pemasaran : Oke, mari kita lakukan. Mungkin ini akan membantu saya mengubah pikiran saya tentang teknologi ini. Tapi saya pikir penting bagi kita untuk memulai dengan strategi. Tanpa visi AI, kita akan kehilangan waktu dan uang.
- Pemimpin AI : Senang Anda semua bersemangat mengikuti pelatihan AI, dan saya sangat setuju dengan Anda bahwa kita memang perlu memiliki visi. Di sisi lain, saya rasa kita membutuhkan lebih banyak pengalaman untuk membangun visi AI. Pelatihan saja tidak akan cukup: kita perlu mulai terjun langsung sebelum meluangkan waktu untuk perencanaan jangka panjang. Saya sarankan setelah pelatihan, kita mencari beberapa proyek AI untuk mulai bereksperimen. Proyek-proyek ini akan memandu kita dalam mendefinisikan visi AI kita.
- CEO : Saya suka arah perkembangannya. Apakah kita semua sepakat untuk mendapatkan pelatihan AI dan mulai bereksperimen dengan teknologi ini? Saya ingin uji cobanya berjalan empat bulan dari sekarang; lalu kita bisa

berhenti sejenak, melihat apa yang telah kita pelajari, dan mulai membahas strategi dan visi kita secara keseluruhan. Bagaimana menurut Anda?

Percakapan ini mencakup banyak elemen organisasi yang baru memulai dengan AI. Memang, beberapa bagian dari naskah ini telah terekam di depan mata kita berkali-kali selama pengalaman konsultasi kita. Mari kita lihat bagian terpentingnya. Pertama, kami ingin Anda berfokus pada orang-orangnya. Ini adalah percakapan, bukan monolog atau strategi yang dipaksakan oleh seseorang. Untuk mewujudkan perubahan yang bermanfaat, Anda membutuhkan tim, bukan orang yang bekerja sendiri.

Selanjutnya, naskah ini memiliki empat karakter dengan karakteristik spesifik:

- ❖ *Pemrakarsa* — Seseorang yang memulai diskusi. Bisa jadi seseorang yang pernah mendengar tentang AI di berita, seseorang yang berpengalaman dan melihat peluang, atau hanya seseorang yang ingin memulai proyek baru untuk memajukan karier mereka. Dalam kasus kami, kami membayangkan orang ini adalah CEO perusahaan. Ini adalah skenario ideal, tetapi pemrakarsanya juga bisa seorang manajer atau karyawan.
- ❖ *Teknisi* — Seorang teknolog yang memiliki pendekatan teknologi pertama terhadap AI. Dalam kasus kami, CTO. Orang seperti ini biasanya cenderung melihat aspek teknis AI dan melupakan dampak bisnisnya. Penting untuk menyeimbangkan perspektif teknisi dengan perspektif orang yang lebih berorientasi bisnis.
- ❖ *Orang yang skeptis* — Seseorang yang sama sekali tidak percaya bahwa AI memiliki nilai. Seringkali, orang-orang ini berargumen bahwa AI hanyalah tren sesaat atau "hanya statistik." Terkadang orang-orang ini akan tertarik pada AI setelah mempelajarinya lebih lanjut. Terkadang mereka perlu melihat hasilnya dengan mata kepala sendiri. Dalam kasus kami, orang yang skeptis adalah seorang manajer pemasaran, tetapi bisa juga siapa pun di dalam organisasi.
- ❖ *Pemimpin AI* — Ini adalah orang yang memahami AI, diharapkan memiliki pengalaman dengannya, dan memastikan bahwa teknologi tersebut digunakan dengan cara yang memberikan nilai bagi bisnis. Mereka memiliki pengetahuan yang mendalam tentang bisnis dan mengetahui prinsip-prinsip inti AI untuk menemukan titik temu antara keduanya. Semoga, orang ini bisa menjadi Anda setelah membaca buku ini dan memperoleh pengalaman.

Elemen penting dalam percakapan ini adalah pemimpin AI. Dalam beberapa saat, percakapan ini bisa saja salah arah, yang membawa organisasi ke jalan buntu potensial dalam upaya AI-nya. Beruntung bagi perusahaan fiktif kita, pemimpin AI turun tangan dan mengusulkan strategi yang baik untuk mencegah hal ini terjadi.

Alarm pertama dari percakapan ini adalah tidak ada satu pun peserta (inisiator, skeptis, dan teknisi) yang memiliki pengetahuan AI yang cukup untuk membuat keputusan yang sadar. Dalam situasi seperti ini, risikonya adalah organisasi akan mengikuti siapa pun yang bersuara paling lantang, meskipun mereka memimpin tim secara membabi buta. Lebih parah lagi, skeptis mungkin mengajukan keberatan yang didasarkan pada prasangka dan bukan pengetahuan yang sebenarnya. Ini adalah resep bencana.

Untungnya, pemimpin AI mengusulkan untuk menjalankan pelatihan nonteknis bagi semua pengambil keputusan. Pendidikan adalah elemen fokus kedua dari percakapan ini. Memastikan bahwa "para pebisnis" (CEO, CMO, dan sebagainya) memahami AI penting untuk implementasi AI yang sukses dalam suatu organisasi. Kecuali Anda berada dalam pengembangan teknologi inti, "para pebisnis" berada di pucuk pimpinan, dan Anda ingin mereka tahu bagaimana samudra baru yang mereka arungi bekerja.

Setelah fondasi pengetahuan tim bisnis dibangun, manajer yang skeptis dengan tepat mengingatkan semua orang bahwa AI harus digunakan sebagai alat yang melayani bisnis, dan penting untuk memulai dari awal dengan visi AI yang jelas: jalur yang jelas yang mendefinisikan apa yang akan dilakukannya di era AI, bagaimana ia akan membantu orang-orangnya bekerja lebih efisien, dan bagaimana ia akan membentuk produknya untuk melayani pelanggan dengan lebih baik.

Memiliki visi AI tentu krusial bagi setiap organisasi, dan kami sangat mendorong setiap orang untuk menemukan visi mereka sendiri. Namun, seperti yang ditunjukkan oleh pemimpin AI dalam percakapan kami, visi AI Anda tidak akan datang kepada Anda dalam mimpi, atau dalam presentasi PowerPoint yang disusun oleh perusahaan konsultan.

Sebaliknya, itu adalah hasil dari pengalaman dan eksperimen, keberhasilan dan kemunduran. Bagaimana Anda bisa menghasilkan visi AI yang komprehensif jika Anda belum mengambil langkah apa pun dalam menggunakan AI? Berdasarkan pengalaman kami, kami mengamati bahwa perusahaan yang mengabaikan eksperimen demi fokus pada strategi jangka panjang cenderung berakhir di "lembah kematian curah pendapat": mereka terus membuat hipotesis tentang apa yang harus mereka lakukan, tetapi karena kurang bertindak, mereka kehilangan informasi yang dibutuhkan untuk membuktikan bahwa mereka berada di jalur yang benar. Hasilnya adalah aliran pemikiran yang berkelanjutan tanpa cukup tindakan untuk memvalidasi strategi mereka.

Cara yang baik untuk mulai membangun visi AI jangka panjang Anda adalah dengan mulai berfokus pada proyek AI: inisiatif berskala luas yang menggunakan AI untuk membangun produk atau fitur baru, atau untuk mengoptimalkan proses yang ada. Dengan menjalankan proyek AI, Anda akan mendapatkan pengalaman yang dibutuhkan dan mempelajari apa yang belum Anda ketahui untuk membentuk visi AI yang bermanfaat.

Proyek AI bisa lebih atau kurang rumit (idealnya, Anda memulai dengan sesuatu yang kecil lebih lanjut tentang itu nanti), tetapi harus selalu memiliki dampak bisnis yang solid dan terukur. Anda dapat melihatnya sebagai upaya dua hingga enam bulan: sebuah tim dari organisasi Anda membentuk gugus tugas AI dan membangun satu produk atau layanan berbasis ML, dengan keluaran yang terdefinisi dengan baik, KPI, dan dampak bisnis yang diharapkan. Semua studi kasus industri yang kami miliki dalam buku ini adalah proyek AI. Contoh sempurna tentang bagaimana proyek AI dapat menghasilkan visi AI adalah studi kasus pertama di bab 2, yang menjelaskan bagaimana Google merevolusi cara mengoptimalkan konsumsi energi pusat datanya.

Mari kita telaah lebih dalam studi kasus Google untuk melihat bagaimana mereka membangun pusat data yang dioptimalkan untuk ML. Percikan pertama yang memicu revolusi

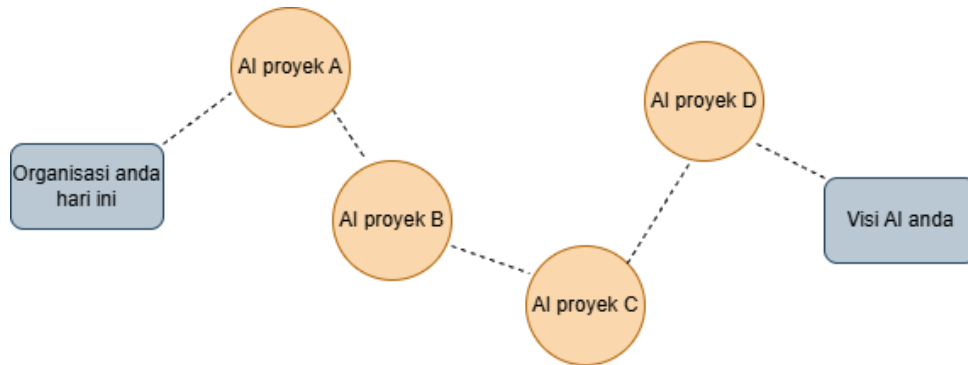
ini adalah proyek sederhana yang digagas dan dipimpin oleh Jim Gao, salah satu insinyur pusat data Google. Gao memanfaatkan kebijakan 20% Google untuk menghabiskan satu hari dalam seminggu kerjanya membangun model ML sederhana guna memprediksi salah satu metrik kinerja utama sistem pendingin pusat data, dengan mempertimbangkan parameter operasional lainnya. Proyek ini tampak menjanjikan, dan langkah selanjutnya adalah menguji model Gao dengan mensimulasikan sistem pendingin dalam berbagai kondisi dan memilih yang terbaik secara manual.

Langkah selanjutnya adalah membiarkan DeepMind, anak perusahaan AI Google, mengambil alih. DeepMind menyempurnakan model Gao dan memungkinkan Google memangkas tagihan energinya hingga 40%. Ini bukanlah langkah terakhir dari inisiatif ini: Google masih mengandalkan para insinyur untuk memeriksa pilihan algoritma dan mengimplementasikannya. Visi yang utuh semakin mendekati kenyataan dengan proyek DeepMind terakhir, yang memungkinkan ML mengambil alih rekomendasi, dengan para insinyur hanya memeriksa bahwa tidak ada yang salah.

Pelajaran utama dari pengalaman pusat data Google adalah bahwa membina proyek-proyek kecil adalah cara terbaik untuk mulai membentuk visi AI yang lebih luas. Dengan memulai dan menyelesaikan proyek-proyek AI, Anda akan mempelajari apa yang perlu Anda ketahui untuk akhirnya mendefinisikan visi AI yang lengkap yang mencakup seluruh organisasi, seperti yang ditunjukkan pada gambar 7.1.

Jika Anda melihat kasus Google, tampaknya tujuan akhir dari sebuah visi AI seharusnya sudah jelas sejak awal, dan setiap proyek AI hanyalah bagian dari rencana yang lebih besar ini. Ini adalah kebohongan yang mudah. Sebenarnya, aliran wawasan mengalir ke arah sebaliknya. Proyek-proyek AI individual dapat (dan seharusnya) memberi tahu Anda dan organisasi Anda di mana terdapat potensi, dan di mana terdapat jalan buntu, menginformasikan investasi dan membantu Anda mengembangkan visi berdasarkan apa yang Anda temukan dari setiap proyek AI. Di bagian selanjutnya, kita akan membahas cara merancang proyek AI baru, mengujinya, memprioritaskannya, dan memilih di mana Anda akan memfokuskan energi. Sebelum melanjutkan, kami ingin memperingatkan Anda tentang elemen lain dari percakapan fiktif kita yang mungkin lebih rumit di dunia nyata.

Pertama-tama, penggagas dalam kasus kita adalah CEO perusahaan. Ini adalah kasus yang paling beruntung; bayangkan jika CEO tersebut adalah seorang skeptis. Dalam kasus ini, Anda harus berjuang untuk mendapatkan dukungan mereka. Elemen keberuntungan lainnya dari karakter fiktif kita adalah bahwa dalam kasus kita, pemimpin AI segera memperkenalkan dua elemen penting: berinvestasi dalam pendidikan dan mengalihkan fokus dari strategi jangka panjang ke proyek AI jangka pendek. Bagaimana jika Anda tidak memiliki pemimpin AI? Dan bagaimana jika orang yang skeptis, penggagas, dan teknisi tidak semudah diyakinkan seperti karakter fiktif kita yang berpikiran terbuka? Seperti yang telah kami katakan di awal bagian ini, inovasi AI harus dilakukan oleh sebuah tim, bukan oleh seorang diri. Artinya, jika Anda adalah pemimpin AI, Anda harus berupaya mengajak orang lain dan memberi pencerahan agar mereka menghargai nilai pendidikan dan eksperimen.



Gambar 7.1 Visi AI Anda akan merupakan hasil dari beberapa proyek AI. Visi tersebut jarang mengikuti alur yang lurus: Anda mungkin harus membangun beberapa proyek AI yang bergerak ke arah yang berbeda-beda.

7.2 PENEMUAN: Mencari Peluang AI

Maju cepat dua minggu dari percakapan sebelumnya. Tim fiktif kita telah menjalani beberapa pelatihan tentang AI dan mereka sekarang siap membentuk gugus tugas AI untuk meluncurkan proyek-proyek pertama. Berikut percakapan mereka setelah pelatihan:

CEO : Saya sangat menyukai pelatihan ini. Saya merasa jauh lebih memahami apa itu AI sebenarnya dan bagaimana AI dapat membantu kita. Bagaimana menurut kalian?

CTO : Sejujurnya: meskipun saya seorang teknisi, saya menyadari bahwa saya memiliki beberapa kesalahpahaman tentang AI.

Manajer

Pemasaran : Sama. Namun, saya ingin melihat hasil nyata sebelum menjadi penggemar.

Pimpinan AI : Keren, teman-teman; mari kita mulai bertukar pikiran tentang beberapa proyek yang bisa kita kerjakan untuk mendapatkan momentum, ya? Apakah kalian punya ide?

CEO : Saya melihat Google memiliki fitur baru di ponsel saya yang mengurutkan gambar berdasarkan isinya. Bagaimana jika kita membantu orang-orang mengurutkan gambar rumah mereka secara otomatis berdasarkan ruangan? Dengan cara ini, akan jauh lebih mudah bagi penjual untuk mengunggah gambar, dan bagi pembeli untuk menemukan apa yang mereka minati (kamar mandi, dapur, dan sebagainya).

CTO : Itu ide yang keren. Saya juga selalu bertanya-tanya apa yang bisa kita lakukan dengan semua data yang kita kumpulkan. Tim saya memiliki ratusan ribu catatan rumah yang terjual, dan saya bertanya-tanya apakah kita bisa menggunakan data ini untuk menjadi penasihat keuangan bagi investor real estat; saya pasti akan menggunakannya.

Mungkin langkah pertama bisa berupa algoritma yang memprediksi harga jual rumah.

Manajer

Pemasaran : Di pihak saya, saya bertanya-tanya bagaimana kita bisa menggunakan AI untuk mengotomatiskan proses kita. Mungkin kita bisa mencoba memangkas biaya dengan mengotomatiskan beberapa tugas yang dilakukan broker kita, atau mungkin bahkan memikirkan broker AI yang sangat canggih?

Pimpinan AI : Ini semua ide yang bagus! Saya sarankan kita mencoba membuat sketsa beberapa proyek dan memprioritaskannya. Bagaimana menurut Anda?

Merencanakan langkah pertama untuk menerapkan AI dalam suatu organisasi bisa jadi menakutkan. Dari mana Anda mulai? Bagaimana jika Anda tidak punya ide untuk proyek? Bagaimana jika Anda memiliki terlalu banyak ide? Apakah ide Anda bagus atau buruk? Meskipun kreativitas dan insting bisnis Anda sangat berharga, memiliki kerangka kerja mental untuk memandu upaya Anda selalu bermanfaat.



Gambar 7.2 Tempat-tempat di mana Anda dapat mencari peluang AI, baik di dalam maupun di luar organisasi Anda

Ada tiga fase untuk beranjak dari nol hingga mendefinisikan proyek AI yang

solid:

1. *Invensi*—Jenis proyek apa yang bisa kita bangun?
2. *Prioritas*—Proyek mana yang harus kita mulai fokuskan?
3. *Validasi*—Apakah proyek-proyek ini layak untuk dikejar?

Dalam percakapan sebelumnya, Anda melihat contoh langkah pertama: invensi. Tokoh fiksi kita kreatif dan langsung menghasilkan beberapa ide menarik dan berbeda, tetapi mengandalkan kreativitas individu tidak selalu merupakan strategi yang baik. Inilah sebabnya kami merancang model mental yang dapat membantu menginspirasi kreativitas dan ketajaman bisnis Anda. Model mental berikut membantu Anda mencari proyek AI yang kemudian dapat Anda prioritaskan dan validasi. Dimulai dengan membagi ruang eksplorasi Anda menjadi dua bagian: di dalam dan di luar organisasi Anda, seperti yang ditunjukkan pada gambar 7.2.

Saat Anda melihat ke dalam organisasi Anda, Anda dapat mengikuti dua pendekatan:

1. Lihatlah hal-hal yang dilakukan bisnis Anda saat ini dan tanyakan, "Bagaimana jika saya bisa melakukannya dengan lebih baik/lebih murah/lebih cepat?"

2. Lihatlah hal-hal yang tidak dilakukan bisnis Anda saat ini dan tanyakan, "Dapatkah AI membantu saya menawarkan produk/layanan ini?"

Jika Anda melihat kembali percakapan fiktif dari tim real estat kita, manajer pemasaran melakukan #1: ia menyarankan untuk menganalisis proses perusahaan saat ini dan memikirkan bagaimana AI dapat membantu organisasi menjadi lebih baik. Pendekatan ini biasanya akan membawa Anda ke proyek-proyek inovasi inkremental yang akan memajukan organisasi Anda, menjadi lebih produktif, efektif, dan efisien.

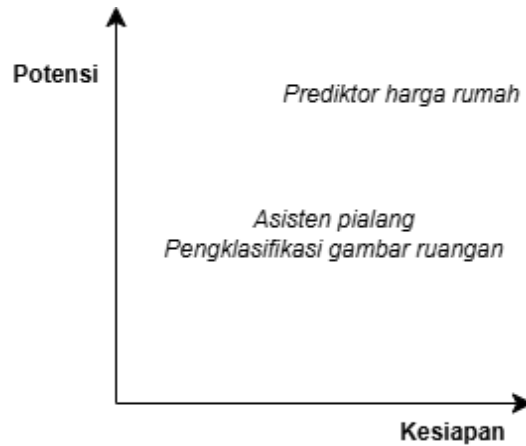
Jenis proyek kedua muncul dari aktivitas, produk, atau layanan yang belum pernah dibangun atau ditawarkan oleh organisasi Anda, tetapi berpotensi dibangun menggunakan AI. Inilah yang dilakukan CTO: ia mengatakan ia ingin memiliki penasihat keuangan pribadi yang didukung AI, dan mulai bertanya-tanya bagaimana hal itu dapat dibangun. Proyek-proyek ini lebih disruptif dan berpotensi mendefinisikan kategori produk yang inovatif. Contohnya adalah algoritma kelayakan pinjaman Square Capital yang dijelaskan di bab 2. Pinjaman usaha kecil tidak pernah menjadi bisnis yang menarik bagi bank, karena potensi keuntungannya kecil, risikonya tinggi, dan proses evaluasinya mahal. Square menyadari bahwa AI adalah kunci untuk mengakses pasar yang hampir tak tersentuh ini.

Dalam hal mencari ide di luar organisasi Anda, berdasarkan pengalaman kami, kami menemukan dua opsi efektif: mengamati apa yang dilakukan industri lain dan apa yang dimungkinkan oleh teknologi baru. CEO fiktif dalam percakapan kami melakukan keduanya secara bersamaan: ia mengamati apa yang dilakukan Google, yang dimungkinkan oleh teknologi visi komputer baru. Sering kali, mengamati industri lain berarti mengamati perusahaan teknologi. Jika Anda berfokus pada perusahaan digital di Silicon Valley, kebanyakan dari mereka 10 tahun lebih maju daripada mayoritas organisasi besar yang mapan di industri "tradisional". Hal ini memungkinkan kami untuk mengintip masa depan, mengamati apa yang mampu dilakukan AI ketika diterapkan pada masalah skala besar, dan mencoba melihat apa yang dapat dipelajari dan diterapkan ke industri yang Anda minati. Contohnya adalah pekerjaan Google di pusat datanya. Google adalah pemimpin dalam penerapan AI untuk produk-produknya, dan pengalaman ini menginspirasi Gao untuk menerapkan pendekatan tersebut ke bidang baru yang belum dieksplorasi, yaitu mengoptimalkan konsumsi energi.

Contoh menarik lainnya dari inovasi yang terinspirasi oleh teknologi baru adalah FaceApp, sebuah aplikasi yang menggunakan AI untuk menghasilkan transformasi wajah yang sangat realistis dalam foto. Aplikasi ini dapat mengubah wajah agar tersenyum, tampak lebih muda, tampak lebih tua, atau mengubah jenis kelamin. Aplikasi ini telah beredar di pasaran cukup lama, tetapi menjadi viral beberapa bulan setelah NVIDIA merilis model AI sumber terbuka dengan kemampuan serupa. Kemungkinan besar, para insinyur di balik FaceApp memanfaatkan teknologi baru yang tersedia untuk membangun versi aplikasi yang lebih canggih, mendapatkan daya tarik yang dibutuhkan untuk menjadi viral.

Jika Anda mengadakan sesi curah pendapat dengan tim Anda, empat pendekatan yang baru saja kita bahas dan yang dapat Anda temukan pada Gambar 7.3 adalah tempat yang baik untuk memulai. Kemungkinan besar, di akhir sesi curah pendapat, Anda akan mendapatkan

beberapa ide yang harus Anda prioritaskan. Di bagian selanjutnya, kami akan memperkenalkan dua konsep dasar yang akan membantu Anda melakukan penilaian umum pertama terhadap potensi ide-ide baru Anda.



Gambar 7.3 Perbandingan ide proyek yang muncul dari sesi curah pendapat fiktif

7.3 PRIORITAS: MENGEVALUASI PROYEK AI

Kita semua pernah mengalaminya: Anda mengadakan rapat curah pendapat dengan beberapa orang, dan setiap orang memiliki ide dan menganggapnya yang terbaik. Hal ini mungkin terjadi pada kelompok fiktif kita. Idealnya, Anda ingin menguji setiap ide, tetapi kenyataannya, sebagian besar waktu, Anda harus memutuskan apa yang akan diprioritaskan. Di bagian ini, kami akan membantu Anda melakukannya.

Hal paling sederhana yang biasanya dipertimbangkan kebanyakan orang dalam fase ini adalah laba atas investasi (ROI) yang diharapkan dari sebuah proyek: dari semua proyek, memprioritaskan proyek dengan potensi tertinggi tampaknya merupakan ide yang bagus. Berdasarkan pengalaman kami, berfokus hanya pada ROI suatu proyek adalah strategi yang picik. Ada variabel lain, yang mungkin bahkan lebih penting, yang perlu dipertimbangkan saat memilih proyek: kesiapan Anda. Tanpa mempertimbangkan seberapa siap Anda untuk menangani proyek AI, Anda berisiko terjebak. Untuk setiap proyek, Anda dapat memperkirakan seberapa siap Anda menghadapinya dengan mempertimbangkan hal-hal berikut:

- ❖ *Kompleksitas*—Seberapa sulit proyek AI kita untuk dibangun? Bisakah kita mendapatkan bukti konsep dalam satu bulan, atau apakah kita membutuhkan dua tahun R&D? (Anda akan lebih memahami hal ini setelah membaca dua bab berikutnya.)
- ❖ *Keterampilan*—Apakah Anda sudah memiliki pengalaman dalam proyek AI dan talenta AI untuk mengimplementasikan ide-ide baru Anda?
- ❖ *Data*—Apakah Anda memiliki semua data yang dibutuhkan?
- ❖ *Pengujian*—Apa lagi yang kita butuhkan untuk menguji proyek ini? (Misalnya, apakah kita memerlukan aplikasi seluler atau situs web?)

Berdasarkan jawaban atas pertanyaan-pertanyaan ini, Anda dapat membandingkan kesiapan Anda untuk menyelesaikan setiap proyek yang Anda buat dan membuat keputusan. Untuk

membuat konsep ini lebih konkret, kami akan menceritakan kisah sebuah organisasi besar yang kami bantu memulai upaya AI-nya.

Kami akan menyebutnya ACME, dan katakanlah mereka memproduksi barang konsumsi yang cepat laku seperti tisu toilet. Ketika kami mulai bertukar pikiran, kepala produksi perusahaan sangat bersemangat untuk membangun algoritma yang dapat meningkatkan perencanaan produksi. Keuntungannya pasti sangat besar: lagipula, ACME memproduksi ratusan truk tisu toilet setiap hari, dan penjadwalan produksi yang lebih efisien akan meningkatkan efisiensi, dengan potensi keuntungan jutaan rupiah. Namun, ada beberapa tantangan:

- Masalahnya sangat kompleks dan memiliki begitu banyak variabel yang berperan sehingga membutuhkan upaya penelitian yang panjang, mahal, dan berisiko.
- ACME belum memiliki talenta teknologi atau pengalaman AI.
- Departemen TI ACME baru saja meningkatkan infrastruktur pengumpulan datanya setahun sebelumnya, sehingga hanya memiliki data untuk satu tahun saja.
- Menguji proyek ini sangat sulit: Anda harus mengubah jadwal produksi, yang lebih mudah diucapkan daripada dilakukan di perusahaan yang benar-benar memproduksi berton-ton barang setiap hari.

Sekalipun ROI-nya bisa sangat besar, kami menyarankan untuk menempatkan proyek ini di urutan terbawah daftar ACME, dan melanjutkannya lagi setelah mereka memiliki lebih banyak data, lebih banyak pengalaman, dan lebih percaya diri dengan AI. Sebaliknya, kami berfokus pada proyek-proyek yang lebih kecil yang mungkin menghasilkan imbal hasil yang lebih kecil, tetapi kami yakin dapat menyelesaikannya dalam hitungan minggu, bukan tahun. Kompromi antara ROI dan kecepatan ini sangat penting bagi perusahaan yang baru memulai perjalanan AI-nya.

Jika kompleksitas, keterampilan, data, dan kemudahan pengujian adalah empat variabel yang perlu dipertimbangkan untuk mengevaluasi kesiapan Anda menghadapi suatu proyek, bagaimana Anda mengevaluasi ROI? Terkadang Anda dapat dipandu oleh ketajaman bisnis Anda, tetapi panduan akan sangat membantu.

Potensi sebuah proyek AI berkorelasi langsung dengan seberapa besar ia memanfaatkan dua kekuatan utama AI: skala dan akurasi. Skala mengacu pada fakta sederhana bahwa terdapat keuntungan yang jelas dalam membiarkan komputer mengambil alih suatu tugas. Keuntungan yang paling umum adalah kecepatan dan efisiensi biaya, tetapi Anda mungkin mendapatkan keuntungan yang berbeda berdasarkan kebutuhan Anda. Misalnya, skala dapat membantu Anda mengurangi kebosanan karyawan dengan mengotomatiskan tugas-tugas yang mengganggu, atau meningkatkan keselamatan mereka untuk tugas-tugas berbahaya di lini produksi yang dapat Anda otomatisasi. Tentu saja, skala merupakan keunggulan setiap aplikasi perangkat lunak, tetapi AI sangat memperluas apa yang dapat disentuh oleh perangkat lunak (misalnya, kita telah melihat bagaimana perangkat lunak tradisional biasanya tidak dapat memecahkan masalah visi komputer).

Proyek yang memanfaatkan skala biasanya merupakan tugas-tugas otomatisasi yang memberikan manfaat utama berupa pemotongan biaya atau pengurangan waktu. Contohnya

adalah proyek penyortiran mentimun yang kami perkenalkan di bab 5 yang menghemat banyak waktu bagi seorang petani Jepang.

Akurasi mengacu pada kemampuan ML untuk belajar dari data dan membuat model dengan kesalahan yang lebih kecil daripada yang dapat dicapai dengan rekayasa perangkat lunak konvensional. Beberapa masalah memiliki hubungan langsung antara akurasi dan hasil, sedemikian rupa sehingga Anda dapat menghubungkan jumlah dolar dengan setiap peningkatan akurasi prediksi. Contoh klasik yang jelas adalah di bidang keuangan: semakin baik Anda memprediksi harga saham Apple di masa mendatang, semakin banyak uang yang akan Anda hasilkan. Semakin erat hubungan antara keuntungan dan akurasi, semakin besar kemampuan AI dalam memodelkan dunia dari data yang akan memberikan hasil positif bagi proyek Anda.

Dalam buku ini, Anda telah melihat beberapa proyek yang berhasil, terutama karena AI memungkinkan pengambilan keputusan dengan akurasi yang lebih tinggi. Salah satunya adalah proyek Target di bab 3; peritel Amerika tersebut menggunakan data dari kartu loyalitas untuk menargetkan keluarga yang akan memiliki bayi. Mungkin, alternatif tradisionalnya adalah menargetkan secara luas semua keluarga dalam rentang usia tertentu. Jelas bahwa akurasi penargetan yang lebih baik dapat menghasilkan imbal hasil yang jauh lebih tinggi daripada status quo.

Proyek AI terbaik memanfaatkan skalabilitas dan akurasi AI. Ini adalah proyek-proyek dengan potensi disrupsi tinggi yang dapat sepenuhnya mengubah sebuah industri, seperti kasus Square Capital di bab 2. Square menggunakan data keuangan usaha kecil untuk memutuskan apakah akan memberi mereka pinjaman. Keberhasilan Square terkait langsung dengan seberapa baik mereka dapat memahami bagaimana keuangan bisnis memengaruhi kemampuan mereka untuk membayar kembali pinjaman: semakin tinggi akurasi mereka, semakin sukses mereka dalam mengelola risiko dan akan memberikan pinjaman kepada bisnis yang layak kredit.

Dalam kasus Square, skala sebenarnya adalah alasan utama keberhasilan perusahaan, karena memungkinkannya untuk melayani pinjaman usaha kecil yang panjang yang tidak mungkin dilayani sebaliknya. Proyek pusat data Google memanfaatkan skalabilitas dan akurasi AI. Sebelum sistem AI diimplementasikan, konsumsi pusat data dioptimalkan oleh para insinyur menggunakan pengetahuan termodinamika mereka. Bahkan para insinyur harus tidur, jadi sistem yang skalabel yang memantau pabrik dan membuat keputusan secara real-time sangat berharga. Terlebih lagi, algoritma Google lebih akurat daripada yang dihasilkan oleh insinyur manusia, sehingga menghasilkan hasil luar biasa yang dicapai perusahaan: penghematan 40% adalah jumlah yang sangat besar!

Jika skalabilitas dan akurasi tinggi merupakan dua pendorong utama proyek Anda, kemungkinan besar Anda sedang melihat potensi ROI yang sangat tinggi. Setelah mengetahui hal ini, Anda dapat mempertimbangkan kesiapan Anda untuk setiap proyek spesifik guna memutuskan di mana Anda akan memfokuskan energi dalam beberapa bulan ke depan dalam eksperimen AI Anda.

Untuk memperjelas seperti apa proses ini, mari kita lihat kembali proyek-proyek yang dihasilkan dari sesi curah pendapat gugus tugas AI real estat fiktif kita:

- ✓ *Pengklasifikasi gambar kamar*—Fitur yang secara otomatis mengurutkan gambar berdasarkan kamar
- ✓ *Prediktor harga rumah*—Fitur yang memprediksi harga jual rumah
- ✓ *Asisten broker*—Sejenis chatbot yang dapat menyederhanakan pekerjaan broker atau menggantikannya sepenuhnya

Mari kita lihat potensinya terlebih dahulu. Pengklasifikasi gambar kamar memiliki faktor skala yang tinggi, karena memungkinkan orang mengunggah dan mengurutkan gambar dalam sepersekian detik, alih-alih mengurutkannya secara manual. Namun, faktor akurasi rendah: kami tidak peduli jika algoritma kami lebih baik daripada manusia dalam mengidentifikasi kamar mandi dari dapur, selama hasilnya cukup baik.

Prediktor harga rumah memiliki faktor skala yang tinggi, karena saat ini menebak harga jual adalah tugas yang biasanya dilakukan oleh broker rumah berpengalaman yang sibuk dan dapat memakan waktu sehari-hari untuk membuat prediksi. Proyek ini dapat membantu klien mendapatkan penawaran harga dalam hitungan detik, bukan hari. Faktor akurasi juga sangat tinggi, karena semakin baik Anda dalam memprediksi nilai rumah, semakin Anda dapat membantu klien dalam perencanaan keuangan mereka.

Asisten broker juga memiliki nilai skala yang tinggi, karena mempercepat waktu respons klien dan menghemat waktu broker dalam menjawab pertanyaan sederhana, sehingga membantu mereka fokus pada aktivitas yang paling dapat mereka lakukan. Namun, faktor akurasi tidak terlalu tinggi, karena broker sudah cukup baik dalam menjawab pertanyaan klien.

Sekarang mari kita evaluasi kesiapan Anda. Anda sudah memiliki gambar ruangan untuk membangun pengklasifikasi gambar ruangan. Selain itu, teknologinya tidak terlalu sulit untuk dibangun, dan bahkan dapat dibeli dari penyedia pihak ketiga. Di sisi lain, mengujinya membutuhkan beberapa koneksi antara model dan platform Anda saat ini.

Untuk prediktor harga rumah, Anda sudah memiliki semua data yang dibutuhkan. Ini juga merupakan proyek data bisnis inti dengan data yang terstruktur dengan baik, yang tidak menghadirkan banyak tantangan teknis dan tidak memerlukan keahlian teknis yang luar biasa (Anda akan lebih mahir dalam mengevaluasi kompleksitas suatu proyek di bab berikutnya). Untuk mengujinya, Anda juga perlu menghubungkan situs web Anda dengan model Anda (atau trik cerdas lainnya tunggu bagian selanjutnya).

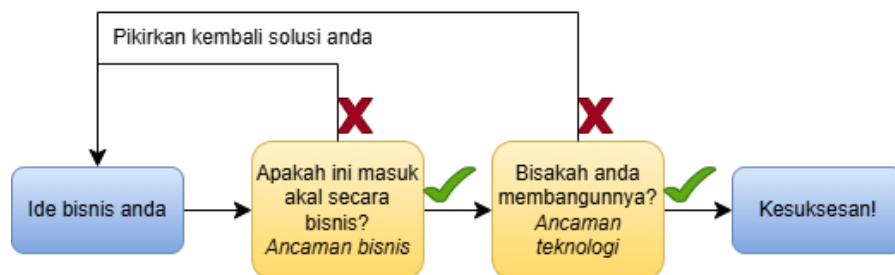
Asisten broker adalah yang paling rumit untuk dibangun, karena seperti yang Anda pelajari dari bab 5, data teks adalah yang paling rumit dari semua jenis data. Anda mungkin memiliki beberapa data untuk itu, dan pengujiannya dapat dilakukan menggunakan platform Facebook Messenger.

Jika pada fase Invensi kita membuat daftar proyek, pada fase Prioritas kita membangun tebakan awal yang matang tentang potensinya berdasarkan skala dan faktor akurasi, lalu mengurutkannya berdasarkan potensi dan kesiapan Anda untuk mengimplementasikannya. Sekarang kita tahu bahwa prediktor harga rumah adalah proyek terbaik untuk difokuskan saat

ini, karena memiliki potensi tinggi dan kita juga lebih siap untuk membangunnya dibandingkan yang lain. Sekarang, mari kita lihat bagaimana kita dapat memvalidasinya dan mulai membangunnya.

7.4 VALIDASI: MENGANALISIS RISIKO

Dua jenis ancaman utama yang dapat menghambat keberhasilan proyek AI Anda: ancaman bisnis dan teknologi (gambar 7.4). Mari kita lihat apa arti masing-masing. Ancaman bisnis berkaitan dengan nilai dan kegunaan produk atau layanan baru Anda yang didukung AI sebagaimana dipersepsikan oleh pelanggan. Sekalipun produk Anda memiliki teknologi yang luar biasa, belum tentu orang-orang membutuhkannya atau ingin menggunakannya. Dalam konteks prediktor harga rumah kita, terdapat banyak potensi ancaman terhadap kesuksesan bisnisnya. Misalnya, ada kemungkinan pelanggan tidak menginginkannya atau tidak mempercayainya. Dalam kedua kasus tersebut, mereka tidak akan menggunakannya, betapa pun hebatnya teknologi Anda. Perlu diingat bahwa kita menggunakan definisi pelanggan yang cukup luas. Asumsikan Anda sedang membangun alat internal untuk organisasi Anda; misalnya, sebuah algoritma yang membantu departemen penilaian kualitas sebuah pabrik.



Gambar 7.4 Dalam perjalanan menuju kesuksesan, proyek AI Anda harus menghadapi ancaman bisnis dan teknologi.

Dalam hal ini, karyawan departemen tersebut dapat dianggap sebagai pelanggan, dan sebagai pelanggan "eksternal", mereka mungkin menganggap produk baru Anda bermanfaat atau mungkin tidak. Ancaman teknologi berkaitan dengan kinerja model AI Anda dan apa yang dapat diberikannya kepada pelanggan Anda. Anda sekarang sudah familier dengan konsep akurasi model ML, dan fakta bahwa tidak ada penerapan ML yang sempurna. Besarnya dan sifat kesalahan yang dibuat oleh algoritma ML dapat diterima atau tidak, tergantung pada aplikasi spesifiknya. Misalnya, aplikasi yang kritis terhadap keselamatan seperti otomasi industri memiliki persyaratan akurasi yang lebih ketat daripada kebanyakan aplikasi seluler.

Pendekatan yang baik untuk mengurangi risiko implementasi adalah dengan memandang potensi ancaman seolah-olah merupakan serangkaian asumsi yang harus dipenuhi agar proyek berhasil. Saat Anda mengembangkan proyek AI Anda, teruslah membuat daftar asumsi bisnis dan teknis yang harus Anda buat. Untuk prediktor harga rumah kami, asumsi-asumsi ini akan terlihat seperti ini:

Asumsi bisnis:

- Pelanggan menginginkan estimasi cepat nilai rumah mereka.

- Pelanggan mempercayai estimasi harga otomatis dan bersedia mempertaruhkan keuangan mereka pada hasil mereka, terlepas dari seberapa akuratnya secara teknis.

Asumsi teknologi:

- Kami memiliki data yang tepat untuk membangun model, dan kami dapat memperbaruinya seiring waktu.
- Akurasi model cukup tinggi sehingga estimasinya dapat bermanfaat.

Karena seringkali ada cara untuk mengatasi hambatan teknologi, kami sarankan Anda menguji asumsi bisnis Anda terlebih dahulu. Tidak ada solusi untuk produk yang tidak diinginkan siapa pun. Ada cara untuk menguji hipotesis bisnis sebelum menulis baris kode apa pun atau mengumpulkan byte data apa pun. Misalnya, Anda dapat menambahkan tombol berlabel Dapatkan Penawaran Harga untuk Rumah Anda dalam 5 Menit dan mengarahkan permintaan ke broker manusia. Kita dapat menyebut strategi validasi ini sebagai AI concierge: Anda menggunakan orang untuk menguji daya tarik fitur baru yang ingin Anda terapkan dengan AI. Ini dimaksudkan sebagai solusi sementara yang tidak akan diskalakan ke jutaan pengguna, tetapi Anda dapat menggunakannya untuk mengukur minat pelanggan pada sampel kecil calon pengguna.

Uji coba yang lebih sederhana lagi yang bisa Anda jalankan adalah dengan menyisipkan tombol palsu berlabel Dapatkan Penawaran Harga Instan untuk Rumah Anda yang mengarahkan pengguna ke halaman web sederhana yang bertuliskan, "Maaf, fungsi ini akan segera hadir" atau "Sedang dalam perbaikan." Anda bisa mencoba trik ini dengan beberapa pengguna, dan hitung berapa banyak dari mereka yang mengklik tombol umpan Anda. Memalsukan fungsi memang tidak ideal, tetapi memungkinkan Anda memvalidasi potensi proyek AI Anda dalam hitungan hari (dan pada dasarnya tanpa anggaran), alih-alih berbulan-bulan.

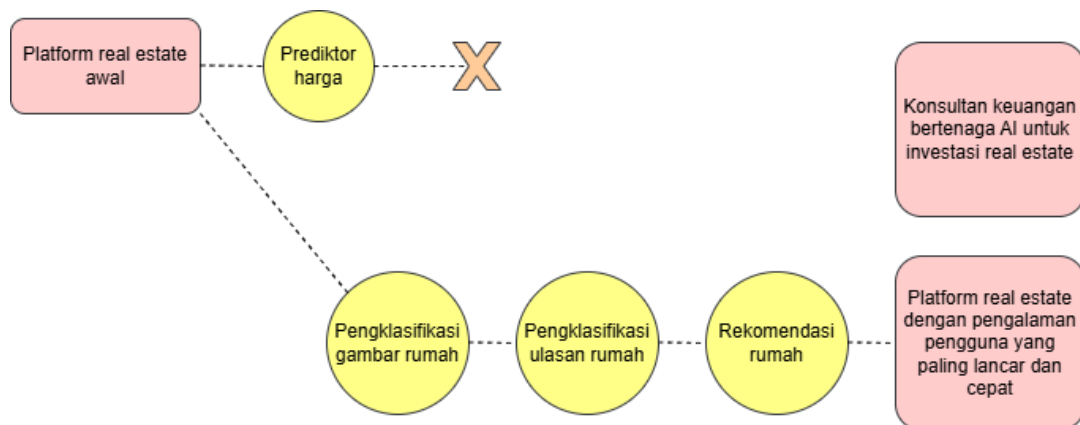
Hasil uji coba semacam itu bisa positif atau negatif. Jika banyak pengguna mengklik tombol Dapatkan Penawaran Harga dalam 5 Menit, saatnya untuk beralih ke konten bab berikutnya dan mulai membangun fitur tersebut! Dalam hal ini, visi AI yang dipaparkan manajer pemasaran kami tampaknya membuahkan hasil: ada ruang untuk mengubah platform real estat Anda menjadi konsultan keuangan lengkap untuk investasi real estat.

Ada kemungkinan juga hasilnya negatif. Jika tidak cukup banyak orang yang tertarik dengan fungsi baru ini, mungkin tidak ada yang merasa membutuhkan estimasi nilai rumah secepat kilat. Dalam hal ini, jangan putus asa: eksperimen pra-produk Anda memungkinkan Anda menghemat banyak uang dan waktu untuk membangun sesuatu yang tidak diinginkan siapa pun. Ini berarti visi AI manajer pemasaran Anda perlu dialihkan ke hal lain, dan Anda dapat mulai mengerjakan proyek AI berikutnya; dalam kasus kami, pengklasifikasi gambar ruangan.

Mari kita asumsikan Anda menguji pengklasifikasi gambar kamar, dan hasilnya sangat sukses. Peluncurannya telah mengajarkan Anda bahwa pelanggan memang sangat menghargai pengalaman listing yang menyenangkan. Berbekal pengetahuan tambahan ini, Anda dapat memutuskan untuk fokus menciptakan platform real estat dengan pengalaman pengguna yang paling lancar. Pada saat Anda mengimplementasikan dan merilis fitur tersebut, Anda akan

bebas untuk terjun ke proyek-proyek mendatang dengan keyakinan bahwa semua proyek AI tersebut memiliki tempat dalam visi AI secara keseluruhan.

Gambar 7.5 menunjukkan bagaimana alur eksperimen dan pembelajaran ini dapat membantu Anda memfokuskan visi AI Anda. Tentu saja, ini hanyalah contoh yang disederhanakan. Dunia nyata adalah tempat yang lebih berantakan, dengan banyak upaya yang gagal dan rasa pusing di setiap kesempatan. Namun, poin utamanya tetap sama: sebagian besar perusahaan asli AI yang sukses tidak memulai dengan strategi yang lengkap, melainkan menghubungkan titik-titik hingga misi AI mereka muncul.



Gambar 7.5 Evolusi visi AI dari contoh platform real estat kami, dan berbagai proyek AI yang kami bahas untuk mencapainya.

Jika Anda ingin menggunakan AI sebagai teknologi transformatif, Anda juga perlu mengubah pola pikir Anda. Kami mendorong setiap organisasi, baik kecil maupun besar, untuk mengadopsi pola pikir eksperimen berbasis data dan berinvestasi dalam AI secara bertahap, belajar dari setiap proyek, dan menyusun visi mereka di sepanjang prosesnya. Cara terbaik untuk melakukannya adalah dengan berfokus pada proyek AI di awal, dan mencoba belajar dari setiap proyek. Jika Anda cermat dalam menyimak, proyek-proyek itu sendiri akan memberi tahu Anda ke mana visi AI Anda harus mengarah, dan mereka akan menjadi fondasi visi Anda.

Menggali lebih dalam metodologi eksperimen produk berada di luar cakupan buku ini, tetapi Anda dapat menemukan banyak sumber daya daring atau di buku lain mengenai topik ini. Tempat yang baik untuk memulai adalah *The Lean Startup* karya Eric Ries atau karya Steve Blank. Yang ingin kami sampaikan kepada Anda adalah gagasan bahwa tidak ada makan siang gratis dalam hal mengembangkan produk baru, bahkan jika Anda menggunakan teknologi canggih seperti AI. Jika Anda belum menguji asumsi bisnis Anda dan ide Anda cacat, bahkan AI yang paling canggih pun tidak akan menyelamatkan Anda.

Setelah asumsi bisnis Anda teruji, saatnya untuk mengubah ide Anda menjadi proyek ML. Inilah yang akan kita lakukan di bagian dan bab selanjutnya.

7.5 MENDEKONSTRUKSI PRODUK AI

Misalkan perusahaan real estat kita telah memvalidasi ide untuk membangun bot guna menyederhanakan interaksi antara pengguna dan platform. Idennya adalah membuat platform

terasa seperti broker sungguhan, sehingga Anda dapat mengajukan pertanyaan seperti "Apa saja rumah dengan tiga kamar tidur di Distrik Keuangan Manhattan?" dan dapat menerima daftar hasil kurasi, seperti yang dilakukan broker manusia. Ini bisa dibilang Siri milik Apple untuk real estat. Karena Anda telah memvalidasi ide tersebut, saatnya untuk berbicara dengan beberapa pakar AI untuk membangun produk tersebut. Beginilah kira-kira percakapan hipotetisnya dalam kasus ini:

Manajer : Hai, John, kami ingin membuat Siri untuk properti. Apakah itu memungkinkan?

Ilmuwan data

(John) : Kedengarannya menarik, tapi saya butuh informasi lebih lanjut. Apa saja data masukannya?

Manajer : Pengguna kami akan langsung berbicara ke aplikasi atau situs web kami dan mendapatkan hasil yang disajikan dengan baik.

John : Hmm, jadi ini berarti data masukannya berupa suara, dan keluarannya bisa... apa saja?

Manajer : Ya, masukannya berupa suara, dan keluarannya adalah jawaban atas pertanyaan pengguna. Kami tahu NLP itu sulit, jadi kami pikir kami bisa mengurangi kerumitannya dengan mengurangi kedalaman aplikasi ini: fitur ini hanya akan membantu pengguna mencari informasi rumah dan tidak lebih. Untuk saat ini.

John : Oke, ini membantu, tapi saya rasa kita sedang membicarakan dua proyek AI yang berbeda. Manajer: Apa maksudmu?

John : Yah, kami tidak bisa langsung memproses audio dan menghasilkan jawaban; ada langkah perantara. Pertama-tama, kita perlu mentranskripsi audio menjadi teks; ini adalah aplikasi umum yang disebut ucapan ke teks. Kemudian, setelah pertanyaan pengguna tersampaikan dalam teks biasa, kita harus memahami maksudnya. Anda dapat menyebut aplikasi ini klasifikasi teks atau deteksi maksud. Hal ini dapat dilakukan, tetapi harus dipecah menjadi dua bagian.

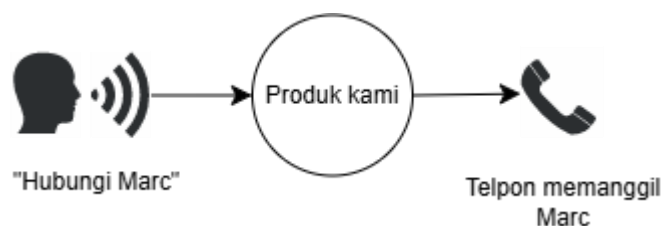
Mempersiapkan proyek AI Anda untuk diproses oleh tim teknis mungkin tidak semudah yang Anda bayangkan. Seperti dalam beberapa contoh kami, banyak masalah yang dapat ditangani AI relatif mudah dan dapat dideskripsikan sebagai satu ide. Misalnya, model prediksi harga rumah merupakan satu produk. Dalam kasus lain, tujuan akhirnya lebih kompleks dan perlu dipecah lebih lanjut sebelum kita dapat menerjemahkannya ke dalam istilah yang ramah ML. Dalam kasus ini, manajer kami menemukan bahwa proyek AI-nya sebenarnya merupakan gabungan dari dua algoritma ML yang terpisah dengan baik. Sekarang, kami akan mengajarkan Anda cara memahami apakah proyek AI Anda merupakan kombinasi dari beberapa algoritma ML. Mari kita ambil contoh sesuatu yang mungkin Anda gunakan sehari-hari: asisten suara seperti Siri dari Apple.

Katakanlah kita ingin asisten pribadi kita hanya melakukan dua hal: menelepon orang dan mengirim pesan. Berkat apa yang telah Anda pelajari di buku ini, Anda tahu bahwa hal pertama yang harus dilakukan untuk mengatasi masalah pembelajaran mesin apa pun adalah memikirkan data yang terlibat. Dalam kasus kita, data masukan adalah suara orang yang menggunakan asisten, yang direkam melalui mikrofon. Keluarannya adalah tindakan yang diambil oleh telepon: mengirim pesan atau menelepon orang tertentu.

Perangkat lunak tradisional akan mengurus pelaksanaan tindakan tersebut: satu-satunya hal yang perlu kita capai adalah menggunakan ucapan untuk memutuskan tindakan mana yang akan diambil, seperti yang ditunjukkan pada gambar 7.6.

Pada dasarnya, yang kami jelaskan adalah masalah klasifikasi terbimbing: masukannya adalah data audio, dan keluarannya adalah sebuah kelas. Apa saja kelas-kelasnya? Secara teoritis, Anda seharusnya memiliki kelas-kelas seperti "kirim pesan ke Marc," "telepon Marc," "kirim pesan ke Marc," "jawab Marc," dan "telepon balik Marc" untuk setiap orang di daftar kontak Anda. Kedengarannya agak gila, bukan? Masalahnya bukan hanya kita memiliki banyak kelas, tetapi juga memetakan berkas audio ke kelas-kelas secara langsung sangatlah sulit. Bahkan, jika Anda mencoba mengingat semua aplikasi yang telah kita bahas dalam buku ini, tidak ada yang seperti ini.

Mari kita lihat apa yang terjadi jika kita memecah masalah menjadi langkah-langkah yang lebih kecil, menggunakan teknologi yang dibahas dalam buku ini dan mulai dari data masukan. Kita membahas data audio di bab 3, di mana kita mengatakan bahwa pembelajaran mendalam dapat digunakan untuk membangun sistem pengenalan suara, mengubah data audio menjadi teks. Teknologi ini cukup solid dan matang, terutama untuk bahasa-bahasa paling populer seperti Inggris dan Spanyol, serta untuk penutur asli (kami, para penulis, terkadang masih kesulitan dipahami oleh Siri dengan intonasi bahasa Italia kami). Ini berarti kami cukup yakin bahwa jika pengguna kami mengatakan "SMS Marc", kami dapat membangun sistem yang mengambil suara tersebut dan mentranskripsikannya menjadi kata-kata tertulis.



Gambar 7.6 Skema dasar produk asisten suara

Setelah langkah ini, data masukan yang kita tangani berubah secara mendasar: kita sekarang bekerja dengan data teks, dan bukan lagi dengan audio. Kita masih perlu memahami apakah pengguna ingin menelepon atau mengirim pesan teks, dan dengan siapa mereka ingin berkomunikasi. Mari kita tetap menggunakan pendekatan kita untuk memecah masalah menjadi komponen-komponen intinya, dan fokus pada bagian pertama. Kita ingin membangun

sesuatu yang memetakan kalimat tertulis ke salah satu dari dua tindakan: menelepon atau mengirim pesan teks.

Ketika kita membahas data teks di bab 5, kita telah menjelaskan secara ekstensif bagaimana Anda dapat menggunakan ML untuk mengklasifikasikan teks ke dalam kelas-kelas. Dalam hal ini, kita dapat melatih model ML dengan puluhan kalimat dan label seperti ini:

- "Telepon Marc" -> telepon
- "Bisakah Anda mengirim pesan ke Marc?" -> teks
- "Jawab Marc" -> teks
- "Telepon balik Marc" -> telepon
- "Kirim SMS ke Marc" -> teks

Seperti yang mungkin Anda ingat dari bab 5, meskipun ada puluhan atau ratusan cara untuk mengungkapkan maksud yang sama, penyematan kata sangat menyederhanakan pekerjaan kita dan memungkinkan kita menyelesaikan tugas ini. Beginilah keadaan kita sekarang: kita tahu tindakan apa yang ingin dilakukan pengguna, dan kita tinggal memahami orang mana yang harus dipilih dari daftar kontak. Ini cukup mudah, dan bukan tugas ML. Karena kita memiliki transkripsi dari apa yang dikatakan pengguna, kita cukup memeriksa semua kata yang mereka ucapkan dan melihat kontak mereka. Ketika kita menemukan nama yang cocok dengan apa yang dikatakan pengguna, kita tahu siapa yang harus dikirim SMS atau dihubungi.

Seperti yang Anda lihat, masalah yang tampaknya mustahil dapat menjadi mudah setelah dipecah menjadi tugas-tugas kecil, yang masing-masing diselesaikan dengan kombinasi ML dan rekayasa perangkat lunak tradisional. Gambar 7.7 adalah diagram arsitektur dari produk yang lengkap.

Anda dapat mengikuti proses yang sama untuk memecah proyek yang lebih kompleks. Untuk contoh berikutnya, kita akan mengambil kasus yang sangat rumit: membangun mobil tanpa pengemudi. Bayangkan langkah-langkah yang akan Anda ambil sebagai manusia untuk mengendarai mobil dari halaman belakang rumah Anda di San Francisco ke Los Angeles, dan pahami apa yang harus dihadapi oleh pengemudi komputer.

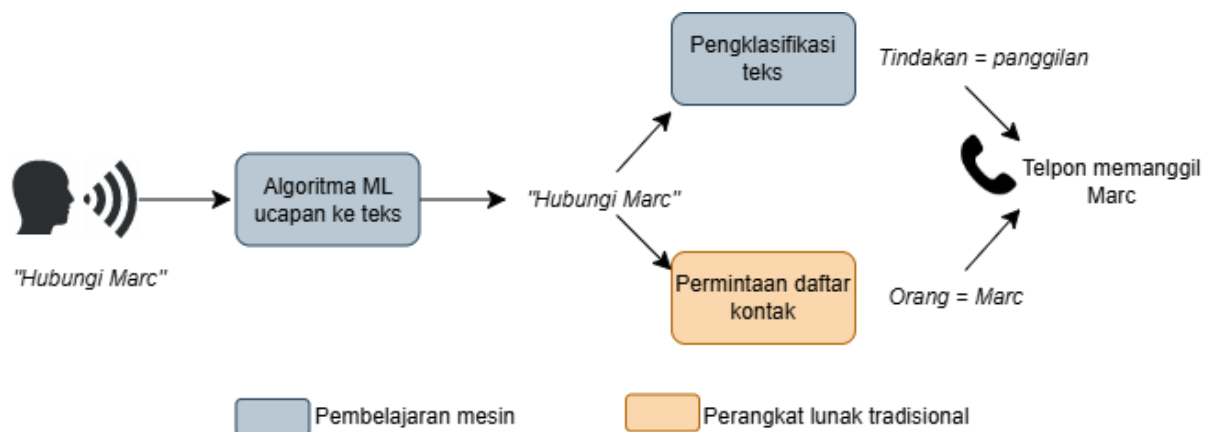
Saat Anda duduk di kursi pengemudi, Anda akan mulai menguraikan langkah-langkah untuk mencapai tujuan akhir Anda: keluar dari garasi, masuk ke jalan bebas hambatan, pilih stasiun radio, usahakan untuk tidak bosan selama beberapa jam, dan, akhirnya, menepi di halaman rumah teman Anda di Los Angeles. Sebelum Anda dapat mencapai tujuan, semua tujuan tingkat tinggi ini harus dipecah lebih lanjut: jalan mana yang harus dituju, jalur mana yang harus dipilih, dan sebagainya. Selain itu, Anda harus mematuhi peraturan lalu lintas dan memperhatikan apa yang dilakukan pengemudi lain agar tidak menyebabkan kecelakaan.

Intuisi ini akan membantu dalam mulai menguraikan sistem AI untuk mencapai mengemudi otonom. Perutean jalan tidak benar-benar membutuhkan komponen AI, karena kita cukup menggunakan kembali algoritma navigasi yang sama yang digunakan di Google Maps. Di sisi lain, akan ada banyak pembelajaran mesin dalam komponen persepsi, yang bertugas mencari tahu di mana pejalan kaki, pesepeda, dan mobil lain berada dalam kaitannya

dengan kendaraan itu sendiri. Karena beberapa sensor yang paling umum digunakan dalam mobil self-driving adalah kamera, pilihan yang wajar adalah menggunakan pembelajaran mendalam untuk klasifikasi gambar dan model lokalisasi objek. Untuk komponen ini, data masukan adalah gambar yang diambil dari kamera depan dan samping yang terpasang pada mobil, dan keluarannya adalah lokasi objek di dekatnya.

Memahami bahwa mobil dan pejalan kaki berada di jalan tidaklah cukup; Anda juga perlu memprediksi apa yang akan dilakukan oleh pengguna jalan lain, sehingga mobil self-driving dapat membuat keputusan terbaik dan teraman. Misalnya, selain mengetahui bahwa pejalan kaki berada di trotoar, Anda juga ingin AI membantu Anda memprediksi apakah mereka akan menyeberang jalan dan memperlambat laju kendaraan demi keselamatan. Tugas ML ini sedikit lebih eksotis daripada yang sebelumnya (mengklasifikasikan objek di sekitar kendaraan), jadi mari kita lihat bagaimana kita akan membingkainya dalam istilah ML.

Kita bahkan tidak yakin tentang apa yang kita inginkan dari model ini: apakah cukup dengan klasifikasi sederhana seperti "kemungkinan bergerak" dan "kemungkinan tidak bergerak", atau apakah model juga harus memprediksi arah pergerakan? Kedua pilihan tersebut valid (dan proyek nyata kemungkinan akan menjalankan keduanya), tetapi yang terakhir lebih berguna dengan mengorbankan kompleksitas dan data pelatihan yang lebih banyak. Dari sudut pandang data masukan, bit yang paling relevan kemungkinan besar adalah lokasi (kota atau jalan raya), lingkungan (penyeberangan jalan, trotoar, pintu masuk kantor), dan jenis entitas (pejalan kaki, pesepeda, atau kendaraan).



Gambar 7.7 Asisten suara dapat dipecah menjadi beberapa komponen, yang mengubah suara pengguna menjadi perintah perangkat lunak.

Semua informasi ini dimasukkan ke dalam komponen pengambilan keputusan yang menentukan cara bergerak aman di tengah kemacetan menuju tujuan. Misalnya, meskipun mobil perlu berbelok ke kiri untuk masuk ke jalan raya, mobil harus mengalah jika jalur tersebut sedang ramai dan manuver mendadak dapat menyebabkan kecelakaan. Mungkin paling masuk akal untuk mengimplementasikan modul pengambilan keputusan sebagai gabungan antara rekayasa perangkat lunak konvensional dan pembelajaran mesin. Hal ini karena beberapa aturan bersifat keras dan cepat (lampu lalu lintas merah berarti "berhenti"),

sementara banyak perilaku lain lebih bernuansa (rem mendadak berbahaya jika mobil membuntuti dari belakang).

Terakhir, keputusan tentang perilaku di jalan, seperti belok kanan atau berhenti di lampu lalu lintas, harus diubah menjadi tindakan untuk roda kemudi dan gas yang benar-benar membuat mobil bergerak (dan berhenti!). Tugas ini tidak terlalu cocok untuk pembelajaran mesin, karena kita memiliki aturan matematika yang tepat yang menyatakan seberapa keras tindakan pengereman seharusnya berdasarkan jarak berhenti, berat penumpang, dan sebagainya. Ini berarti bahwa komponen ini merupakan kandidat utama untuk rekayasa perangkat lunak konvensional.

Gambar 7.8 adalah ikhtisar semua komponen yang telah kita bahas, dan bagaimana komponen-komponen tersebut saling bergantung untuk menghasilkan perilaku yang kompleks. Deskripsi ini jelas tidak lengkap, tetapi menguraikan proyek kompleks seperti ini membantu Anda memahami bagaimana rekayasa konvensional dan pembelajaran mesin dapat saling melengkapi dengan optimal. Bahkan masalah yang paling rumit sekalipun dapat diatasi jika Anda meluangkan waktu untuk meninjau data yang berpotensi Anda gunakan sebagai masukan dan membingkai keluarannya sebagai nilai yang terukur.

Sekarang, mobil self-driving adalah semacam proyek ambisius yang sangat intensif teknologi dan memiliki beberapa persyaratan yang tidak dapat kita bantah: mobil otonom harus dapat dikendarai setidaknya seaman pengemudi manusia, titik. Ketika Anda menerjemahkan peluang bisnis ke dalam satu atau lebih proyek ML, Anda biasanya memiliki lebih banyak kebebasan dalam menentukan batasan proyek ML Anda. Misalnya, dalam kasus "Siri untuk real estat" awal kami, Anda dapat dengan mudah memutuskan untuk mengubah pengalaman pengguna dari ucapan menjadi teks; alih-alih membiarkan orang berbicara dengan platform Anda, mereka dapat menulis kueri dalam bahasa alami. Hal ini dapat memberikan sebagian besar nilai yang Anda harapkan kepada pengguna, sekaligus membatalkan separuh proyek AI Anda (Anda tidak perlu lagi membangun teknologi ucapan-ke-teks!).



Gambar 7.8 Mobil tanpa pengemudi merupakan produk kompleks dengan banyak subkomponen, sebagian menggunakan AI dan sebagian lainnya paling baik diimplementasikan menggunakan rekayasa perangkat lunak konvensional.

Dimulai dengan ide untuk sebuah proyek AI, bagian ini mengajarkan Anda cara memecahnya menjadi subkomponen yang lebih kecil yang dapat ditangani secara individual. Bagian selanjutnya melengkapi pekerjaan persiapan ini dengan menunjukkan cara mengembangkan deskripsi yang mudah dipahami oleh para insinyur untuk masing-masing komponen ini.

7.6 MENERJEMAHKAN PROYEK AI KE DALAM ISTILAH YANG RAMAH ML

Di bagian sebelumnya, ilmuwan data kami, John, membantu kami menemukan bahwa proyek AI "Siri untuk real estat" ternyata lebih kompleks daripada yang kami duga. Kami kemudian menyadari pentingnya memecah proyek AI menjadi komponen-komponen teknologi minimal agar dapat menjadi satu upaya teknologi tunggal. Mari kita bayangkan percakapan lain dengan John, kali ini membahas prediktor harga rumah yang lebih sederhana:

Pimpinan AI : Hai John, kami baru-baru ini menguji fitur baru di situs web kami: prediktor harga rumah. Pada dasarnya, pengguna memasukkan informasi tentang properti mereka, dan kami langsung memberikan perkiraan harganya. Kami menjalankan uji coba terbatas dengan broker manusia yang kembali setelah 10 menit, dan banyak orang menyukainya. Sekarang, kami ingin membangun AI yang dapat melakukannya dalam hitungan detik.

John : Kedengarannya menarik! Jadi, variabel yang perlu kita prediksi adalah harga jual rumah, bukan?

Pimpinan AI : Tepat sekali.

John : Berdasarkan apa? Apakah Anda punya gambaran kasar tentang fitur apa yang sebaiknya kita gunakan untuk model ML kita? Pemimpin AI: Ya, kami telah mewawancarai broker ahli kami dan mencatat semua yang mereka pertimbangkan saat melakukan tugas ini. Beberapa fitur sederhana, seperti luas meter persegi dan jumlah kamar tidur. Beberapa lebih kompleks, seperti kedekatan transportasi umum, reputasi lingkungan, dan sebagainya. Saya dapat membagikan laporan tentang semua yang mereka katakan.

John : Sepertinya kita sudah memulai dengan baik. Hal terbaik untuk membangun model ini adalah menggunakan data transaksi sebelumnya. Dengan cara ini, model pembelajaran mesin (ML) akan mempelajari harga sebenarnya rumah yang dijual. Apakah kita memiliki dataset seperti itu?

Pemimpin AI : Ya, saya sudah memeriksanya dengan tim TI. Kami telah mengumpulkan informasi tentang transaksi selama enam tahun terakhir. Kami telah mencatat ratusan ribu transaksi.

John : Saya pikir kita bisa membangun sesuatu yang hebat dari ini. Apakah kita sudah tahu kinerja seperti apa yang harus kita capai agar proyek ini berhasil?

Pemimpin AI : Kami memperkirakan pengguna akan menerima prediksi yang $\pm 3\%$ lebih rendah dari harga sebenarnya.

Perhatikan betapa apiknya percakapan ini. Yang membuatnya begitu produktif adalah karena pemimpin AI kita telah melakukan pekerjaan rumah mereka: mereka sudah tahu apa yang dibutuhkan tim ilmu data untuk membangun proyek AI ini, dan telah melakukan beberapa analisis awal. Inilah bahan-bahan terbaik untuk kerja sama yang lancar antara pebisnis dan teknisi.

Ketika Anda bukan orang yang mengode algoritma AI, mudah untuk melupakan cara kerjanya dan berakhir dalam percakapan yang sia-sia dengan para insinyur, membuang-buang waktu, dan merasa frustrasi. Sebagai gambaran, berikut adalah titik awal yang buruk untuk proyek AI yang akan membuat tim teknis mana pun merasa ngeri:

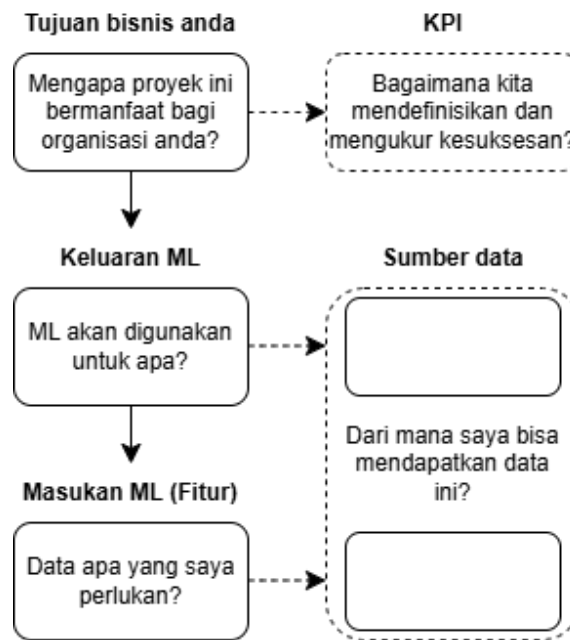
"Kami akan menggunakan data internet untuk memprediksi apa yang akan dibeli orang."

"Kami akan menggunakan data pelanggan kami untuk menemukan yang terbaik."

Sebenarnya, mari kita ulangi. Ide-ide ini tidak salah; hanya saja terlalu samar. Ide-ide ini mungkin cocok untuk materi pemasaran tetapi akan membuat setiap ilmuwan data mengeluh, karena tidak memberikan informasi bermanfaat apa pun tentang teknologinya. Secara umum, dapat dikatakan bahwa berikut adalah karakteristik yang harus dimiliki setiap aplikasi AI:

- Keluaran yang objektif dan terdefinisi dengan baik
- Satu atau lebih set data yang menghubungkan masukan dan keluaran
- Metrik keberhasilan yang terukur dan terdefinisi dengan baik

Teknisi yang familiar dengan ML terbiasa berpikir dalam istilah-istilah ini, dan biasanya dapat dengan mudah meringkaskan masalah ML dengan tepat dengan mempertimbangkan poin-poin penting ini. Namun, berdasarkan pengalaman kami, kami menyadari bahwa orang nonteknis terkadang dapat tersesat dalam prosesnya, dan membutuhkan sedikit bantuan dalam menyusun pemikiran mereka. Untuk alasan ini, kami merancang Kanvas Pembimbingan yang digambarkan pada gambar 7.9: sebuah alat lima langkah yang Anda isi dengan informasi tentang proyek yang sedang Anda kerjakan. Alat ini akan memandu Anda melalui semua elemen yang diperlukan untuk memulai proyek AI yang memiliki cakupan yang tepat.



Gambar 7.9 Kanvas Framing memandu Anda melalui langkah-langkah yang diperlukan untuk mengubah proyek bisnis menjadi proyek ML.

Untuk memandu Anda memahami konsep Kanvas Framing, kami akan kembali mengambil dua studi kasus dan sebuah contoh dari bagian 1, dan meninjaunya melalui sudut pandang Kanvas. Bagian pertama adalah tujuan bisnis Anda. Tujuan bisnis dari ketiga proyek yang akan kami gunakan di seluruh bagian ini adalah sebagai berikut:

- *Mengurangi churn pelanggan* — Kami membahas contoh di bab 3, di mana kami mempertimbangkan model yang mengidentifikasi pelanggan mana yang akan membatalkan layanan berlangganan (misalnya, paket telepon).
- *Meningkatkan efisiensi sistem pendingin pusat data kami* — Dalam studi kasus di akhir bab 2, Google menggunakan pembelajaran terawasi untuk memangkas tagihan energinya sebesar 40%.
- *Manfaatkan pasar pinjaman usaha kecil dengan menawarkan pinjaman langsung kepada usaha kecil* — Ini seperti kasus Square Capital yang dijelaskan di bab 2.

Untuk setiap tujuan bisnis, penting untuk mengingat indikator kinerja utamanya, sebuah metrik kuantitatif yang digunakan untuk mengevaluasi keberhasilannya. Langkah pertama adalah menemukan parameter yang akan digunakan untuk melacak ROI proyek. Untuk ketiga proyek ini, tiga KPI dapat berupa:

- ❖ Tingkat churn pelanggan (bulanan atau tahunan)
- ❖ PUE (efektivitas penggunaan daya, parameter yang mengevaluasi efisiensi energi pusat data)
- ❖ Persentase usaha kecil yang memenuhi syarat yang teridentifikasi

Sekarang, kita perlu menetapkan ambang batas penerimaan untuk masing-masing variabel ini. Dengan kata lain, kita perlu mengidentifikasi nilai untuk setiap KPI yang mewakili kinerja minimum yang diterima, di bawah nilai tersebut produk belum siap untuk dikirim.

Titik awal bisa berupa solusi status quo: lihat solusi Anda saat ini untuk masalah tersebut, ukur kinerjanya berdasarkan KPI yang Anda identifikasi, dan anggap itu sebagai ambang batas. Jika Anda memperkenalkan produk atau layanan baru, Anda mungkin menjalankan kalkulasi manfaat finansial yang memberi tahu Anda jenis kinerja yang Anda butuhkan sebelum model mulai masuk akal secara finansial.

Poin penting lainnya adalah Anda harus memilih cara mengukur KPI Anda. Kita telah membahas berbagai metrik utama dan bagaimana positif palsu atau negatif palsu dapat memiliki implikasi yang sangat berbeda bagi pelanggan Anda. Mari kita asumsikan Anda sedang mengembangkan sistem AI yang memproses pemindaian otak dan mendiagnosis keberadaan tumor. Jelas, positif palsu dan negatif palsu tidak memiliki bobot yang sama dalam situasi ini: positif palsu akan membuat pasien menjalani tes tambahan yang tidak diperlukan, sementara negatif palsu berarti pasien yang sakit secara keliru dipulangkan untuk menghadapi konsekuensi yang jauh lebih serius.

Positif palsu dan negatif palsu juga memiliki bobot yang berbeda dalam situasi yang tidak terlalu tragis. Misalnya, jika kita kembali ke tiga contoh kita, baik prediktor churn maupun estimator kelayakan pinjaman kita mengharuskan kita untuk berpikir mendalam tentang metrik mana yang ingin kita optimalkan. Dalam kasus prediktor churn, kita telah membahas perbedaan antara negatif palsu (pengguna yang ditandai oleh algoritma ML sebagai "OK" tetapi yang sebenarnya akan meninggalkan layanan Anda) dan positif palsu (pengguna yang ditandai oleh algoritma ML sebagai "berisiko" tetapi yang sebenarnya OK). Dalam kasus pertama, Anda akan kehilangan kesempatan untuk mencoba mempertahankan pelanggan Anda, sementara dalam kasus kedua Anda akan melakukan tindakan retensi yang sebenarnya bisa Anda hindari. Ada biaya yang terkait dengan kedua kesalahan tersebut, dan menemukan keseimbangan antara keduanya adalah cara untuk memutuskan KPI terbaik Anda.

Hal yang sama berlaku untuk algoritma kelayakan pinjaman. Dalam kasus ini, jika kita mendapatkan hasil positif palsu, kita memberikan uang kepada bisnis yang tidak akan membayar kembali pinjamannya, dan ini merupakan kerugian bagi kita. Hasil negatif palsu berarti kita menolak pinjaman kepada bisnis kecil yang sebenarnya akan membayar kita, yang merupakan peluang kerugian bagi kita (dan peluang yang terlewatkan bagi pelanggan kita). Di sini, Anda juga dapat menghitung dan menemukan titik optimal antara kedua kesalahan tersebut yang akan mewakili KPI Anda.

Sebagai contoh, mari kita asumsikan metrik-metrik berikut ini yang kita anggap dapat diterima untuk tiga contoh mainan kita:

- Recall lebih tinggi dari 90%
- Prediksi dalam $\pm 3\%$ dari PUE sebenarnya 95% dari waktu
- Presisi lebih tinggi dari 95%

Blok Output ML di kolom kiri Gambar 7.9 adalah tempat ML berperan. Di sini, Anda perlu spesifik tentang apa yang Anda inginkan dari ML untuk membantu Anda mencapai tujuan. Kembali ke tiga contoh kita, berikut adalah pilihan potensial untuk tiga proyek yang sedang kita kerjakan:

- ✓ Kita akan memprediksi probabilitas churn untuk setiap pelanggan.

- ✓ Kita akan memprediksi kinerja sistem pendingin di masa mendatang.
- ✓ Kita akan mengevaluasi kelayakan pinjaman setiap pelanggan.

Setelah Anda mengetahui apa yang akan diberikan oleh algoritma ML Anda untuk membangun aplikasi, Anda dapat mulai memikirkan input ML-nya. Kami menyebutnya sebagai fitur di bagian pertama buku ini. Ingatlah bahwa titik awal yang baik untuk menghasilkan fitur yang relevan adalah dengan bertanya pada diri sendiri, "Informasi apa yang saya perlukan jika saya harus mengambil alih tugas algoritma ini?" Daftar ini mungkin tidak lengkap, dan Anda selalu dapat menambahkan (atau menghapus) item, terutama dengan bantuan ilmuwan data atau setelah Anda mendapatkan pengalaman dengan model tersebut. Namun, mengisi draf pertama fitur-fitur ini memberi Anda titik awal untuk memastikan bahwa Anda merancang sesuatu yang dapat dibangun dan dipahami oleh pengembang. Berikut beberapa contoh cara mengisi kotak-kotak ini dengan contoh-contoh sebelumnya:

- ✚ Penggunaan produk, demografi, jenis langganan
- ✚ Suhu luar ruangan, kelembapan, tekanan air, beban pusat data
- ✚ Arus kas, ukuran bisnis, lokasi, industri

Langkah selanjutnya dan terakhir adalah memikirkan sumber data. Dalam dua langkah sebelumnya, Anda telah mengidentifikasi jenis informasi yang perlu dipelajari oleh model; sekarang saatnya memikirkan dari mana mendapatkan data ini. Anda dapat memiliki sumber data sebanyak yang dibutuhkan, dan Anda dapat menggambar semua kotak yang Anda inginkan di Kanvas untuk memperhitungkan semua sumber informasi yang ingin Anda gunakan. Untuk ketiga contoh tersebut, berikut adalah beberapa kemungkinan jawaban:

- ❖ Untuk output, pelanggan yang sebelumnya berhenti berlangganan. Untuk input, data CRM.
- ❖ Log peralatan pusat data dan data cuaca.
- ❖ Data CRM, data transaksi, data pasar keuangan.

Kanvas Framing memandu Anda melalui proses yang menghasilkan spesifikasi ML yang terdefinisi dengan jelas yang dapat Anda teruskan kepada tim teknis untuk diimplementasikan. Di bagian selanjutnya, kita akan membahas implementasi praktis Kanvas Framing berdasarkan studi kasus yang kami sajikan di bagian 1.

7.7 LATIHAN STUDI KASUS

Cara terbaik bagi Anda untuk benar-benar memahami cara membingkai proyek AI Anda adalah dengan mengisi Kanvas Framing dengan berlatih pada studi kasus yang telah kami sajikan di seluruh buku ini. Kami menyarankan Anda menerapkan Kanvas Pembingkai pada ketiga studi kasus ini:

- ❖ Tingkatkan penargetan dengan menganalisis perilaku pelanggan (studi kasus Opower di bab 3)
- ❖ Otomatiskan proses industri dengan visi komputer (studi kasus pemilahan mentimun di bab 4)
- ❖ Bantu pelanggan memilih konten (studi kasus Netflix di bab 6)

Untuk masing-masing studi kasus ini, ikuti langkah-langkah yang telah kami lakukan di bagian 7.3, dan coba pikirkan apa yang akan Anda lakukan jika Anda memimpin proyek tersebut. Mulailah dengan menuliskan deskripsi singkat tentang tujuan bisnis proyek, dan tentukan KPI-nya serta kriteria penerimaan yang sesuai. Temukan definisi proyek yang ramah ML dalam hal input dan output ML, dan lakukan curah pendapat tentang sumber data yang berpotensi Anda gunakan. Dengan membaca studi kasus itu sendiri, Anda seharusnya memiliki semua informasi yang dibutuhkan untuk latihan ini. Jangan takut untuk mengambil langkah yang berbeda. Pada tahap proses ini, hanya sedikit pilihan yang salah, dan eksperimen adalah satu-satunya cara untuk mengujinya. Setelah selesai, bandingkan pemikiran Anda dengan komentar kami dan kanvas yang telah selesai untuk setiap latihan.

Meningkatkan Penargetan Pelanggan

Mari kita mulai dengan studi kasus praktik pertama: Opower dan upayanya untuk memahami pola penggunaan energi pelanggannya guna menyempurnakan aktivitas penargetan. Sekilas, tujuan bisnisnya jelas: "memahami bagaimana pelanggan kami menggunakan energi." Ini adalah tujuan yang layak, yang bahkan mungkin menjadi presentasi awal yang menarik. Namun, jika Anda berhenti sejenak, Anda akan menyadari bahwa ini masih merupakan pernyataan yang cukup abstrak yang perlu diklarifikasi sebelum Anda dapat melanjutkan. Misalnya, jika tujuan bisnis Anda adalah "memahami bagaimana pelanggan kami menggunakan energi," bagaimana Anda akan melacak hasil dan manfaat model Anda? Bahkan, kami harap Anda menyadari hal ini segera setelah Anda mencoba menentukan KPI untuk tujuan ini.

Sebaliknya, cobalah untuk berfokus pada istilah yang lebih konkret untuk proyek Anda. "Memahami kebutuhan energi pelanggan kami" cukup terfokus untuk pengambilan keputusan tingkat eksekutif yang luas, tetapi belum cukup konkret untuk memulai proyek AI. Bagaimana dengan ini?

Kelompokkan pelanggan berdasarkan pola konsumsi energi mereka.

Berinvestasi dalam upaya untuk menentukan tujuan yang lebih konkret akan langsung membuahkan hasil, karena Anda akan merasa mudah untuk menentukan KPI yang tepat guna melacak keberhasilan proyek. Dalam kasus Opower, metrik yang jelas adalah sebagai berikut:

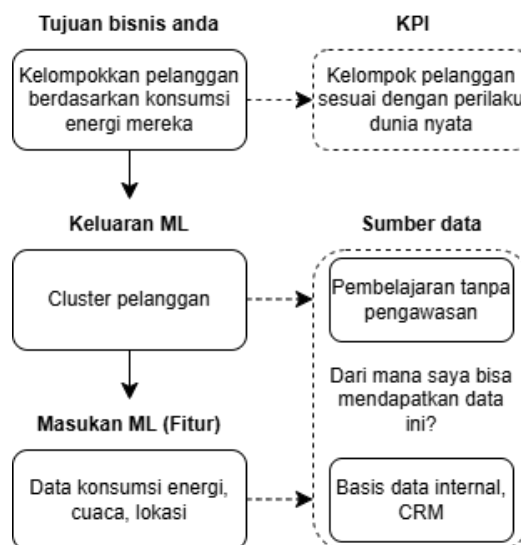
Perilaku yang diekstraksi dari pola konsumsi energi akurat.

Perhatikan bahwa menuliskan KPI berfungsi sebagai pemeriksaan kewarasan terhadap tujuan bisnis yang kita temukan di langkah sebelumnya. Sebagai aturan praktis, jika tujuan tersebut tidak mudah diukur, masih ada beberapa pekerjaan yang harus dilakukan. Selain KPI, Anda juga perlu menentukan kriteria penerimaan yang akan memberi tahu Anda kapan proyek cukup baik untuk dikirimkan kepada pelanggan. Tergantung pada proyeknya, Anda mungkin mempertimbangkan beberapa faktor saat menetapkan angka-angka ini: seberapa penting

proyek tersebut bagi organisasi, risiko keselamatan dan keuangan, dan sebagainya. Berikut ini yang kami anggap sebagai titik awal yang baik untuk proyek ini:

Kurang dari 10% pelanggan salah diklasifikasikan jika dibandingkan dengan wawancara perilaku.

Ini adalah KPI klasik yang dinyatakan dalam hal akurasi model. Kita akan melihat lebih banyak cara gabungan untuk mendefinisikan KPI dan kriteria penerimaan dalam dua contoh berikut.



Gambar 7.10 menunjukkan Framing Canvas yang telah selesai untuk latihan ini.

Pada tahap ini dalam proses Framing Canvas, Anda seharusnya sudah memiliki cukup detail di kepala untuk memulai dengan pertanyaan-pertanyaan spesifik ML. Berdasarkan tujuan bisnis yang baru saja kita uraikan, keluaran ML yang kita harapkan dari model ini adalah pengelompokan konsumen energi. Dengan kata lain, kita ingin model ini mengelompokkan pelanggan dengan pola konsumsi yang serupa. Dalam hal masukan ML, kandidat yang jelas untuk proyek ini adalah data meteran energi yang telah dikumpulkan Opower pada granularitas kedua. Namun, kita juga dapat membayangkan menambahkan data demografi, cuaca, dan lokasi ke dalam campuran, karena data-data tersebut kemungkinan akan meningkatkan akurasi model yang dihasilkan. Ini berarti bahwa data akan dikumpulkan melalui jaringan meteran energi yang ada, dan dilengkapi dengan sumber data eksternal yang dibeli dari pihak ketiga.

Otomatisasi Proses Industri

Mari kita lanjutkan ke studi kasus kedua. Seperti yang Anda ingat, proyek ini membahas otomatisasi pemilahan mentimun di fasilitas pertanian skala kecil. Tujuan bisnis yang tepat, meskipun cukup luas, adalah "meningkatkan efisiensi pertanian." Namun, dari pembahasan kita tentang contoh sebelumnya, Anda telah belajar untuk lebih spesifik, sehingga kami mengusulkan tujuan bisnis berikut untuk proyek ini:

Otomatiskan pemilahan mentimun berdasarkan tingkat kualitas.

KPI yang baik untuk mendukung hal ini adalah sebagai berikut:

Mentimun dipilah tanpa campur tangan manusia.

Ini adalah tujuan yang cukup ambisius, setidaknya saat Anda sedang menjalankan proyek ini. Namun, ini memberi kita dasar yang baik untuk menetapkan kriteria penerimaan:

Kurang dari 5% mentimun yang salah diklasifikasikan ke dalam kelas kualitas yang lebih tinggi (yang berdampak buruk bagi reputasi), dan kurang dari 10% yang salah diklasifikasikan ke dalam kelas kualitas yang lebih rendah (yang merugikan pendapatan).

Alternatifnya, kita bisa membingkai KPI dan kriteria penerimaan dalam hal penghematan sumber daya manusia, atau bahkan peningkatan laju produksi. Keluaran Pembelajaran Mesin (ML) untuk proyek ini akan menjadi sebagai berikut:

Klasifikasi mentimun ke dalam tingkat kualitas yang digunakan untuk dijual

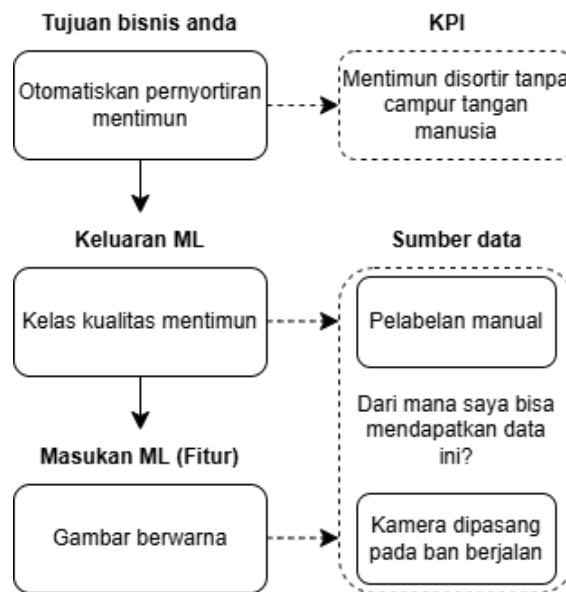
Ini adalah masalah klasifikasi citra yang cukup standar seperti yang dijelaskan di seluruh bab 4. Untuk proyek ini, kandidat yang paling jelas untuk input ML juga merupakan yang benar:

Gambar berwarna beresolusi tinggi mentimun dari perspektif atas-bawah dan dari kedua sisi

Berat mentimun juga kemungkinan berguna untuk meningkatkan akurasi klasifikasi. Solusi langsung untuk sumber data adalah sebagai berikut:

Gambar dikumpulkan dari kamera yang terpasang di ban berjalan pemilihan mentimun. Pencahayaan, pengaturan waktu, dan orientasi dapat disesuaikan untuk hasil terbaik.

Simpulannya, latihan ini merupakan aplikasi buku teks dari masalah klasifikasi citra, di mana bagian dari pekerjaan menyiapkan strategi pengumpulan data akan melibatkan pengaturan kamera di sepanjang ban berjalan.



Gambar 7.11 Kanvas Pembimbingan lengkap untuk latihan klasifikasi mentimun menunjukkan Kanvas Pembimbingan yang telah selesai untuk latihan ini.

Membantu Pelanggan Memilih Konten

Latihan terakhir membahas tentang halaman beranda konten video Netflix yang dipersonalisasi. Secara umum, visinya adalah "membantu konsumen menemukan konten untuk ditonton di platform Netflix." Berikut tujuan bisnisnya:

Mengurangi waktu yang dihabiskan pemirsa untuk memilih konten yang akan ditonton.

Cara alami untuk menemukan KPI untuk tujuan ini adalah dengan membandingkan kinerja halaman beranda yang dipersonalisasi dengan pilihan konten acak di platform Netflix. Ini adalah cara yang bagus untuk segera melacak dampak proyek, dan cara yang akan menyulitkan seluruh organisasi untuk mengabaikan proyek Anda. Beginilah tampilan pilihan ini:

Pengguna yang menelusuri halaman beranda yang dipersonalisasi lebih cepat memilih konten yang akan ditonton.

Dan berikut adalah kriteria penerimaan yang sesuai:

Pengguna yang menelusuri halaman beranda yang dipersonalisasi 20% lebih cepat dalam memilih konten yang akan ditonton dibandingkan dengan daftar acak.

Secara umum, Anda harus selalu mencoba menghubungkan keberhasilan proyek Anda dengan metrik perilaku pengguna, karena itulah cara paling langsung untuk mengukur dampak upaya Anda. Konsep ini dikenal luas sebagai pengujian A/B, karena Anda membandingkan varian baru suatu fitur atau produk (berdasarkan model AI baru Anda) dengan status quo. Kandidat terbaik

untuk pengujian A/B adalah proyek-proyek yang memiliki aliran pengguna atau item yang konstan, sehingga Anda dapat membaginya menjadi dua kelompok: satu yang Anda kirim ke sistem lama, dan satu yang Anda kirim ke sistem baru. Dengan membandingkan kinerja di kedua kelompok pengguna, Anda akan mendapatkan gambaran terbaik tentang bagaimana proyek Anda berjalan.

Beralih ke bagian Keluaran ML di Canvas, kita perlu spesifik tentang bagaimana keluaran model ML akan digunakan untuk mempersonalisasi halaman beranda. Berikut opsi yang paling mudah:

Daftar lima film yang kemungkinan besar akan disukai pengguna

Namun perlu diingat bahwa, seperti yang Anda lihat di akhir bab 6, implementasi di dunia nyata akan memiliki beberapa daftar berbeda untuk meningkatkan peluangnya menarik perhatian pengguna. Kami yakin Anda menyadari bahwa contoh ini merupakan penerapan umum sistem rekomendasi, yang berarti masukan ML akan terlihat seperti ini:

*Riwayat tontonan pengguna target dan semua pengguna lain di platform.
Informasi tentang film (sutradara, tahun, genre, dan sebagainya).*

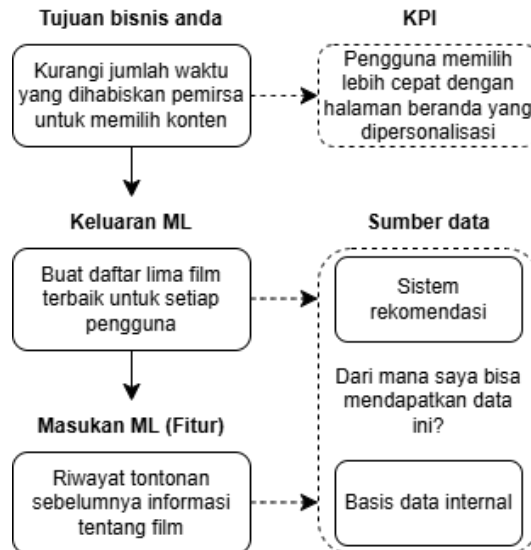
Dari mana Anda akan mengumpulkan data ini? Kami tidak memiliki banyak opsi di sini:

*Riwayat tontonan yang dikumpulkan dari server video. Informasi film yang
dikumpulkan dari basis data internal.*

Faktanya, contoh ini tipikal: Anda tidak perlu mencari terlalu jauh untuk aset data, karena aset tersebut sudah krusial bagi operasional organisasi, sehingga kemungkinan besar akan dikumpulkan dan dikelola dengan cermat. Sederhananya, setiap platform streaming video sudah menyimpan informasi tentang genre film, aktor, dan sutradara sehingga dapat menggunakan data tersebut saat menayangkan film kepada pengguna. Tidak sulit untuk mengambil informasi ini dan menggunakannya untuk pembelajaran mesin. Demikian pula, riwayat penayangan juga mudah diakses karena harus dilacak untuk keperluan akuntansi (misalnya, untuk membayar royalti). Gambar 7.12 menunjukkan Kanvas Pembimbingan yang telah selesai untuk latihan ini.

Kita telah membahas banyak materi dalam bab ini. Anda telah mempelajari perbedaan antara visi AI jangka panjang dan proyek AI individual, serta bagaimana proyek AI individual membantu organisasi Anda membentuk visi AI. Kita juga mengeksplorasi atribut-atribut yang membuat proyek AI berharga bagi organisasi, dan memperingatkan Anda tentang potensi risiko dan jebakan yang dihadapi saat meluncurkan proyek baru. Terakhir, kami memperkenalkan Framing Canvas sebagai alat penting yang membantu Anda menyaring "bahasa bisnis" menjadi "bahasa pembelajaran mesin".

Bab selanjutnya akan menunjukkan cara menggunakan definisi yang ramah pembelajaran mesin ini dan menyusun tim serta rencana yang dapat digunakan organisasi Anda untuk mengimplementasikan proyek Anda.



Gambar 7.12 Kanvas Pembingkaian yang telah selesai untuk latihan Netflix

- ❖ Visi AI organisasi Anda adalah jalur yang jelas yang mendefinisikan bagaimana AI akan terlihat di era AI, bagaimana AI akan membantu karyawannya bekerja lebih efisien, dan bagaimana AI akan membentuk produk Anda untuk melayani pelanggan.
- ❖ Proyek AI adalah inisiatif yang terdefinisi dengan baik yang menggunakan AI untuk membangun produk atau fitur baru, atau untuk mengoptimalkan proses yang ada.
- ❖ Beberapa proyek AI yang kompleks perlu dibagi menjadi tugas-tugas ML yang independen, yang mendefinisikan KPI, input dan output, serta sumber data yang ingin Anda gunakan. Kanvas Framing membantu Anda melakukannya.
- ❖ Kanvas Framing membantu Anda merumuskan tujuan dan visi untuk proyek AI dalam istilah ML.

BAB 8

SIAPKAN DATA, TEKNOLOGI, DAN SDM

Bab ini membahas

- Mengidentifikasi sumber data potensial, baik di dalam maupun di luar organisasi
- Menilai kualitas dan kuantitas data
- Membentuk tim AI yang efektif

Bab ini melanjutkan pembahasan dari bab 7. Setelah Anda mengetahui cara menggunakan Kanvas Pembingkaian untuk menciptakan visi yang ramah ML untuk proyek Anda, saatnya untuk menyiapkan bahan-bahan lain yang Anda butuhkan. Mewujudkan sebuah proyek membutuhkan tiga bahan utama: model ML, data, dan sumber daya manusia. Meskipun memilih model ML yang baik merupakan tugas tim teknis, tugas Anda adalah merekrut mereka dan menyusun strategi data agar mereka dapat bekerja. Bab ini berfokus pada cara menemukan dan mengelola data, serta cara merekrut tim yang terdiri dari orang-orang berbakat dengan keterampilan yang tepat untuk proyek Anda.

8.1 STRATEGI DATA

Salah satu tujuan kami dalam menulis buku ini adalah untuk membuat Anda berpikir kritis tentang data dan memahami bagaimana para insinyur menggunakannya untuk membangun model ML. Karena data sangat krusial, mengembangkan strategi data yang koheren sangat penting bagi keberhasilan setiap proyek. Di bagian 1, kami membahas data dengan cara yang mengasumsikan bahwa Anda memilikinya secara pasti untuk membangun model Anda. Anda mungkin bisa menebak bahwa hal ini jarang terjadi; bab ini akan mengisi kekosongan tersebut dan membantu Anda memahami berapa banyak data yang Anda butuhkan, di mana Anda bisa mendapatkannya, dan bagaimana mengelolanya.

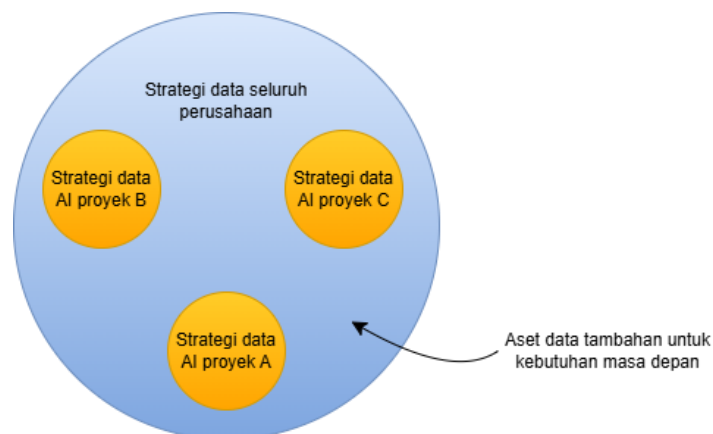
Satu perbedaan yang harus kita buat adalah antara strategi data organisasi Anda dan strategi data proyek AI Anda. Ketika media bisnis atau eksekutif berbicara tentang strategi data, mereka umumnya berbicara tentang strategi perusahaan secara keseluruhan untuk memperoleh, menyimpan, dan menggunakan data. Strategi ini didorong oleh tujuan jangka panjang dan visi perusahaan untuk masa depan.

Buku ini berfokus pada strategi data proyek Anda, yang spesifik untuk satu inisiatif AI. Memiliki fokus yang spesifik pada satu proyek lebih baik daripada fokus yang luas karena dua alasan. Pertama, mengembangkan strategi data, bahkan untuk satu proyek, memaksa Anda untuk bersikap konkret tentang data spesifik yang perlu dikumpulkan untuk membangun proyek Anda, berapa banyak yang dibutuhkan, dan di mana Anda akan mendapatkannya. Untuk mengembangkan strategi AI, Anda mungkin berkata, "Kita akan membangun infrastruktur untuk mengumpulkan data dari interaksi pengguna dengan platform." Namun, untuk membangun proyek AI, Anda perlu berkata, "Kita akan menggunakan klik pengguna, waktu yang mereka habiskan di setiap halaman, dan pembagian sosial, lalu

menggabungkannya dengan data CRM untuk merekomendasikan konten." Tingkat spesifikasi ini membantu Anda memikirkan apa yang benar-benar perlu dikumpulkan untuk membangun proyek Anda. Seseorang harus membuat keputusan tentang apa yang harus dikumpulkan, dan jika arahan yang Anda berikan kepada tim TI terlalu luas, mereka akan membuat keputusan itu untuk Anda, seringkali tanpa mengetahui konteks bisnisnya.

Alasan kedua (dan yang lebih penting) untuk memikirkan strategi data proyek AI adalah bahwa strategi di seluruh organisasi harus diinformasikan oleh kebutuhan setiap proyek AI yang akan Anda kejar. Mulai memikirkan strategi data di seluruh perusahaan tanpa pengalaman sama seperti mulai memikirkan rumah keluarga Anda saat masih lajang. Anda sama sekali tidak tahu apa yang akan Anda butuhkan. (Apakah Anda akan punya nol, satu, atau lima anak? Lokasi mana yang lebih nyaman bagi Anda dan pasangan? Apa selera pasangan Anda? Apa yang mampu Anda beli?) Sebagaimana proyek AI individual membantu Anda membentuk visi AI yang lebih terinformasi, membangun strategi data Anda sebagai kumpulan data minimal yang dibutuhkan untuk proyek-proyek tersebut juga membantu Anda memastikan bahwa Anda mengumpulkan data yang Anda butuhkan.

Jika Anda memulai dari proyek AI, membangun strategi data organisasi akan menjadi latihan sederhana dengan menggabungkan pengalaman yang Anda peroleh melalui implementasi setiap proyek (seperti yang ditunjukkan pada gambar 8.1). Untuk membangun strategi data untuk proyek AI, Anda perlu tahu di mana mendapatkan data dan berapa banyak yang Anda butuhkan. Dua subbagian berikutnya akan membahas setiap topik secara mendalam.



Gambar 8.1 Strategi data organisasi menggabungkan kebutuhan masing-masing proyek individu (ditambah beberapa insidental yang disertakan untuk masa mendatang).

Dari Mana Saya Mendapatkan Data?

Data telah menjadi konsep fundamental di seluruh buku ini. Kami telah membahas berbagai jenis data: data inti terstruktur, gambar, serta suara dan bahasa alami. Jika Anda ingin membangun proyek AI, perbedaan yang lebih penting akan memengaruhi Anda: data yang Anda miliki, dan data yang tidak Anda miliki.

Mari kita mulai dengan asumsi bahwa organisasi Anda memiliki data yang Anda butuhkan untuk membangun proyek AI Anda. Karena data ini diproduksi dan dimiliki oleh

organisasi Anda, kami akan menyebutnya data internal. Bab 2 telah membahas nilai dari jenis data internal tertentu yang kami sebut data bisnis inti: data yang berdampak langsung pada laba atas atau laba bersih organisasi. Untuk platform real estat kami, ini bisa berupa data harga rumah; untuk Square Capital, ini adalah transaksi pelanggan mereka; atau untuk Google, data konsumsi pusat datanya (pada dasarnya, satu-satunya biaya variabelnya).

Organisasi Anda kemungkinan juga menghasilkan data dari proses bisnis yang mungkin tidak terkait langsung dengan sumber pendapatan atau biaya utama Anda. Kami akan menyebut jenis data ini sebagai data tambahan. Dalam contoh real estat kita, foto rumah dan ulasan rumah dapat dianggap sebagai data tambahan: tetap menarik dan bermanfaat, tetapi tidak sepenting karakteristik rumah yang dijual dan harga jualnya.

Sebagai contoh lain yang lebih konkret, mari kita pertimbangkan platform e-commerce seperti Amazon, Zappos di Amerika Serikat, atau Zalando di Eropa. Dalam kasus mereka, data bisnis inti adalah pembelian setiap pelanggan, karena berkorelasi langsung dengan pendapatan perusahaan. Data tambahan dapat berupa halaman yang dikunjungi pelanggan, email yang dibuka, ulasan produk, dan sebagainya. Anda dapat membangun proyek yang luar biasa dengan data tambahan, tetapi kemungkinan besar dampaknya tidak sebesar yang dapat Anda bangun dengan data bisnis inti.

Sebelum melanjutkan, kami ingin mengingatkan Anda bahwa memiliki kumpulan data tidak berarti Anda dapat menggunakannya tanpa kesulitan. Anda mungkin berencana menggunakan data pelanggan tetapi tidak memiliki izin untuk melakukannya—entah pelanggan menolak izin mereka atau, seperti yang sering terjadi, kebijakan privasi Anda tidak ditulis untuk memperhitungkan penggunaan yang Anda rencanakan saat ini. Sekalipun Anda memiliki izin untuk menggunakan data, berdasarkan pengalaman kami, kebanyakan orang meremehkan waktu dan tantangan yang dihadapi dalam proses transisi dari "kami memiliki data" hingga benar-benar memasukkannya ke laptop ilmuwan data, siap untuk diproses. Sekalipun Anda 100% yakin bahwa organisasi Anda memiliki data yang dibutuhkan untuk suatu proyek, saran tulus kami adalah Anda mencoba memperhitungkan segala hal yang berpotensi memperlambat Anda dalam penggunaannya.

Dalam satu proyek yang kami kerjakan, departemen hukum sebuah perusahaan mempersulit proses ekspor data yang kami butuhkan sehingga kami harus naik pesawat, masuk ke gedung mereka, mencolokkan laptop, mengenkripsi data, dan membawanya pulang dalam bentuk hard disk. Di proyek lain, kami tahu perusahaan memiliki data yang kami butuhkan, tetapi kami semua kesulitan memahami arti variabel-variabel dalam kumpulan data tersebut. Masalahnya adalah orang yang merancang proses pengumpulan data telah pensiun, dan kami harus menghabiskan sebulan penuh untuk melakukan wawancara guna memahami data yang kami lihat.

Sekarang mari kita bahas kasus di mana Anda belum memiliki data yang dibutuhkan untuk sebuah proyek. Anda memiliki tiga pilihan:

- Mulai mengumpulkannya dan tunggu hingga Anda memilikinya.
- Cari data terbuka atau kumpulkan dari internet secara gratis.
- Beli dari penyedia.

Pilihan pertama adalah yang paling lambat; terkadang juga yang paling mahal, tetapi dalam beberapa kasus Anda mungkin tidak punya pilihan lain. Misalkan Anda menjalankan toko pakaian fisik dan tidak memiliki sistem kartu loyalitas atau cara lain untuk melacak siapa yang membeli apa. Jika seorang pelanggan datang, Anda tidak melacak bahwa namanya adalah John dan dia baru saja membeli sepatu Nike Air Jordan, ukuran 10, seharga Rp. 1.290.000. Jika Anda belum menyimpan informasi ini, apakah menurut Anda ada perusahaan lain yang memilikinya? Tentu saja, dalam kasus ini, satu-satunya pilihan Anda adalah membangun proses baru untuk mengumpulkan informasi ini, lalu menunggu hingga Anda mengumpulkan cukup data.

Di sisi lain, terkadang proyek Anda membutuhkan data yang tidak dapat Anda kumpulkan atau tidak ingin Anda upayakan untuk kumpulkan. Pilihan kedua dalam kasus ini adalah menggunakan data terbuka.

Sungguh menakjubkan betapa banyak data gratis yang bisa Anda dapatkan dari internet. Contohnya adalah data terbuka yang seringkali dijadikan sumber terbuka oleh pemerintah, yang dapat Anda gunakan secara bebas untuk proyek Anda. Misalnya, Anda dapat menggunakan data pendapatan negara Anda untuk memfokuskan pemasaran Anda di wilayah-wilayah terkaya. Berikut beberapa tempat bagus lainnya untuk mencari kumpulan data terbuka:

- *kaggle.com*—Situs web tempat perusahaan atau individu dapat menyelenggarakan kompetisi ML dan mengunggah kumpulan data.
- *arxiv.org*—Repositori gratis untuk makalah ilmiah. Saat para peneliti mengumpulkan kumpulan data baru, mereka menulis dan merilis makalah ilmiah untuk dipresentasikan kepada komunitas ilmiah dan seringkali juga merilis kumpulan data mereka.
- *github.com*—Repositori kode sumber terbuka.

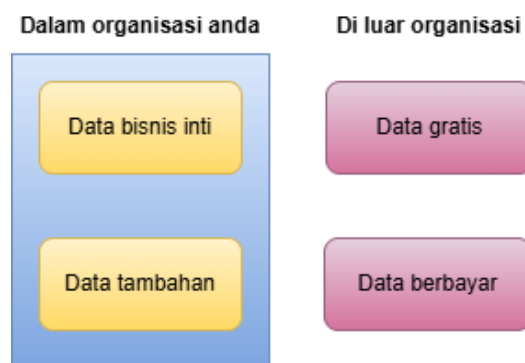
Saat menggunakan dataset pihak ketiga, Anda perlu memperhatikan kualitas data dan masalah hukumnya. Banyak dari dataset ini diproduksi berdasarkan "upaya terbaik" dan dikumpulkan sesuai kebutuhan untuk mendukung algoritma atau bidang aplikasi baru; seringkali kualitas data tidak terjamin. Selain itu, banyak dataset yang tersedia untuk umum dirilis dengan lisensi nonkomersial, yang berarti tidak dapat digunakan untuk tujuan bisnis; selalu pastikan Anda dapat menggunakan dataset terbuka untuk tujuan yang Anda inginkan.

Pilihan ketiga adalah membeli data dari penyedia. Strategi ini terkadang bisa sangat mahal, dan juga dapat menempatkan Anda pada posisi berbahaya karena Anda mungkin akan bergantung pada penyedia ini selamanya. Sebelum memutuskan untuk menempuh jalur ini, kami sarankan Anda meluangkan waktu yang cukup untuk mengevaluasi biaya jangka panjang terkait dan menentukan apakah ada beberapa penyedia yang dapat Anda beli. Jika hanya ada satu, Anda perlu menemukan strategi untuk melindungi diri Anda jika mereka gulung tikar atau strategi bisnis mereka berubah dan mereka berhenti menjual data mereka kepada Anda.

Dalam kedua kasus tersebut, kami mendorong Anda untuk berpikir kritis tentang kumpulan data yang berasal dari luar organisasi Anda. Menurut definisi, jika beberapa data tersedia secara terbuka baik gratis maupun berbayar berarti semua orang juga dapat memperolehnya. Seringkali, proyek yang didasarkan pada sumber data eksternal akan lebih

sulit di pasaran karena dapat dengan mudah direplikasi oleh pesaing. Beberapa proyek AI terkuat dibangun di atas fondasi kumpulan data kepemilikan yang sulit direplikasi oleh pihak lain. Singkatnya, dinding organisasi Anda (atau dinding virtual pusat datanya) membuat perbedaan besar dalam strategi data proyek AI Anda, seperti yang ditunjukkan pada Gambar 8.2. Jika data berada dalam batasan organisasi Anda, Anda memiliki kumpulan data unik dan sangat berharga yang dapat Anda gunakan. Jika tidak, Anda harus mendapatkannya secara gratis atau membelinya. Dalam kedua kasus tersebut, pertimbangkan liabilitas yang menyertai pilihan tersebut.

Tentu saja, Anda dapat menggabungkan data internal dan eksternal. Sebenarnya, ini seringkali merupakan ide yang bagus. Contohnya adalah prediktor harga rumah real estat kami: kami dapat menggunakan data pemerintah gratis tentang pendapatan setiap lingkungan untuk meningkatkan model. Anda bahkan dapat menggunakan data dari OpenStreetMap atau dari Google Maps untuk memeriksa keberadaan layanan dan transportasi umum di berbagai lingkungan, menambahkan dimensi lain pada prediktor harga rumah.



Gambar 8.2 Dinding digital atau fisik organisasi Anda menentukan batasan antara jenis data: data inti dan tambahan dalam organisasi Anda, dan data gratis atau berbayar di luarnya.

Untuk proyek yang kami bangun untuk organisasi besar, kami menggunakan semua jenis data. Perusahaan memiliki data penjualan dari toko-tokonya, tetapi kami ingin melihat korelasi antara kinerja penjualan dan demografi, jadi kami mengumpulkan data sensus gratis. Data sensus yang tersedia gratis sebagian besar cukup baik, tetapi untuk beberapa kota, kami membutuhkan gambaran populasi yang lebih rinci. Kami kemudian beralih ke penyedia eksternal khusus dan mengintegrasikan data mereka juga. Perhatikan bahwa meskipun data penjualan diperbarui setiap hari, demografi berubah jauh lebih lambat, sehingga mengandalkan sumber eksternal bukanlah masalah.

Dari mana pun Anda mendapatkan data, sepotong informasi dapat menentukan keberhasilan atau kegagalan proyek Anda: label. Ingatlah bahwa jika Anda melatih algoritma pembelajaran terawasi, komputer perlu belajar menghasilkan angka atau kelas (label) berdasarkan angka lain (fitur). Sering kali, Anda dapat mengatasi fitur yang hilang, tetapi Anda mungkin akan kesulitan jika tidak memiliki label. Dengan kata lain, dalam hal data, label jauh lebih penting daripada fitur. Mari kita gunakan contoh prediktor harga rumah. Label kita adalah

harga jual rumah, dan fitur kita dapat berupa luas persegi, jumlah kamar, keberadaan taman, dan sebagainya.

Mari kita asumsikan bahwa ketika Anda merancang antarmuka untuk situs web real estat Anda, Anda tidak terpikir untuk menambahkan kolom yang memungkinkan pengguna menentukan apakah rumah mereka memiliki taman. Oleh karena itu, Anda tidak memiliki informasi tersebut di basis data Anda dan tidak dapat menggunakannya untuk membangun model. Namun, Anda masih memiliki fitur relevan lainnya yang disertakan dalam formulir daftar rumah, termasuk luas persegi, jumlah kamar, dan lokasi. Bahkan tanpa kolom "taman", Anda kemungkinan besar masih dapat membangun model yang cukup akurat.

Di sisi lain, jika Anda lupa menanyakan harga jual kepada pengguna, Anda akan benar-benar tersesat. Tanpa label, tidak ada cara untuk membangun model pembelajaran terawasi. Label dapat dikumpulkan dengan tiga cara:

- Secara Alami
- Peretasan
- Pembayaran

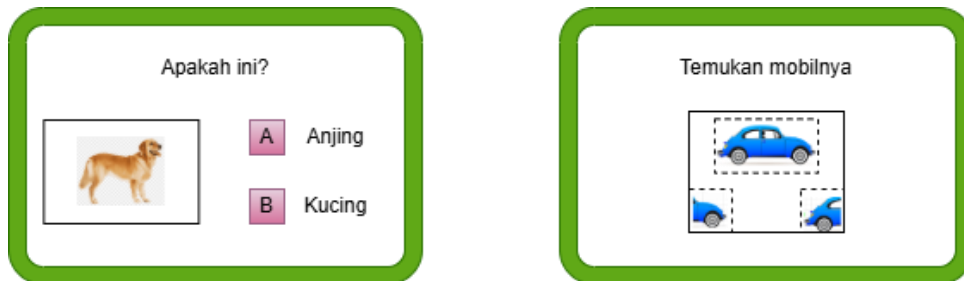
Label alami dihasilkan oleh proses bisnis Anda. Misalnya, jika platform real estat Anda meminta klien untuk memasukkan harga jual rumah mereka saat mereka menghapus daftar mereka, Anda akan mendapatkan label tersebut secara alami. Google secara alami menyimpan data tentang kinerja energi pusat datanya karena Google yang menjalankannya. Amazon menyimpan semua barang yang Anda beli dalam basis data. Semua informasi ini disimpan untuk menjalankan bisnis, dan dapat digunakan sebagai label jika diperlukan. Terkadang, label tidak mudah didapat, tetapi Anda masih dapat menemukan kiat cerdas untuk mendapatkannya. Contohnya adalah apa yang dilakukan Amazon dengan ulasan produk.

Saat Anda menulis ulasan tentang kecintaan Anda pada penyedot debu baru, Anda juga menambahkan peringkat bintang (misalnya, dari 1 hingga 5). Skor tersebut dapat digunakan sebagai label untuk sistem analisis sentimen. Pada dasarnya, Anda memberi Amazon input (ulasan teks) dan label (skor bintang) yang dapat digunakan untuk membangun teknologi analisis sentimen, gratis. Contoh lain adalah Facebook, yang pada awalnya meminta pengguna untuk menandai teman dalam gambar dengan mengklik wajah mereka. Facebook bisa saja meminta Anda untuk menulis siapa yang ada di gambar, tetapi dengan mengklik wajah, Anda memberi Facebook label perkiraan untuk algoritma pengenalan gambar.

Akhirnya, mungkin terpikir oleh Anda bahwa saat mendaftar untuk layanan internet baru, Anda diminta dengan tugas yang membosankan, seperti menemukan mobil dalam gambar untuk membuktikan bahwa Anda manusia. Layanan ini disebut Google reCAPTCHA, dan Anda mungkin sudah bisa menebak kegunaannya: Anda memberi perusahaan label untuk algoritma ML-nya, gratis.

Dalam beberapa kasus, satu-satunya pilihan Anda adalah membayar orang untuk memberi label pada contoh. Pola umum untuk memberi label data adalah menggunakan platform crowdsourcing seperti Amazon Mechanical Turk, yang memberi Anda akses sesuai permintaan, dibayar per menit, ke tenaga kerja sementara yang tersebar di seluruh dunia. Dengan Mechanical Turk, satu-satunya hal yang dapat Anda andalkan adalah kontraktor akan

memiliki koneksi internet: karena mereka biasanya tidak terlatih, Anda harus menyiapkan materi pelatihan dan antarmuka pelabelan yang dapat digunakan pekerja untuk memilih label yang tepat untuk contoh pelatihan. Gambar 8.3 menunjukkan contoh dua antarmuka pelabelan untuk mengumpulkan label untuk tugas klasifikasi gambar dan pelokalan objek.



Gambar 8.3 Antarmuka pelabelan untuk pengenalan gambar dan pelokalan objek. Pilihan dapat dibuat menggunakan pintasan keyboard untuk meningkatkan kecepatan entri data.

Secara umum, platform crowdsourcing cocok untuk tugas pelabelan yang tidak memerlukan banyak pelatihan. Jika Anda mengerjakan proyek yang membutuhkan penalaran manusia tingkat tinggi (misalnya, menemukan sel kanker pada pemindaian mikroskop), Anda akan lebih baik jika mengumpulkan tenaga kerja terampil Anda sendiri.

Tabel 8.1 merangkum biaya dan waktu yang dibutuhkan untuk mengumpulkan label untuk ketiga strategi pelabelan.

Tabel 8.1 Kebutuhan biaya dan waktu untuk ketiga strategi pelabelan

Strategi pelabelan	Biaya	Waktu yang dibutuhkan
Label alami (gratis).	Nol-Anda sudah mengumpulkan label ini.	Nol-Anda sudah memilikinya.
Label yang diretas	Rendah dan tetap-Anda hanya perlu menyiapkan proses pengumpulan data baru.	Tergantung pada lalu lintas Anda.
Label berbayar	Tinggi dan bervariasi-Anda membayar berdasarkan jumlah label yang Anda inginkan.	Tergantung pada waktu yang dibutuhkan untuk memberi label pada contoh dan jumlah pemberi label yang Anda miliki.

Setelah Anda memiliki gambaran yang jelas tentang di mana mendapatkan data yang Anda butuhkan, langkah selanjutnya adalah menentukan berapa banyak yang Anda butuhkan.

Berapa Banyak Data yang Saya Butuhkan?

Berdasarkan pengalaman kami sebagai konsultan, kami sering melihat orang terjebak dalam big data. Sungguh melegakan mengetahui bahwa memiliki banyak data adalah kunci untuk membuka peluang luar biasa. Namun, seperti yang akan Anda lihat di bagian selanjutnya, kualitas data seringkali lebih penting daripada kuantitas data.

Jumlah data yang Anda butuhkan untuk membangun produk AI sangat bergantung pada produk itu sendiri. Begitu banyak variabel yang berperan sehingga memberikan aturan yang tepat seperti "Anda memerlukan data dari 10.523 pelanggan untuk membangun model

prediksi churn yang 93% akurat" mustahil dilakukan dalam praktik. Yang dapat kami berikan adalah panduan yang dapat membantu mengembangkan intuisi Anda tentang persyaratan data untuk jenis masalah paling umum yang akan Anda temukan di dunia bisnis. Kami rasa masuk akal untuk membagi presentasi kami berdasarkan jenis data yang akan ditangani proyek Anda, seperti yang telah kami lakukan di bagian pertama buku ini.

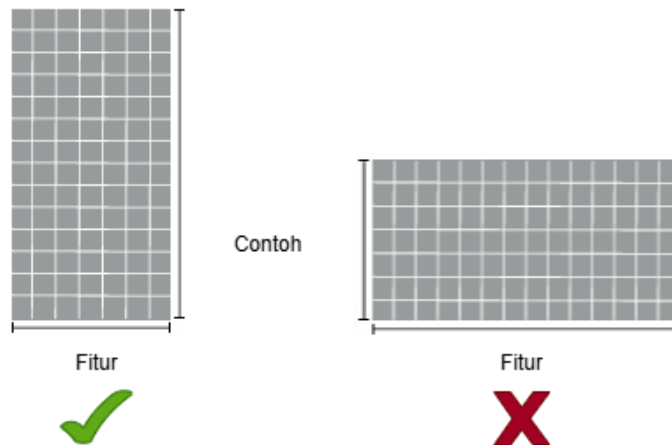
Mari kita bahas terlebih dahulu proyek-proyek yang membutuhkan data terstruktur, seperti prediksi harga rumah atau churn pelanggan. Anda perlu mempertimbangkan tiga faktor:

- Apakah target (hal yang ingin Anda prediksi) berupa angka (regresi) atau pilihan (klasifikasi)
- Untuk masalah klasifikasi, jumlah kelas yang Anda minati
- Jumlah fitur yang memengaruhi target

Mari kita mulai dengan fitur. Ingat bahwa data terstruktur adalah jenis data yang Anda lihat saat membuka lembar Excel, dan data tersebut disusun dalam baris dan kolom. Cara yang baik untuk memikirkan kebutuhan data adalah dengan membayangkan bahwa, jika Anda dapat melihat kumpulan data Anda dalam satu layar, Anda ingin data tersebut terlihat sangat tipis dan tinggi, seperti yang ditunjukkan pada gambar 8.4. Anda menginginkan lebih banyak baris daripada kolom. Hal ini karena baris menunjukkan contoh, sementara kolom menunjukkan fitur yang harus dipelajari model. Secara intuitif, semakin banyak informasi yang harus dipelajari model (semakin banyak fitur), semakin banyak contoh yang perlu dilihat untuk memahami bagaimana fitur tersebut memengaruhi target (yang berarti lebih banyak baris).

Jika Anda tidak memiliki cukup contoh dan terlalu banyak fitur, beberapa kolom mungkin tidak berguna atau menyesatkan! Ambil contoh, kumpulan data harga rumah seperti yang ada di bab 2. Menambahkan kolom dengan zodiak penjual kemungkinan besar tidak akan meningkatkan akurasi prediksi harga. Namun, model ML tidak dapat menarik kesimpulan yang masuk akal secara apriori: mereka perlu mengetahuinya dari data. Selama Anda memiliki cukup baris (contoh), sebagian besar keluarga model memang mampu melakukannya, dan akan "mengabaikan" kolom zodiak.

Namun, jika Anda memiliki terlalu sedikit data pelatihan, model akan tetap berusaha sebaik mungkin untuk memperkirakan bagaimana zodiak memengaruhi harga, sehingga mencapai kesimpulan yang secara numerik benar tetapi menyesatkan. Sebagai contoh ekstrem, bayangkan satu-satunya rumah seharga Rp. 10 miliar dalam dataset tersebut dijual oleh seorang Gemini. Tentu saja, ini tidak berarti prediksi harga rumah yang dijual oleh Gemini harus lebih tinggi daripada rumah yang dijual oleh Aries. Sebagian besar model akan dapat menghindari kesalahan ini jika dataset tersebut berisi banyak vila seharga jutaan dolar (dataset dengan banyak contoh), karena pembeli akan memiliki banyak zodiak yang berbeda dan pengaruhnya dapat diperkirakan dengan tepat. Namun, jika Anda tidak memiliki banyak contoh, model tersebut dapat "berpikir" bahwa zodiak adalah pendorong nilai rumah.

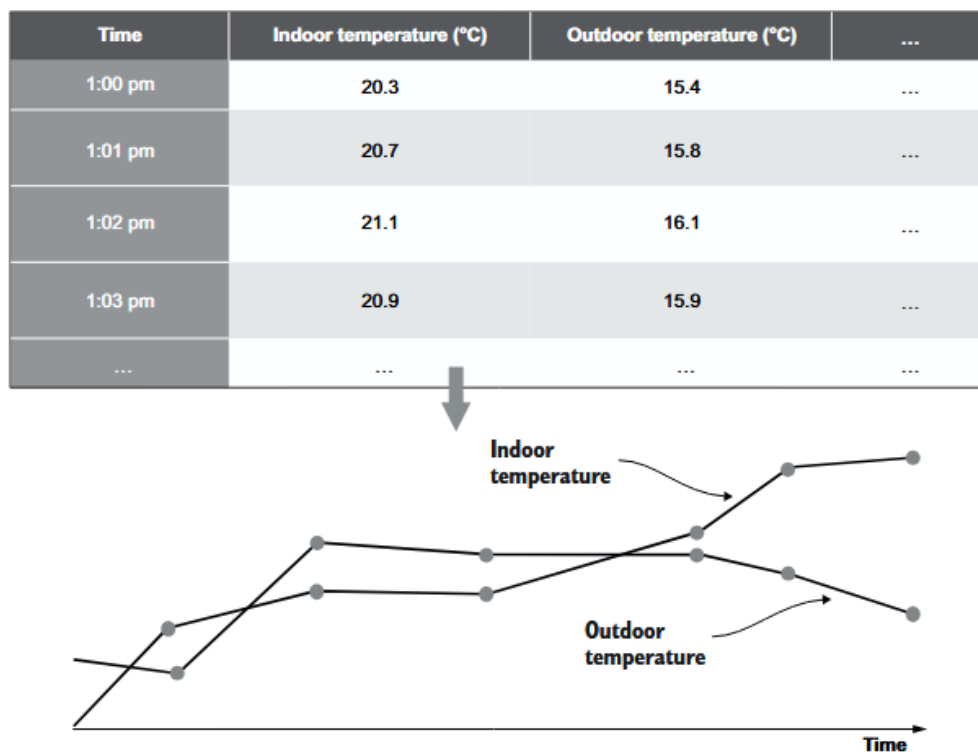


Gambar 8.4. Dataset di sebelah kiri, yang memiliki banyak contoh dan beberapa fitur (tinggi dan tipis), merupakan dataset yang baik untuk ML. Dataset di sebelah kanan, yang memiliki banyak fitur dan beberapa contoh (pendek dan lebar), bukanlah dataset yang baik untuk ML.

Mari kita bahas secara spesifik masalah klasifikasi atau regresi. Dalam masalah klasifikasi, Anda mencoba memprediksi apakah suatu contoh termasuk dalam dua kelas atau lebih. Contoh klasifikasi pertama dalam buku ini adalah sebagai berikut:

- Algoritma kelayakan pinjaman Square di bab 2 yang memberikan pelanggan salah satu dari dua kelas: memenuhi syarat untuk pinjaman atau tidak memenuhi syarat
- Prediksi churn di bab 3 yang memberi label pelanggan sebagai akan meninggalkan layanan atau tidak

Dengan asumsi Anda memiliki jumlah fitur yang sederhana (katakanlah 10), Anda harus menganggarkan setidaknya 1.000 contoh untuk setiap kelas yang Anda miliki dalam masalah Anda. Misalnya, dalam model churn pelanggan yang hanya memiliki dua kelas (loyal versus nonloyal), Anda mungkin ingin merencanakan setidaknya 2.000 contoh. Secara intuitif, semakin banyak kelas yang harus ditangani model, semakin banyak contoh yang perlu dilihat untuk mempelajari cara membedakan semua kelas. Lebih sulit memberikan aturan praktis serupa untuk model regresi karena model tersebut dapat memodelkan skenario dan fenomena yang lebih kompleks. Banyak model regresi didasarkan pada data deret waktu, jenis data khusus yang menggambarkan bagaimana pengukuran atau angka berkembang seiring waktu.



Gambar 8.5 Data deret waktu tampak seperti aliran pengukuran yang dilakukan dari waktu ke waktu (misalnya, suhu). Nilai di sebelah kanan plot lebih baru daripada nilai di sebelah kiri.

Contohnya adalah studi kasus pusat data Google dari bab 2 yang mengumpulkan pengukuran lingkungan sepanjang hari. Contoh lain yang mungkin lebih Anda kenal adalah data keuangan (misalnya, harga saham). Anda dapat melihat seperti apa data deret waktu pada Gambar 8.5.

Dalam data deret waktu, jumlah titik data yang Anda kumpulkan tidak seberapa rentang waktu pengumpulannya. Alasannya cukup intuitif: misalkan Anda adalah Jim Gao, teknisi pusat data Google dari bab 2, dan Anda mengumpulkan satu titik data per detik dari sistem pendingin udara DC Anda. Jika Anda mengumpulkan data dari Januari hingga Maret, Anda akan memiliki hampir 8 juta titik data (3 bulan $\times$ 30 hari per bulan $\times$ 24 jam per hari $\times$ 60 menit per jam $\times$ 60 detik per menit). Dengan begitu banyak titik, Anda mungkin berpikir Anda memiliki kumpulan data yang hebat, dan memang model Anda mungkin sangat akurat... selama beberapa hari. Tetapi apa yang terjadi ketika musim panas semakin dekat dan suhu meningkat? Model tidak mengetahui bagaimana pusat data berperilaku selama bulan-bulan terpanas dalam setahun karena telah dilatih hanya menggunakan data musim dingin.

Berbicara tentang AI untuk media, tugas klasifikasi gambar yang paling sederhana (bayangkan kucing versus anjing) membutuhkan sekitar beberapa ratus contoh untuk setiap kelas (himpunan data ImageNet yang terkenal memiliki sekitar 700–800 gambar per kelas). Jika Anda memilih kelas yang lebih mirip satu sama lain (seperti ras anjing), jumlah sampel yang dibutuhkan melonjak menjadi ribuan karena model akan membutuhkan lebih banyak contoh untuk mengasimilasi perbedaan yang lebih halus.

Jika Anda memanfaatkan pembelajaran transfer, Anda memulai pelatihan dari model yang sudah ada yang dilatih pada himpunan data besar seperti ImageNet (seperti yang dijelaskan di bab 4), dan Anda dapat memperoleh kinerja yang baik dalam ratusan gambar (atau bahkan puluhan untuk tugas yang sangat sederhana). Memberikan panduan serupa untuk aplikasi bahasa alami lebih sulit, karena lanskap tugasnya lebih beragam. Untuk tugas yang lebih sederhana seperti analisis sentimen atau klasifikasi topik, 300–500 contoh untuk setiap kelas mungkin sudah cukup, dengan asumsi Anda menggunakan pembelajaran transfer dan penyisipan kata.

Setelah Anda memiliki panduan tentang seberapa banyak data yang perlu dikumpulkan, Anda perlu memahami bahwa tidak semua titik data memiliki kepentingan yang sama saat melatih model. Menambahkan contoh yang sangat buruk bahkan dapat menjadi bumerang dan mengurangi akurasi model secara keseluruhan. Hal ini akan dibahas di bagian selanjutnya.

8.2 KUALITAS DATA

Banyak diskusi tentang AI di media populer menunjukkan bahwa keputusan berbasis AI secara inheren lebih rasional daripada manusia karena didasarkan pada matematika dan data. Dengan kata lain, mereka bersikeras bahwa AI didorong oleh data dan dengan demikian kebal terhadap bias dan prasangka manusia. Tiga cerita di bagian ini bertujuan untuk meyakinkan Anda sebaliknya.

Pembelajaran mesin berfokus pada menemukan pola dalam data, dan pelatihan pada data yang bias akan menyebabkan model membuat keputusan yang bias. Apa yang kita maksud dengan data yang bias? Anda telah melihat contoh di bab 4, di mana kami memberi tahu Anda tentang para peneliti yang mengumpulkan kumpulan data berisi gambar husky dan serigala, dan mulai membangun model untuk mengklasifikasikan keduanya. Sekilas, model tersebut bekerja dengan baik, tetapi para peneliti menemukan bahwa hal itu terjadi hanya karena model tersebut mengandalkan fakta bahwa semua gambar husky memiliki latar belakang salju, sementara gambar serigala tidak. Ternyata, bahkan set data standar pun terpengaruh oleh masalah yang sama: jaringan saraf cenderung berhalusinasi domba ketika disajikan gambar padang rumput tanpa domba, seperti yang ditunjukkan pada Gambar 8.6.

Demikian pula, jika Anda menunjukkan gambar seorang anak yang menggendong domba kepada jaringan saraf, kemungkinan besar jaringan tersebut akan mengklasifikasikannya sebagai anjing. Hal ini terjadi karena set data pelatihan bias: domba hanya muncul di padang hijau, dan sebagian besar hewan yang dipangku manusia adalah anjing. Karena model pembelajaran mesin (ML) sebagian besar merupakan mesin pencocokan pola yang sangat canggih, model tersebut mungkin telah mempelajari bahwa "domba" hanyalah setitik kecil putih di lanskap pegunungan, dan "anjing" adalah apa pun berkaki empat yang dekat dengan manusia.



Kawanan domba sedang merumput di lereng bukit yang hijau subur. Tag: merumput, domba, gunung, sapi, kuda

Gambar 8.6 Jaringan saraf cenderung mengasosiasikan gambar bukit hijau dengan domba, karena di sanalah sebagian besar gambar tersebut muncul dalam set data pelatihan.

(Sumber: <https://aiweirdness.com/post/171451900302>.)

Masalah yang mendasarinya adalah set pelatihan tidak lengkap: tidak ada kombinasi faktor-faktor penting yang akan membantu model dalam pencarian (otonom) untuk menemukan karakteristik terpenting dalam gambar tersebut. Misalnya, masalah ini akan diselesaikan dengan menambahkan gambar orang yang menggendong hewan lain, dan gambar domba lainnya di tempat-tempat yang tidak biasa (di pantai, di dalam rumah, di atas kapal, dan sebagainya). Dengan cara ini, model akan belajar memisahkan domba dari konteksnya dan dengan benar “memahami” apa sebenarnya domba itu.

Mari kita bahas kasus bias data lain yang lebih halus dan berbahaya. Tujuan proyek ini adalah menciptakan model yang dapat secara otomatis menyaring pasien untuk kanker kulit dengan melihat foto tubuh mereka. Ternyata, ketika dokter kulit mengambil gambar pasien yang sakit, mereka sering menempatkan penggaris di samping kanker untuk skala. Ini berarti bahwa sebagian besar foto pasien yang sakit juga memiliki penggaris di dalam bingkai. Tak perlu dikatakan lagi, model tersebut belajar untuk mendeteksi keberadaan penggaris, karena itu jauh lebih mudah dikenali daripada bercak kecil kulit yang lebih gelap.

Karena tidak memiliki kemampuan untuk mengenali bahwa penggaris adalah petunjuk yang membingungkan, model tersebut sama sekali tidak berguna pada pasien di masa mendatang. Kami menyukai kasus ini karena membuktikan bagaimana detail yang tampaknya tidak signifikan dapat memengaruhi kualitas dataset pelatihan, dan dengan demikian memengaruhi model yang dihasilkan. Dengan kata lain, model AI saat ini memiliki pemahaman yang sangat dangkal tentang dunia sehingga kesalahan mereka bisa sangat bodoh.

Dalam kasus pasien kanker dan penggaris, setidaknya seorang insinyur berpengalaman akan mampu menemukan dan menentukan masalahnya. Namun, beberapa sumber bias lebih sulit dideteksi. Katakanlah semua gambar pasien sakit diambil di ruangan dengan pencahayaan fluoresens, dan semua gambar subjek sehat diambil di ruangan dengan lampu LED. Sedikit perubahan warna pada gambar sulit ditangkap dengan mata telanjang, tetapi merupakan

petunjuk yang menarik bagi jaringan saraf apa pun. Akibatnya, Anda dihadapkan pada potensi bencana: Anda mengira model Anda telah belajar mendeteksi kanker, tetapi sebenarnya model tersebut belajar mengklasifikasikan cahaya LED.

Tema umum dari dua cerita pertama ini adalah bahwa data pelatihan memiliki dampak langsung pada perilaku model Anda. Ketidakberpihakan berbasis AI dan rasionalitas berbasis data sudah cukup! Dalam kedua kasus tersebut, masalahnya adalah kurangnya contoh yang kontradiktif dalam kumpulan data (misalnya, domba berjalan di aspal, anak-anak memeluk harimau kecil, dan kanker kulit dalam kondisi pencahayaan yang berbeda). Namun, kualitas data juga dapat menurun karena asumsi yang salah. Mari kita lihat contoh terakhir.

Selama Perang Dunia II, baik Sekutu maupun Poros terus-menerus mencari cara untuk meningkatkan pesawat tempur dan pesawat pengebom mereka. Salah satu tujuan terpenting adalah membuat pesawat lebih ringan sehingga bisa terbang lebih jauh. Karena pesawat membutuhkan mesin dan tangki bahan bakar untuk terbang, mengurangi berat berarti menghilangkan lapisan pelindung yang dapat melindungi kendaraan selama tembakan musuh, yang berpotensi menyelamatkan nyawa pilot. Para insinyur bertanya-tanya tentang area mana di pesawat yang paling rentan dan dengan demikian membutuhkan lapisan pelindung yang lebih berat. Militer mulai memeriksa rangka pesawat yang rusak dan menghasilkan gambar seperti gambar 8.7.

Tampaknya badan pesawat dan sayap mengalami kerusakan terburuk. Ini berarti bahwa perancang pesawat harus memperkuat lapisan pelindung di area yang terkena dampak keras, bukan? Berlawanan dengan intuisi, ini adalah hal yang salah untuk dilakukan! Kita hanya melihat pesawat yang berhasil pulang, dan bukan pada yang ditembak jatuh oleh musuh.

Semua ini dapat mengajarkan kita bahwa bertentangan dengan apa yang kita pikirkan di awal, pesawat dapat menahan tembakan hebat di badan pesawat tanpa kehilangan kemampuan untuk terbang. Faktanya, para perancang seharusnya memasang lapisan baja terberat pada mesin, meskipun mereka hanya menemukan beberapa lubang peluru di sana, karena itu berarti pesawat yang ditembakkan ke mesin akan jatuh.



Gambar 8.7 Selama Perang Dunia II, pesawat Sekutu yang pulang setelah menjalankan misi sering kali mendapat pukulan keras di sayap dan badan pesawat.

Dibandingkan dengan dua contoh sebelumnya, di mana datanya tidak seimbang atau tidak lengkap, kisah ini mengajarkan kita bahwa pengumpulan data bisa menyesatkan, dan membawa kita ke jalur yang sepenuhnya salah. Fenomena ini umumnya disebut sebagai bias survivorship, dan Anda akan mendapati bahwa hal ini juga sering berlaku di luar AI. Secara umum, kapan pun Anda memilih subkelompok item atau orang, Anda harus mewaspadaai efek bias survivorship yang akan menggagalkan upaya Anda.

Katakanlah Anda terus-menerus terkejut dengan banyaknya mobil berusia 20 tahun yang Anda lihat berkeliaran. Anda mungkin cenderung berpikir bahwa model yang lebih baru tidak akan bertahan lama, karena mobil masa kini terus-menerus rusak. Bias survivorship menyarankan Anda juga harus menghitung mobil-mobil tua yang sudah teronggok di tempat rongsokan setelah rusak bertahun-tahun yang lalu.

Anda tidak akan melihat satu pun mobil seperti itu berkeliaran, sehingga angkanya akan miring. Ketiga contoh ini telah menunjukkan kepada Anda bagaimana upaya pengumpulan data Anda dapat membuat Anda melewatkan gambaran besarnya dan berakhir dengan model yang tidak efektif. Namun, jangan putus asa; Ini adalah beberapa bidang penelitian yang paling aktif saat ini. Sementara itu, penawar terbaiknya adalah mengikuti pendekatan yang mendalam dan luas. Mendalam mengacu pada pengumpulan data yang efisien yang Anda butuhkan untuk membangun model yang akurat. Luas adalah tentang melengkapi pendekatan mendalam dengan kumpulan data (yang kemungkinan lebih kecil) berisi contoh-contoh yang tidak biasa dan rumit yang dapat Anda gunakan untuk memeriksa ulang hasilnya.

Ingat: tantangan utamanya adalah kita memiliki alat yang terbatas untuk menganalisis bagaimana model AI membuat keputusan. Sama seperti para ilmuwan yang tidak tahu banyak tentang bakteri sebelum penemuan mikroskop, para insinyur ML harus menebak apa yang dilakukan model tersebut. Karena satu-satunya cara untuk mempelajari bagaimana model akan bereaksi terhadap input tertentu adalah dengan menjalankannya, model harus diuji dan divalidasi berdasarkan data yang realistis. Sebagai contoh praktis, perusahaan mobil self-driving mengumpulkan data di keempat musim dalam setahun dan waktu dalam sehari untuk memastikan model deteksi objek mereka dapat mendeteksi pejalan kaki dengan andal, terlepas dari apakah mereka mengenakan celana pendek atau mantel. Dengan kata lain, pastikan Anda sedikit mengombinasikan berbagai hal dan mengambil foto-foto anjing husky di pantai. Setelah kita membahas aspek-aspek utama strategi data, sisa bab ini akan membahas jenis sumber daya yang kurang teknis, meskipun sama menantanginya: manusia.

8.3 MEREKRUT TIM AI

Bagian ini memandu Anda melalui proses perekrutan tim dengan perpaduan bakat dan pengalaman yang tepat yang dibutuhkan untuk implementasi proyek AI. Tiga kategori keterampilan utama akan bermanfaat untuk dibahas:

- *Rekayasa perangkat lunak*—Mengekstrak data dari berbagai sumber, mengintegrasikan komponen pihak ketiga dan AI, serta mengelola infrastruktur
- *Pembelajaran mesin dan ilmu data*—Memilih algoritma yang tepat untuk masalah tertentu, menyempurnakannya, dan mengevaluasi akurasi

➤ *Matematika tingkat lanjut*—Mengembangkan algoritma pembelajaran mendalam yang mutakhir

Anda mungkin tergoda untuk mencari seseorang yang memiliki semua keterampilan ini, sosok yang sering disebut sebagai unicorn: pada dasarnya tidak ada. Namun, Anda dapat menemukan berbagai orang (atau tim) yang unggul dalam satu atau lebih keterampilan ini.

Dalam buku ini, ketika membahas implementasi teknis proyek AI, kita sering merujuk pada sosok mitologis ilmuwan data. Ditetapkan oleh Harvard Business Review pada tahun 2012 sebagai "Pekerjaan Terseksi di Abad ke-21", setiap perusahaan mencarinya, tetapi tidak ada yang bisa menjelaskan secara pasti apa itu.

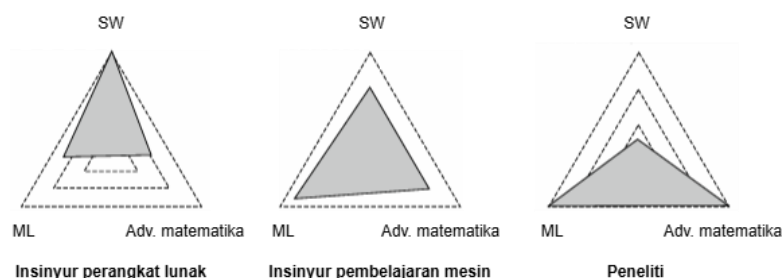
Masalahnya, dalam benak banyak orang, seorang ilmuwan data adalah orang yang serba bisa: seseorang yang dapat memecahkan segala jenis masalah yang melibatkan data, mulai dari analisis sederhana hingga perancangan model pembelajaran mesin yang kompleks. Selain itu, kita sering kali menginginkan ilmuwan data yang ideal juga menunjukkan ketajaman bisnis. Ekspektasi tentang apa yang dapat dilakukan seorang ilmuwan data begitu tinggi secara teori sehingga seringkali tidak terpenuhi dalam praktik.

Kami menyarankan Anda mencari ilmuwan data dalam kasus-kasus berikut:

- ❖ Anda belum yakin tentang arah yang akan diambil organisasi Anda. Memiliki seseorang dengan keahlian yang luas dapat membantu Anda bersikap fleksibel hingga Anda memiliki arah yang lebih jelas.
- ❖ Anda tidak memiliki kebutuhan yang sangat kompleks di sisi teknis. Jika yang Anda butuhkan hanyalah seseorang yang dapat menganalisis kumpulan data berukuran sedang (hingga beberapa gigabita) untuk membuat keputusan bisnis dan membangun beberapa model pembelajaran mesin (ML), orang yang menyebut dirinya ilmuwan data biasanya cocok.

Jika Anda secara khusus ingin membangun tim AI yang dapat menangani proyek AI yang lebih kompleks dan juga menerapkannya di dunia nyata, Anda mungkin perlu mulai merekrut orang-orang yang memiliki spesialisasi di berbagai keahlian. Kita dapat mengidentifikasi tiga kategori utama talenta, masing-masing dengan keahliannya sendiri:

- Insinyur perangkat lunak
- Insinyur pembelajaran mesin
- Peneliti



Gambar 8.8 Insinyur perangkat lunak, insinyur pembelajaran mesin (ML), dan peneliti pada umumnya memiliki tingkat keterampilan yang berbeda di bidang rekayasa perangkat lunak (SW), pembelajaran mesin (ML), dan matematika.

Gambar 8.8 membantu Anda memvisualisasikan bagaimana masing-masing peran ini memiliki komposisi yang berbeda dari tiga keterampilan inti AI: rekayasa perangkat lunak (SW), pembelajaran mesin/ilmu data (ML), dan matematika tingkat lanjut.

Mari kita lihat bagaimana masing-masing gambar ini cocok dengan perincian data/model/infrastruktur proyek AI. Insinyur perangkat lunak sebagian besar bekerja di sisi infrastruktur proyek, menyiapkan sistem yang mengambil data (dari sumber internal atau eksternal) dan mengintegrasikan model ML ke dalam produk atau layanan Anda yang sudah ada.

Ada juga jenis figur khusus yang disebut sebagai insinyur data. Orang ini biasanya dipekerjakan ketika Anda berurusan dengan data dalam jumlah besar (berbagai terabyte), dan bahkan menghitung rata-rata sederhana dari kolom basis data pun bisa menjadi tantangan. Karena Anda mungkin akan menggunakan insinyur data untuk infrastruktur proyek Anda, tetapi bukan untuk model ML-nya, kami akan menganggap mereka sebagai jenis insinyur perangkat lunak khusus yang menangani kumpulan data yang sangat besar.

Insinyur pembelajaran mesin memiliki pengetahuan khusus tentang AI dan pembelajaran mesin. Mereka hafal isi bagian 1 buku ini (dan banyak lagi), dan pekerjaan mereka dimulai dengan membersihkan dan menganalisis data hingga memilih model yang paling tepat untuk proyek AI Anda. Mereka juga akan mengurus penulisan kode untuk melatih model dan mengevaluasi kinerjanya. Hasil kerja tim ML seringkali berupa model yang terlatih, beserta deskripsi akurasi dan data yang digunakan untuk pelatihan. Insinyur ML menangani peningkatan kinerja model, dan akan memandu upaya Anda dalam mengumpulkan lebih banyak atau data yang lebih baik. Meskipun beberapa insinyur ML juga tahu cara menerapkan model secara efektif, hal itu biasanya bukan aktivitas favorit mereka. Dari ketiga sosok ini, insinyur ML adalah yang paling mirip dengan deskripsi ilmuwan data kami sebelumnya.

Terakhir, peneliti adalah peran yang paling akademis dari ketiganya. Mereka dapat menggunakan pengalaman riset mendalam mereka di bidang tertentu untuk menemukan solusi bagi masalah-masalah baru yang belum banyak dieksplorasi di industri. Dari ketiga komponen proyek, peneliti berada di posisi yang kokoh di kubu model, di mana mereka dapat menggunakan pengetahuan mereka tentang keadaan terkini untuk membangun algoritma AI yang mendorong batas-batas kemungkinan. Kebanyakan peneliti hebat dalam membangun model, tetapi kurang memiliki keterampilan rekayasa perangkat lunak dan biasanya tidak tertarik untuk menerapkan kreasi mereka atau menemukan semua bug.

Perbedaan utama antara peneliti dan insinyur ML adalah peneliti adalah ilmuwan, sementara insinyur ML adalah insinyur. Perbedaanannya mungkin tampak sewenang-wenang, tetapi mereka memiliki keahlian yang sangat berbeda. Jika proyek Anda tentang masalah atau algoritma apa pun yang telah kami jelaskan dalam buku ini, Anda menginginkan seorang insinyur ML. Semua pekerjaan riset telah dilakukan untuk Anda, dan Anda hanya membutuhkan seseorang yang memahaminya dan dapat menerapkannya pada masalah spesifik Anda. Jika Anda menjelajah di luar ranah permasalahan dan sumber data standar

industri, Anda membutuhkan seorang peneliti yang memiliki pemahaman mendalam tentang teori dan dapat menghasilkan pendekatan inovatif untuk permasalahan baru.

Meskipun demikian, keahlian ketiga jenis orang ini jelas tumpang tindih dan bahkan mungkin dapat dipertukarkan dalam beberapa kasus. Anda dapat memandang ketiga sosok tersebut sebagai jenis koki yang berbeda. Jika Anda ingin menyantap hidangan dasar, mungkin koki jenis apa pun dapat melakukannya. Seiring bertambahnya level, koki mulai berspesialisasi dalam masakan atau disiplin ilmu tertentu, dan di restoran bintang tiga Michelin, Anda mungkin menemukan koki pastry, koki saus, koki ikan, dan sebagainya.

Mari kita lihat siapa yang mungkin telah membangun beberapa proyek yang kami rancang dalam buku ini. Prediktor harga adalah proyek ML yang cukup standar (ada juga banyak contoh daring), jadi Anda dapat menggunakan ilmuwan data atau insinyur pembelajaran mesin. Hal yang sama berlaku untuk prediksi churn dan proyek pengklasifikasi upselling yang telah kita bahas di bab 3. Bab 4 membahas data media, yang sedikit lebih rumit. Jika Anda memiliki banyak pengguna yang menggunakan platform Anda secara bersamaan, Anda mungkin memerlukan seorang insinyur perangkat lunak untuk memastikan infrastruktur Anda stabil dan dapat menahan beban ratusan gambar yang perlu diproses secara bersamaan.

Algoritmenya adalah pengklasifikasi sederhana, dan mungkin setiap insinyur ML yang terampil dapat melakukannya. Di bab 5, kita melihat data teks, dan kita melihat berbagai ide proyek dengan tingkat kesulitan yang berbeda. Untuk proyek yang paling sederhana seperti pengklasifikasi teks atau untuk analisis sentimen, Anda dapat melanjutkan dengan seorang insinyur ML, tetapi jika Anda ingin mendorong batas-batas keadaan seni dan mencoba membangun brokerbot yang sangat kuat, Anda mungkin harus menyewa seorang peneliti juga. Di bab 6, kita membahas tentang sistem rekomendasi.

Tabel 8.2: Insinyur perangkat lunak, insinyur ML, dan peneliti AI melakukan pekerjaan terbaik mereka ketika tugas tersebut sesuai dengan pengalaman dan latar belakang mereka.

Insinyur perangkat lunak	Insinyur ML	AI peneliti
Latar belakang biasa	Ilmu Komputer	Ilmu komputer, statistik
Statistika, matematika, fisika	Tugas	Menerapkan aplikasi, mengintegrasikannya dengan infrastruktur yang ada, membangun database
Membangun dan menguji algoritma ML, mengembangkan strategi data, memantau hasilnya	Mengembangkan algoritma baru, meneliti solusi baru	Kesesuaian proyek
Infrastruktur TI yang kompleks, membutuhkan kinerja yang sangat tinggi (misalnya, mobil self-driving, atau layanan dengan banyak lalu lintas seperti Facebook), kumpulan data yang besar (insinyur data)	Masalah ML terkait dengan masalah bisnis yang dapat dipecahkan dengan ML terkini, seperti kebanyakan contoh yang dibahas dalam buku ini	Kompleks-Ketika teknologi yang dibutuhkan untuk memecahkan masalah bisnis belum dikembangkan (misalnya, tugas NLP yang kompleks)

Model sederhana dapat dibangun oleh seorang insinyur ML, sedangkan jika Anda ingin membangun model yang sangat kompleks, Anda dapat mencari peneliti untuk bergabung dengan tim. Sistem rekomendasi Anda mungkin perlu sangat cepat dan stabil. Hal ini berlaku jika Anda memiliki banyak pengguna yang menggunakan layanan Anda secara bersamaan, dan kecepatan sangatlah penting. Contohnya adalah umpan Facebook, yang perlu merekomendasikan konten hampir seketika saat Anda dan jutaan orang menggulir. Dalam hal ini, insinyur perangkat lunak menjadi sangat penting, karena merekalah yang dapat membangun infrastruktur yang dapat menahan beban tersebut.

Untuk mendapatkan gambaran lengkap tentang peran-peran ini, lihat tabel 8.2, yang menyoroti jenis tugas yang paling sesuai untuk setiap kategori bakat. Seperti biasa, perlu diingat bahwa keahlian optimal tim akan berubah seiring perkembangan proyek, bahkan ketika berbagai komponen proyek diselesaikan. Faktor-faktor seperti ini akan mendorong keputusan Anda untuk mempekerjakan konsultan ahli atau merekrut anggota tim penuh waktu. Kami menyarankan untuk tidak membuat kesalahan umum dengan mempekerjakan peneliti AI yang hebat tanpa memiliki dasar rekayasa perangkat lunak yang solid, karena karyawan baru tersebut kemungkinan akan frustrasi dengan kurangnya infrastruktur yang memadai untuk memindahkan pekerjaan ke tahap produksi dan pada akhirnya akan pergi.

Strategi yang umum adalah memulai perekrutan untuk peran yang lebih umum terlebih dahulu, sehingga mereka mendapatkan pengalaman dengan konteks organisasi yang luas. Lebih mudah mengandalkan konsultan untuk pengetahuan khusus seperti pemodelan AI, karena keahlian mereka lebih mudah ditransfer ke berbagai organisasi (bagaimanapun juga, model klasifikasi gambar tetap sama, terlepas dari apakah Anda mencoba mengklasifikasikan mentimun atau pejalan kaki).

Terakhir, jangan remehkan pentingnya senioritas. Meskipun beberapa tugas seperti pembersihan data dapat dilakukan oleh figur junior, proyek AI dapat dengan cepat menjadi berantakan, dan tanpa pengawasan dari orang yang lebih senior, Anda berisiko berakhir dengan kode yang tidak dapat dipelihara. Menghemat uang sekarang dapat menyebabkan pemborosan uang dan waktu yang lebih banyak di kemudian hari.

Bab ini memperkenalkan unsur-unsur utama dari setiap proyek AI yang sukses: strategi pengumpulan data, infrastruktur perangkat lunak dan integrasi dengan pemrosesan yang ada, serta tim yang terampil. Kami juga telah memperingatkan tentang banyak jebakan yang mengganggu perusahaan yang masih kurang berpengalaman dengan AI. Bab selanjutnya memberikan resep untuk menggabungkan unsur-unsur ini menjadi sebuah proyek AI.

- ❖ Implementasi teknis proyek AI memiliki tiga komponen utama: data, model, dan talenta.
- ❖ Strategi data yang koheren mempertimbangkan kekuatan dan kelemahan relatif data kepemilikan dan eksternal.
- ❖ Selain kuantitas, perhatian utama lainnya dalam pengumpulan data dan label adalah jenis bias, cakupan, dan konsistensi.

- ❖ Tim AI yang baik terdiri dari orang-orang dengan latar belakang dan keahlian yang berbeda: insinyur perangkat lunak, insinyur pembelajaran mesin (atau ilmuwan data), dan peneliti.

BAB 9

STRATEGI IMPLEMENTASI AI

Bab ini membahas

- Kompromi dalam membangun algoritma AI, membeli dari penyedia, atau menggunakan sumber terbuka
- Mengembangkan Strategi AI Lean untuk meminimalkan risiko
- Memanfaatkan siklus AI yang baik
- Mempelajari pelajaran berharga dari kegagalan AI

Dua bab sebelumnya telah memberikan Anda semua fondasi yang dibutuhkan untuk mendefinisikan dan mengimplementasikan proyek AI. Sekarang saatnya untuk menyatukan semuanya. Dalam bab ini, Anda akan mempelajari cara memutuskan komponen mana yang akan dibangun sendiri atau dibeli dari pihak ketiga, dan cara mengelola fase implementasi. Anda juga akan memahami kendala paling umum yang dihadapi oleh pengembang AI yang belum berpengalaman. Di akhir bab ini, Anda akan merasa cukup percaya diri untuk memulai proyek Anda besok.

9.1 MEMBELI ATAU MEMBANGUN AI

Mengutip perkataan orang bijak, terkadang cara terbaik untuk menyelesaikan proyek dengan sukses adalah dengan tidak memulainya sejak awal. Sebagai seorang penggiat AI, salah

satu keputusan paling berdampak yang dapat Anda buat adalah mengandalkan produk dan layanan yang ditawarkan oleh penyedia AI dan menghindari membangun sebagian atau seluruh teknologi yang dibutuhkan untuk proyek AI. Keputusan ini berlaku untuk ketiga komponen proyek: data, model, dan infrastruktur.

Bagian ini akan mengajarkan Anda cara berpikir strategis tentang pertanyaan membangun atau membeli. Sebenarnya, kami yakin bahwa status pasar layanan AI saat ini mengharuskan kami untuk menambahkan opsi ketiga: meminjam. Saat kami menyajikan masing-masing dari ketiga opsi ini di bagian selanjutnya, konsep mendasar yang perlu diingat adalah risiko implementasi. Saat Anda memutuskan untuk membeli atau menyewa teknologi dan infrastruktur siap pakai, Anda melepaskan fleksibilitas dan hak kekayaan intelektual, tetapi jelas menghindari risiko yang terkait dengan investasi teknologi itu sendiri.

Tabel 9.1 merangkum perbedaan utama antara solusi Beli, Pinjam, dan Bangun. Anda dapat mengaksesnya kembali kapan pun dibutuhkan, dan kami akan menjelaskannya secara detail di tiga bagian berikutnya.

Tabel 9.1 Membandingkan strategi Beli, Pinjam, dan Bangun

	Membeli	Meminjam	Membangun
Butuh data	Tidak	Ya	Ya
Perlu pengembangan model	Tidak	Tidak	Ya (berpotensi menggunakan model sumber terbuka)
Butuh infrastruktur	Tidak	Tidak	Ya (berpotensi menggunakan infrastruktur yang dihosting)
Resiko	Risiko teknologi rendah, tetapi Anda terikat pada agenda penyedia	Sama seperti Beli untuk platform ML, risiko teknologi untuk sumber terbuka	Risiko teknologi tertinggi
Keahlian tim diperlukan	Pengembangan web	Pengembangan web + ilmu data kecil	Web + ML
Biaya di muka	Tidak ada	Pengumpulan data	Pengumpulan data dan pengembangan model
Biaya variabel	Kecil	Sedang	Diamortisasi dalam investasi infrastruktur / cloud

Opsi Beli: Solusi Siap Pakai

Opsi paling sederhana untuk menerapkan teknologi AI adalah menggunakan salah satu solusi siap pakai yang ditawarkan oleh penyedia teknologi seperti Google, Microsoft, atau Amazon. Solusi Beli ini juga disebut *Machine Learning as a Service* (MLaaS). Dalam layanan

MLaaS, perusahaan teknologi membangun algoritma ML serbaguna, yang menangani pengumpulan data, pelatihan, dan industrialisasi. Anda menggunakan teknologi mereka untuk menginterpretasikan data Anda dengan imbalan uang.

Seperti yang Anda lihat pada Gambar 9.1, Anda hanya perlu mengirimkan data yang ingin dianalisis, dan Anda akan mendapatkan prediksi dari model ML mereka. Secara teknis, Anda akan menggunakan antarmuka pemrograman aplikasi (API) untuk mengirimkan data Anda ke penyedia dan mendapatkan jawaban. Misalnya, hampir setiap raksasa teknologi (Microsoft, Google, IBM) menawarkan API visi komputer: Anda mengirimkan gambar kepada mereka, dan mereka akan mengirimkan kembali kontennya. Beberapa API visi komputer sangat canggih dan dapat memberi Anda informasi menarik seperti perkiraan suasana hati seseorang dalam gambar berdasarkan ekspresi wajahnya.

Salah satu proyek sampingan berbasis ML kami beberapa tahun yang lalu adalah layanan transkripsi suara untuk jurnalis dan mahasiswa yang menginginkan transkrip wawancara atau kuliah. Setelah kami mengimplementasikan aplikasi web dan mengonfirmasi permintaan untuk produk tersebut, kami dimanjakan dengan banyaknya pilihan dalam hal menemukan penyedia AI yang dapat menjalankan transkripsi yang sebenarnya. Google, Amazon, dan Microsoft semuanya menawarkan layanan transkripsi suara dalam lusinan bahasa, serta beberapa penyedia yang lebih kecil yang berfokus pada niche tertentu.

Sayangnya, ketika menggunakan solusi MLaaS, Anda berhadapan dengan kotak hitam pepatah: Anda tidak memiliki kendali atas bagaimana model dibangun, kinerjanya, dan infrastruktur yang melayaninya. Kabar baiknya adalah Anda hampir tidak perlu melakukan apa pun dalam kasus ini, karena penyedia telah melakukan semua kerja keras pengumpulan data, pelatihan model, dan pengaturan infrastruktur untuk melayaninya (gambar 9.2). Yang tersisa untuk dilakukan adalah mengintegrasikan solusi ke dalam produk Anda, sebuah tugas yang dapat diselesaikan oleh pengembang web mana pun. Ini memungkinkan Anda untuk berpotensi mendapatkan prototipe pertama dan cepat dari aplikasi AI Anda dalam hitungan jam, bukan minggu atau bulan.



Gambar 9.1 Penyedia tempat Anda membeli solusi siap pakai mengurus semua detail teknis: Anda hanya perlu mengirim permintaan dengan data Anda dan mendapatkan respons kembali.



Gambar 9.2 Saat membeli solusi, penyedia mengurus ketiga komponen (data, model, dan infrastruktur).

Untuk memberi Anda gambaran tentang jenis layanan yang ditawarkan sebagai solusi siap pakai, mari kita lihat (tidak lengkap) daftar harga Google saat artikel ini ditulis:

- Rp. 20.000 per menit untuk transkripsi ucapan menjadi teks
- Rp. 10.000 per dokumen untuk pemahaman bahasa alami: mengekstrak data terstruktur dari dokumen dan melakukan analisis sentimen
- Rp. 200.000 per juta karakter untuk penerjemahan
- Rp. 10.000 per gambar untuk mendeteksi logo atau wajah pada gambar

Ada dua poin utama di sini. Pertama-tama, semua solusi siap pakai difokuskan pada penyelesaian masalah umum. Hal ini masuk akal, karena memoles dan memasarkan produk-produk ini membutuhkan banyak investasi yang akan dikomitmenkan oleh penyedia hanya jika mereka yakin akan ada minat yang luas di pasar. Misalnya, Google tidak akan pernah menawarkan "AI pemilah mentimun sebagai layanan" karena permintaan pasar tidak akan pernah cukup tinggi untuk menutup biaya pengembangan. Sebaliknya, lebih masuk akal bagi mereka untuk menawarkan layanan dengan minat yang lebih luas yang dapat mengenali objek sehari-hari seperti kucing, anjing, atau mobil. Kedua, dari perspektif finansial, solusi Beli memiliki biaya tetap yang hampir nol dan biaya variabel yang lebih tinggi dibandingkan strategi lain. Bahkan, dengan solusi ini, tidak ada upaya dalam pengumpulan data, pembersihan data, rekayasa ML, dan penerapan. Yang perlu Anda lakukan hanyalah mengintegrasikan API: tugas yang hanya membutuhkan beberapa jam bagi pengembang web rata-rata. Seperti yang telah Anda lihat di daftar harga sebelumnya, semua biayanya dibayar per penggunaan.

Dari sudut pandang teknis, tingkat risiko yang terkait dengan solusi ini sangat rendah. Karena Anda menggunakan produk yang dibuat oleh perusahaan teknologi ternama, Anda dapat cukup yakin bahwa Anda akan mendapatkan kinerja yang cangguh. Tentu saja, hal ini hanya berlaku selama aplikasi Anda sesuai dengan tugas yang dirancang untuk layanan tersebut. Hati-hati: fakta bahwa Anda mengalihdayakan semua risiko teknis tidak berarti solusi Beli bebas risiko. Semua solusi Beli memiliki ancaman utama yang sama: karena Anda tidak memiliki teknologi yang mendasarinya, Anda terikat pada agenda orang lain.

Mereka mungkin memutuskan untuk menghentikan produk yang Anda gunakan dan membiarkan Anda berjuang sendiri. Kisah peringatan favorit kami melibatkan perusahaan rintisan Smyte yang berbasis di San Francisco. Perusahaan ini mengembangkan model ML dan platform untuk membantu komunitas daring melawan spam, penipuan, dan pelecehan dengan

secara proaktif mengidentifikasi tren berbahaya dalam komentar dan interaksi. Ketika Smyte diakuisisi oleh Twitter pada tahun 2018, pelanggan terbangun dan mendapati bahwa platform tersebut telah ditutup, karena Twitter memutuskan untuk fokus melayani kebutuhan internalnya saja. Pelanggan ternama seperti Meetup dan Zendesk terpaksa berebut mencari pengganti karena solusi siap pakai yang mereka andalkan tiba-tiba berhenti merespons pertanyaan mereka.

Risiko penyedia yang menutup atau mengubah produk mereka menciptakan ketergantungan pada layanan mereka. Efek lock-in juga meluas ke keamanan dan privasi, karena solusi siap pakai terbatas untuk dijalankan di dalam jaringan dan infrastruktur cloud penyedia, yang mungkin tidak dapat diterima dalam industri yang mengutamakan keamanan seperti perawatan kesehatan dan pertahanan. Namun, kecepatan luar biasa dan biaya tetap yang rendah dari solusi ini menjadikannya pilihan yang tepat untuk digunakan dalam prototipe cepat yang dapat memvalidasi permintaan pasar akan produk Anda.

Opsi Pinjam: Platform ML

Seperti yang telah Anda lihat, solusi siap pakai hampir menjamin kinerja mutakhir pada banyak tugas AI generik. Bagaimana dengan situasi di mana kebutuhan Anda lebih spesifik? Misalnya, bagaimana dengan tugas klasifikasi mentimun yang dijelaskan di bab 4? Seperti yang telah kami katakan, Anda tidak akan menemukan produk "klasifikasi kualitas mentimun" di platform AI Amazon, karena pasarnya akan sangat kecil.

Sebaliknya, Anda dapat memilih untuk menggunakan produk platform ML. Dibandingkan dengan solusi siap pakai, platform ML memungkinkan Anda mengunggah data pelatihan Anda sendiri untuk mengoptimalkan model siap pakai untuk kebutuhan spesifik Anda. Dalam kasus klasifikasi mentimun, Anda dapat memanfaatkan model klasifikasi gambar Google, dan menyempurnakannya dengan kumpulan data mentimun yang diberi label.

Dari sudut pandang AI, platform ML sering kali memanfaatkan kekuatan transfer pembelajaran untuk memungkinkan pelanggan seperti Anda memanfaatkan investasi besar penyedia dalam pemodelan sambil beradaptasi dengan kebutuhan khusus (seperti klasifikasi mentimun). Dari segi integrasi, platform ML sama seperti solusi Beli: model dihosting di cloud penyedia, dan Anda tinggal mengirimkan permintaan ke sana (gambar 9.3). Harga untuk platform ML dua hingga empat kali lebih tinggi daripada solusi siap pakai karena kompleksitas yang lebih tinggi. Namun, model penetapan harganya sama: Anda hanya membayar sedikit biaya untuk setiap permintaan, dan sejumlah biaya tertentu berdasarkan jumlah data khusus yang ingin Anda gunakan untuk pelatihan.



Gambar 9.3 Saat meminjam solusi, penyedia mengurus model dan infrastrukturnya, dan Anda menyediakan datanya.

Meskipun platform ML tidak memerlukan investasi awal, Anda perlu menganggarkan pengumpulan dan kemungkinan pelabelan data, sehingga Anda dapat memanfaatkan kemampuan kustomisasinya. Produk platform ML umumnya memecahkan kelas masalah yang sama luasnya dengan solusi siap pakai; misalnya, Anda dapat mengunggah basis data teks Anda sendiri untuk meningkatkan analisis sentimen, atau mengirimkan korpus multibahasa untuk meningkatkan akurasi terjemahan. Kami membuat kategori Pinjam untuk platform ML karena Anda berinvestasi di platform "pinjaman": Anda tetap tidak memiliki IP untuk model tersebut. Namun, Anda memiliki data pelatihan yang dalam banyak kasus merupakan aset Anda yang sebenarnya.

Meskipun demikian, efek penguncian untuk platform ML mungkin lebih kuat daripada solusi siap pakai. Solusi siap pakai sangat umum sehingga, bahkan jika penyedia Anda tutup, Anda kemungkinan akan menemukan penawaran serupa dari pesaing. Di sisi lain, interaksi antara rahasia di platform ML dan data pelatihan Anda bisa lebih sulit direplikasi.

Penyedia AI sering mengiklankan platform ML sebagai pilihan yang tepat untuk proyek yang menggunakan data bisnis terstruktur yang dijelaskan di bab 2 dan 3. Karena kebutuhan komputasi model ini seringkali sederhana, platform ML menawarkan produk ML otomatis yang mencoba beberapa model pada data Anda dan secara otomatis memilih yang terbaik. Perlu disebutkan bahwa, jika suatu masalah cukup sederhana untuk dipecahkan oleh alat otomatis ini, rata-rata insinyur ML kemungkinan akan mendapatkan kinerja yang serupa dengan upaya yang terbatas.

Selama membantu perusahaan dalam proyek mereka, kami menemukan bahwa untuk masalah yang berkaitan dengan data terstruktur (biasanya data pemasaran, penjualan, atau bisnis inti), upaya terbesar sejauh ini adalah membersihkan dan menyiapkan data.

Dalam satu kasus, kami perlu membangun algoritma yang bekerja dengan data pusat panggilan sebuah perusahaan farmasi. Data ini telah dikumpulkan selama 14 tahun, tetapi tidak ada yang berusaha untuk mendapatkan nilai darinya: itu adalah biaya hangus yang digunakan hanya untuk pelaporan. Ketika seorang manajer yang cerdas mengusulkan pembangunan algoritma AI, tantangan pertama yang kami hadapi adalah mencoba menelusuri 14 tahun sejarah untuk memahami logika di balik data, dan membersihkan outlier serta potensi kesalahan entri data. Tantangan ini belum cukup berat, selama bertahun-tahun perusahaan menggunakan dua penyedia CRM, sehingga terdapat ketidakkonsistenan dalam format dan bidang data yang dikumpulkan. Singkat cerita, menyiapkan data untuk digunakan membutuhkan waktu tiga bulan. Setelah data siap, membangun model yang berkinerja baik membutuhkan waktu sekitar dua minggu.

Menemukan model yang tepat dan menyempurnakannya bisa sangat cepat. Platform ML tidak akan menghindarkan Anda dari upaya awal menyiapkan data untuk digunakan dalam aplikasi ML, sehingga hanya memberikan sedikit penghematan dalam hal waktu dan biaya

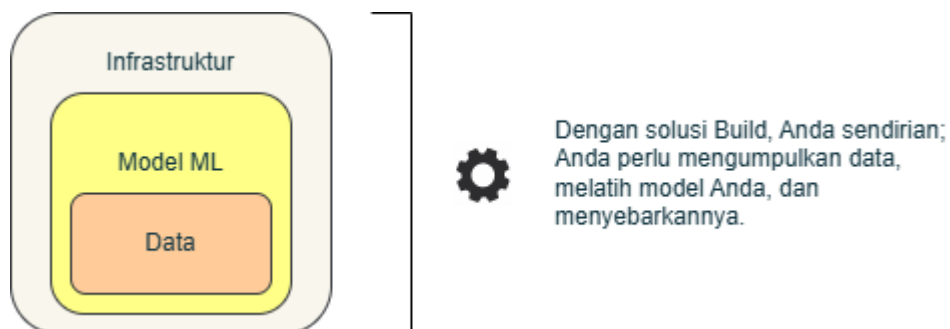
awal. Namun, platform ini menambahkan biaya variabel karena model bisnis mereka adalah bayar per penggunaan.

Opsi Bangun: Bekerja Keras

Jika solusi Beli maupun Pinjam tidak cocok untuk proyek Anda, Anda harus memilih opsi Bangun. Dengan membangun model Anda sendiri, Anda mendapatkan fleksibilitas dan kendali penuh atas teknologi Anda. Secara umum, membangun model Anda sendiri berarti Anda perlu mengurus ketiga komponen: data, model, dan infrastruktur (gambar 9.4).

Ini adalah dua situasi paling umum di mana mustahil menemukan solusi Beli atau Pinjam:

- ✚ Anda menggunakan data spesifik domain—Misalnya, jika Anda mengerjakan MRI 3D atau alat diagnostik medis lainnya, Anda akan bekerja dengan gambar hitam-putih tiga dimensi. Kemungkinan besar Anda tidak akan pernah menemukan produk siap pakai atau platform ML yang dapat mendukungnya.
- ✚ Anda memiliki kebutuhan yang sangat spesifik—Misalnya, Anda mungkin ingin membangun algoritma ML yang mengidentifikasi kanker dari sinar-X paru-paru dan menyoroti lokasi kanker tersebut. Bagian terakhir ini adalah tugas pelokalan objek, aplikasi yang relatif khusus yang hanya dicakup oleh sedikit atau tidak ada penyedia.



Gambar 9.4 Saat membangun solusi, tim Anda harus memperhatikan ketiga komponen: data, model, dan infrastruktur.

Dari sudut pandang risiko pengembangan, pilihan Build adalah yang paling berisiko, karena tim Anda bertanggung jawab penuh atas produk tersebut. Untungnya, mereka seringkali tidak sendirian, dan dapat mengandalkan dua akselerator kemajuan yang penting: model sumber terbuka dan infrastruktur yang dihosting. Mari kita lihat bagaimana masing-masing dapat membantu Anda membangun proyek.

Model sumber terbuka adalah perangkat lunak yang tersedia secara gratis (dan, seringkali, data pelatihan) yang memecahkan masalah tertentu. Selama lisensinya kompatibel dengan tujuan penggunaan Anda (misalnya, penggunaan komersial), teknisi Anda cukup menyalin kode, menyesuaikannya sesuai kebutuhan, dan menggunakannya dalam produk Anda. Jika Anda tidak familiar dengan dunia sumber terbuka, Anda mungkin bertanya-tanya mengapa orang menawarkan karya mereka secara gratis. Nah, dua kontributor terbesar model sumber terbuka adalah perusahaan teknologi besar dan akademisi.

Penyedia teknologi seperti Google dan Microsoft merilis produk-produk mutakhir untuk memperkuat reputasi mereka sebagai pemimpin AI dan menarik orang-orang ke produk siap pakai dan platform ML mereka. Para peneliti di dunia akademis menerbitkan perangkat lunak mereka untuk meningkatkan kesadaran akan karya mereka, menarik pendanaan dan pengakuan. Untuk kasus yang lebih spesifik, Anda mungkin tidak menemukan kode sumber terbuka sama sekali. Pilihan terbaik Anda adalah meminta peneliti Anda untuk mencari makalah ilmiah yang relevan dari komunitas akademis. Seringkali, deskripsi model yang disertakan dalam makalah sudah cukup bagi teknisi ML Anda untuk mereplikasinya dengan mudah, lalu Anda dapat mulai mengumpulkan data pelatihan.

Kode sumber terbuka biasanya merupakan cara yang bagus untuk memulai upaya pengembangan Anda. Tim Anda akan dapat memulai dari solusi yang sudah berfungsi dan mengulanginya untuk beradaptasi dengan masalah spesifik Anda. Dibandingkan dengan platform ML, kemampuan untuk menyempurnakan model memungkinkan teknisi ML di tim Anda untuk men-debug masalah dan bereksperimen dengan perubahan guna meningkatkan akurasi.

Secara umum, jika Anda membangun model sendiri, Anda juga perlu menyediakan dan memelihara infrastrukturnya, seringkali dengan mengelola sumber daya cloud Anda sendiri. Infrastruktur yang dihosting membebaskan teknisi perangkat lunak Anda dari beban ini dengan memungkinkan Anda menjalankan model kustom Anda di cloud penyedia. Selama model Anda sesuai dengan persyaratan teknis penyedia (misalnya, dalam hal bahasa pemrograman), Anda dapat menggunakan perangkat keras dan jaringan mereka untuk menjalankan model Anda. Hal ini berhasil karena, seperti yang telah kami jelaskan di bab 8, sebagian besar keajaiban dalam membangun AI terletak pada ketersediaan data yang tepat dan pembuatan model yang baik. Perangkat lunak untuk mengelola eksekusi telah distandarisasi dengan cepat dan dengan demikian menjadi komoditas dalam beberapa tahun terakhir.

Kode sumber terbuka dan infrastruktur yang dihosting dapat sangat membantu mengurangi investasi yang dibutuhkan untuk membangun model Anda sendiri, sehingga menghemat waktu kerja tim Anda selama bertahun-tahun. Yang terpenting, keduanya juga dapat meningkatkan kepercayaan diri Anda. Di awal proyek, mengetahui bahwa orang lain mampu memecahkan masalah yang sama akan menghilangkan banyak keraguan tentang kelayakannya. Menjelang akhir proyek, memiliki opsi untuk mendelegasikan sebagian besar pekerjaan penerapan ke vendor membuat tim berlari lebih cepat menuju garis akhir. Pada akhirnya, skenario mana (membangun, meminjam, atau membeli) yang Anda pilih bergantung pada spesifikasi setiap proyek AI yang akan Anda kerjakan. Bagian selanjutnya akan memandu Anda dalam mengambil keputusan ini.

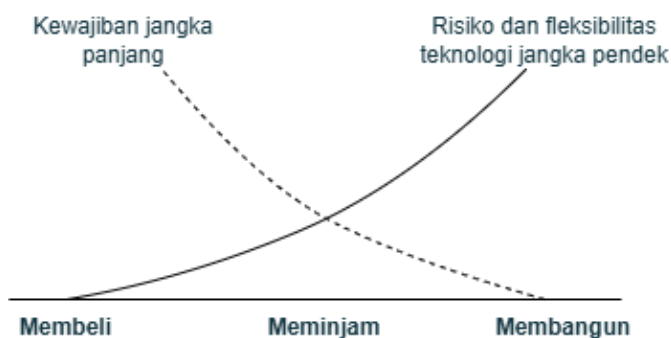
9.2 MENGGUNAKAN STRATEGI LEAN

Bagaimana Anda bisa memilih strategi Beli, Pinjam, dan Bangun yang paling sesuai untuk tim dan proyek yang sedang Anda kerjakan? Mungkinkah menemukan satu strategi yang sesuai untuk setiap tahap proyek? Bagian ini memperkenalkan jawaban kami untuk

pertanyaan ini: Strategi AI Lean, sebuah kerangka kerja yang kami kembangkan untuk membantu Anda membangun produk AI secara bertahap.

Kata lean berasal dari lean manufacturing, sebuah filosofi yang lahir di industri manufaktur Jepang (terutama di perusahaan otomotif seperti Toyota) yang bertujuan meminimalkan pemborosan produksi tanpa mengorbankan efisiensi. Eric Ries mengadaptasi gagasan ini ke dunia kewirausahaan dalam bukunya yang monumental, *The Lean Startup*. Ries memperkenalkan sebuah proses untuk mengelola inovasi yang mendorong pengujian asumsi pasar dan teknologi melalui eksperimen berkelanjutan. Wawasan inti di balik strategi ini adalah bahwa bahkan CEO dan pendiri pun hanya memiliki sedikit informasi tentang apa yang benar-benar dihargai pasar, sehingga mereka sebaiknya segera membangun iterasi pertama produk, melihat apakah pelanggan menyukainya, lalu memperbaikinya berdasarkan masukan mereka. Strategi Lean AI memiliki tujuan yang sama: sebuah proses untuk mengimplementasikan proyek AI secara bertahap sambil meminimalkan risiko dan pemborosan sumber daya (waktu, uang, dan moral tim). Dengan mengikuti Strategi Lean AI, Anda akan belajar memilih rencana implementasi terbaik untuk proyek tertentu (membangun, membeli, atau meminjam).

Saat Anda memulai proyek, langkah pertama adalah memahami risiko teknologi yang terkait dengan setiap jalur pengembangan AI. Setiap opsi memiliki kompleksitas terkait yang disertai risiko jangka pendek, sekaligus memberi Anda tingkat fleksibilitas yang terkait dengan liabilitas jangka panjang. Hubungan antara risiko dan fleksibilitas ini direpresentasikan pada Gambar 9.5.



Gambar 9.5 Poin-poin pada spektrum Bangun versus Beli saling bertumpuk dalam hal risiko dan fleksibilitas.

Sekarang, bayangkan Anda baru saja memulai implementasi proyek AI. Inilah momen di mana ketidakpastian berada pada titik tertinggi, terutama karena tiga faktor:

- *Risiko bisnis*—Apakah solusi saya memiliki nilai bisnis?
- *Risiko alur kerja*—Dapatkah saya berhasil mengintegrasikan solusi saya ke dalam bisnis yang ada?
- *Risiko teknologi*—Apakah teknologinya akan berfungsi dengan cukup baik?

Metodologi yang diperkenalkan di bab 7 dimaksudkan untuk meminimalkan risiko bisnis dan alur kerja, tetapi tidak ada yang sebanding dengan pengujian di dunia nyata dengan orang sungguhan. Untuk menghadirkan produk baru Anda kepada pengguna sungguhan, Anda seringkali perlu membangun setidaknya beberapa teknologi. Lean Startup memperkenalkan konsep produk minimum yang layak (MVP): produk terkecil yang mungkin dapat Anda bangun

sambil tetap memberikan nilai kepada pelanggan. Dalam dunia proyek AI, ada baiknya untuk berpikir dalam kerangka AI minimum yang layak (MVAI): teknologi terkecil dan paling tidak berisiko yang mungkin dapat mendukung MVP pertama Anda.

Memulai dari Solusi Pembelian

Peran terpenting Strategi Lean AI adalah mengarahkan Anda ke jalur yang tepat untuk membangun MVAI. Mengingat tingginya tingkat ketidakpastian, sebaiknya fokus pada strategi jangka pendek dan tunda keputusan jangka panjang setelah kelayakan ide tersebut divalidasi oleh pelanggan. Jika Anda melihat Gambar 9.6, ini berarti memulai dari Solusi Pembelian.



Gambar 9.6 Langkah pertama dalam membangun Strategi AI Lean untuk proyek AI baru adalah mulai mencari solusi Beli, dan ikuti pohon keputusan ini.

Kosakata teknis yang kami ajarkan di bagian 1 sudah cukup untuk mulai mencari solusi Beli di Google. Misalnya, Anda menjalankan startup pengiriman makanan dan ingin mengurutkan ulasan makanan menjadi positif dan negatif. Seperti yang Anda ingat dari bab 5, aplikasi ini disebut analisis sentimen. Query yang bagus untuk mencari solusi Beli adalah "API Analisis Sentimen", yang akan menghasilkan banyak hasil. Dalam banyak kasus, Anda akan menemukan produk siap pakai dari penyedia teknologi besar (seperti Google, Amazon, atau Microsoft) dan perusahaan spesialis yang lebih kecil. Di dunia AI, Anda benar-benar harus mempertimbangkan keduanya. Pemain besar memiliki keuntungan skala, tetapi pemain kecil mungkin memiliki keunggulan karena pengumpulan data dan upaya rekayasa mereka mungkin lebih selaras dengan tujuan spesifik proyek Anda.

Jika Anda menemukan apa yang Anda cari di antara penawaran penyedia AI yang terus berkembang, Anda beruntung. Mengadopsi salah satu dari ini kemungkinan besar merupakan pilihan terbaik Anda untuk mengembangkan iterasi awal proyek AI Anda.

Dalam kasus lain, Anda mungkin tidak menemukan solusi untuk tugas AI yang tepat yang dibutuhkan proyek Anda. Hal ini tidak jarang terjadi; lagipula, penyedia menginvestasikan sumber daya untuk hanya membangun produk-produk yang memiliki potensi pasar yang besar. Jika ini yang terjadi pada proyek Anda, Anda memiliki dua pilihan:

- Membatalkan ulang proyek AI Anda agar sesuai dengan penawaran Beli yang ada.
- Naiki tangga risiko teknologi dan cari solusi Pinjam.

Anda sebaiknya mulai dengan mempertimbangkan opsi 1 terlebih dahulu. Adakah cara untuk merumuskan ulang tugas AI Anda sehingga Anda dapat menggunakan apa yang ada di pasaran tanpa mengorbankan proposisi nilai Anda? Misalnya, mari kita asumsikan Anda adalah perusahaan telekomunikasi dan ingin menguji layanan dukungan pelanggan super cepat: pengguna menulis pesan, dan AI secara otomatis mengarahkan pelanggan ke agen pusat panggilan terbaik; Misalnya, "Internet saya tidak berfungsi" akan menjadi masalah teknis, dan "Saya ingin mengaktifkan penawaran baru" akan menjadi klaim komersial. Selain perusahaan AI khusus dengan produk lengkap untuk proyek spesifik Anda, kecil kemungkinan Anda akan menemukan produk siap pakai untuk tugas ini.

Di sisi lain, Anda akan menemukan beberapa algoritma analisis sentimen yang sangat akurat di pasaran saat ini. Meskipun algoritma ini tidak akan menyelesaikan 100% masalah, bagaimana jika menggunakannya untuk mengarahkan pelanggan yang marah langsung ke dukungan pelanggan, dan pelanggan netral ke penyortiran manual oleh operator manusia? Tentu saja, ini bukan solusi akhir, tetapi memungkinkan Anda menguji asumsi dasar:

- *Asumsi bisnis*—Apakah pengguna bersedia berinteraksi dengan dukungan pelanggan melalui sistem otomatis, atau apakah mereka mengabaikannya sepenuhnya?
- *Asumsi alur kerja*—Apakah sistem baru menghemat waktu bagi operator?

Jika Anda tidak dapat menemukan cara untuk mengurangi masalah Anda seperti yang kami lakukan dalam contoh ini, saatnya mempertimbangkan untuk beralih ke solusi Pinjam. Anda juga bisa mempertimbangkan untuk beralih ke solusi Pinjam ketika Anda sudah mencoba solusi Beli dan performanya tidak cukup baik untuk proyek tersebut. Namun, pertama-tama, Anda harus mempertimbangkan kembali arti "cukup baik": meskipun AI-nya tidak sempurna dan membuat kesalahan yang jelas dari sudut pandang manusia, proyek tersebut mungkin tetap sukses bagi organisasi. Meskipun bukan itu masalahnya, luangkan waktu sejenak dan beri tepuk tangan untuk diri sendiri. Berkat Strategi AI Lean dan fokus pada model siap pakai, Anda telah menemukan informasi teknis penting tentang proyek Anda dengan sedikit atau tanpa investasi awal dalam teknologi.

Selanjutnya, libatkan teknisi ML di tim Anda untuk mencoba mencari tahu mengapa tepatnya produk Beli tidak berhasil. Perlu diingat bahwa produk Beli adalah hasil dari investasi rekayasa yang substansial, sehingga tim Anda dapat berharap untuk bekerja lebih baik hanya dalam satu keadaan: kurangnya data pelatihan khusus masalah (misalnya, klasifikasi mentimun, atau kosakata khusus dalam tugas NLP). Misalnya, jika Anda mendapati model Beli memiliki performa yang buruk dalam menerjemahkan dokumen dari bahasa Inggris ke bahasa Prancis, kecil kemungkinan Anda dapat mengungguli Google dan meningkatkannya dengan menambahkan lebih banyak data pelatihan. Jika Anda merasa bahwa kurangnya data pelatihan khusus tugas menghambat proyek Anda, model Pinjam dapat menjadi solusi.

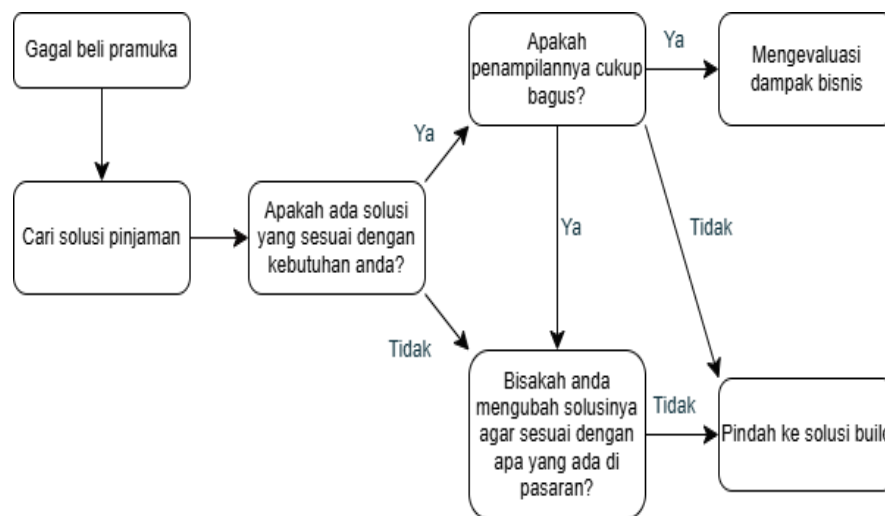
Beralih ke Solusi Pinjam

Ingatlah bahwa kekuatan model Pinjam berasal dari kemampuan Anda untuk melengkapi model yang ada dari penyedia AI dengan data pelatihan Anda sendiri. Kami menyebut produk yang dapat disesuaikan ini sebagai platform ML. Jika Anda berhasil mengumpulkan (dan mungkin memberi label) data yang sesuai untuk proyek Anda, Anda akan

mendapatkan yang terbaik dari kedua dunia: memanfaatkan investasi bertahun-tahun dalam model siap pakai sekaligus mencapai kinerja yang baik pada tugas-tugas spesifik Anda. Sebagai bonus, model Pinjam membebaskan Anda dari keharusan merancang, menerapkan, dan memelihara infrastruktur untuk proyek Anda. Dengan menggunakan solusi Pinjam, Anda juga akan mulai mengembangkan keahlian AI internal yang mendorong upaya pengumpulan data Anda, dan membangun pengetahuan institusional tentang membangun, membandingkan, dan mengevaluasi kinerja berbagai model AI.

Dengan kata lain, beralih dari Beli ke Pinjam memungkinkan Anda untuk mengulangi komponen AI proyek sambil mendapatkan jawaban dengan cepat, karena platform ML pada dasarnya merupakan pengganti langsung untuk model siap pakai. Untuk menemukan solusi Pinjam yang sesuai, mulailah dari pencarian solusi Beli yang telah Anda lakukan dan lanjutkan dari sana. Perlu diingat bahwa opsi untuk platform ML mungkin lebih terbatas daripada model siap pakai, karena menambahkan titik kustomisasi untuk transfer pembelajaran akan menambah beban kerja bagi penyedia AI.

Jika Anda bahkan tidak dapat menemukan solusi Pinjam untuk proyek AI Anda, Anda harus membuat keputusan yang sama seperti saat mengevaluasi opsi Beli: Bisakah Anda membingkai ulang solusi Anda agar sesuai dengan yang ada di pasaran? Jika tidak, satu-satunya pilihan Anda adalah kembali menaiki tangga risiko teknologi dan mulai membangun model Anda sendiri, seperti yang dapat Anda lihat pada Gambar 9.7.



Gambar 9.7 Pohon keputusan yang sama yang kami rancang untuk solusi Beli juga dapat digunakan untuk solusi Pinjam, dengan titik awal dan akhir yang berbeda (kami mulai dari pencarian Beli yang gagal dan berakhir dengan beralih ke solusi Beli).

Melakukan Semuanya Sendiri: Membangun Solusi

Jika Anda mencapai tahap Strategi AI Lean ini, tidak banyak lagi yang perlu diputuskan. Proyek Anda tidak sesuai dengan produk Beli atau Pinjam apa pun di pasaran, dan Anda tidak dapat membingkai ulang untuk melakukannya. Ini berarti anggota tim Anda harus bekerja keras dan mulai membangun model sendiri (kemungkinan besar memanfaatkan komunitas sumber terbuka dan infrastruktur yang dihosting seperti yang dijelaskan di bab 8).

Karena Beli adalah "ujung jalan" untuk Strategi AI Lean, kami ingin mengomentari setiap jalur yang mungkin membawa proyek Anda ke sini. Situasi paling umum terjadi ketika data yang digunakan proyek Anda tidak umum atau sangat spesifik. Bayangkan sebuah aplikasi audio yang berencana mengklasifikasikan spesies burung berdasarkan kicauannya. Tidak ada produk platform ML yang mungkin mendukung pembelajaran transfer dari suara hewan, dan tidak ada solusi siap pakai yang akan menawarkan hal itu.

Dalam kasus lain, kami telah melihat tim dengan solusi Borrow yang berfungsi ingin mencoba membangun model mereka sendiri, mengharapkan pengurangan biaya atau peningkatan kinerja. Meskipun benar bahwa membangun model sendiri (dan mengelola semua aspek operasinya) mungkin memiliki biaya variabel yang lebih rendah, banyak tim membuat kesalahan dengan mengabaikan investasi awal yang substansial yang telah mereka nikmati untuk menciptakan solusi Borrow.

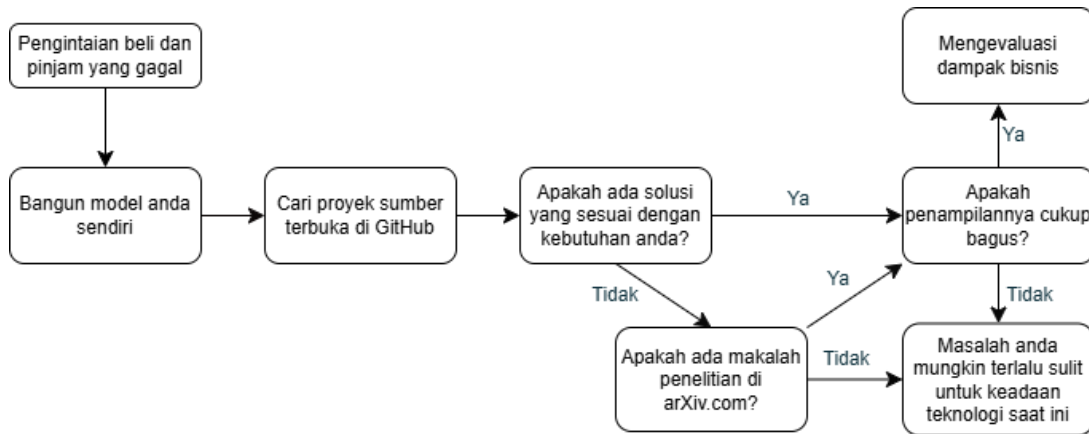
Manfaat kinerja yang dirasakan seringkali juga ilusif: setiap model Borrow di pasaran telah melalui pengujian ekstensif untuk memverifikasi keakuratannya dalam berbagai skenario. Sangat mudah untuk terlalu bersemangat dengan hasil awal yang positif dan meremehkan jumlah upaya rekayasa yang diperlukan untuk menyempurnakan solusi Build. Cara terbaik untuk mendekati migrasi ke solusi Build adalah dengan memperlakukannya seperti proyek sampingan, dan tidak mempertaruhkan seluruh strategi AI padanya. Ini akan membantu mengurangi risiko migrasi melalui pemantauan konstan terhadap upaya internal dibandingkan dengan solusi Borrow.

Selain alasan teknis ini, terkadang masalah organisasi atau bisnis menjadikan Build satu-satunya pilihan. Misalnya, baik model siap pakai maupun platform ML umumnya berjalan di dalam sumber daya cloud penyedia AI. Faktor ini menjadikannya pilihan yang tidak disarankan bagi industri yang mengutamakan privasi atau keamanan seperti kesehatan atau pertahanan, karena data yang berpotensi sensitif harus dibagikan dengan penyedia agar sistem dapat berfungsi. Organisasi lain mungkin memutuskan untuk berinvestasi dalam pembuatan IP pihak pertama meskipun pilihan ini bukan cara termurah atau tercepat untuk mengimplementasikan proyek tertentu.

Kami tidak dapat membantah sudut pandang strategis seperti ini, tetapi ingatlah bahwa Anda akan memiliki banyak waktu untuk mengembangkan IP Anda sendiri setelah Anda membuat prototipe yang berfungsi. Meskipun dalam jangka panjang Anda ingin mengembangkan model Anda sendiri, kami tetap menganjurkan Anda untuk memulai dengan mengikuti Strategi AI Lean dan memulai upaya Anda dengan solusi Beli. Ini mungkin hanya batu loncatan, tetapi tetap akan sangat berharga untuk memiliki prototipe yang berfungsi cepat yang dapat Anda gunakan untuk menguji asumsi bisnis Anda yang paling mendesak. Sebagai peringatan terakhir, kami memperingatkan Anda agar tidak terlalu jauh mendalami riset R&D. Setelah Anda memutuskan untuk membangun model Anda sendiri, mudah untuk menjadi berlebihan dan menjadi semakin ambisius.

Bereksperimen dengan model AI mutakhir memang mengasyikkan dan bermanfaat bagi kemanusiaan. Sebagai bonus tambahan, tim Anda akan sangat bersemangat dan mengerahkan 110%. Namun, kemungkinan besar, jika Anda membaca buku ini, Anda dan

organisasi Anda belum siap untuk memulai proyek R&D multi-tahun untuk memvalidasi penggunaan AI yang baru atau algoritma yang benar-benar baru. Pertimbangkan untuk menunda proyek-proyek "moon-shot" ini dan melanjutkannya setelah budaya AI di organisasi Anda lebih matang.



Gambar 9.8 Pohon keputusan Strategi AI Lean untuk solusi Build

Pohon keputusan untuk opsi Build direpresentasikan pada Gambar 9.8.

Kesimpulannya, Strategi AI Lean mendorong Anda untuk meminimalkan risiko implementasi dengan menunda komitmen terhadap tugas-tugas rekayasa yang mahal dan berisiko. Sebaliknya, strategi ini mendorong Anda untuk memulai proyek dengan menggunakan solusi Beli atau Pinjam yang memungkinkan Anda memanfaatkan skala ekonomi dari penyedia besar. Seiring dengan kematangan proyek dan organisasi Anda, Anda akan memiliki banyak peluang untuk memulai rencana implementasi yang lebih ambisius secara teknis.

Saat Anda beralih dari solusi Beli, Pinjam, dan Build setelah Strategi AI Lean, Anda mungkin kesulitan memutuskan kapan harus memulai dan menerapkan proyek. Bab 2 membahas aspek teknis ini, dan memperkenalkan pengukuran penting dari positif palsu dan negatif palsu. Bagian selanjutnya melengkapi diskusi ini dengan menceritakan kisah yang lebih mendalam tentang bagaimana kinerja berkembang melalui siklus hidup proyek AI pada umumnya.

9.3 MEMAHAMI SIKLUS BAIK AI

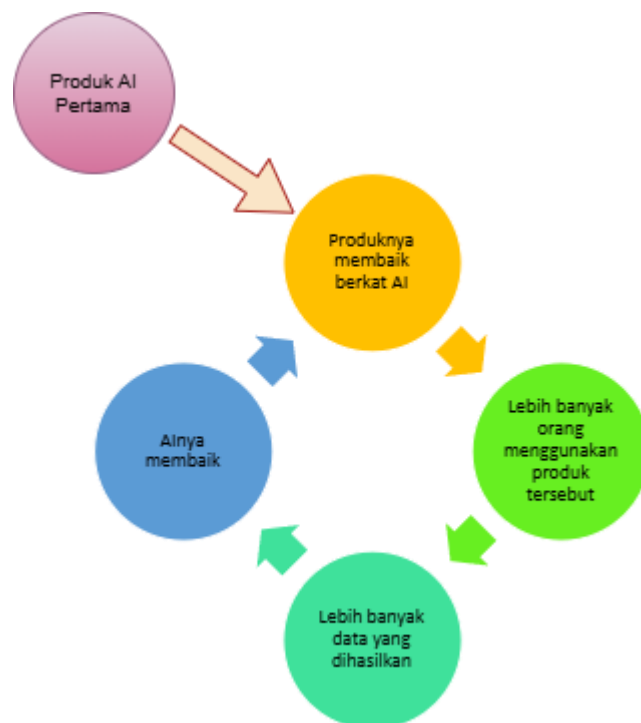
Mari kita kembali ke awal buku ini, ketika kami pertama kali menempatkan diri sebagai pengelola situs web real estat. Kami mengerjakan sebuah platform tempat pemilik rumah dapat mendaftarkan properti mereka untuk dijual, dan calon pembeli dapat melihat penawaran, menjadwalkan kunjungan, dan akhirnya membeli salah satu rumah. Kami membayangkan membangun fitur baru yang memungkinkan model AI untuk secara otomatis dan instan memprediksi harga jual terbaik untuk sebuah rumah.

Asumsi kami adalah pengguna akan menyukai fitur ini, dan ini akan memberi kami keunggulan kompetitif dibandingkan platform real estat lain di pasaran. Jika kami benar, orang-

orang akan mulai menggunakan situs web kami lebih sering daripada pesaing kami, yang mendorong peningkatan pendapatan, lebih banyak pengguna, dan... lebih banyak data.

Apa yang terjadi pada algoritma ML ketika Anda memiliki lebih banyak data? Biasanya, kinerjanya meningkat. Produk Anda bisa menjadi lebih baik lagi, dan prediksi Anda lebih akurat, membuat pengguna lebih puas dan dengan demikian menarik lebih banyak pengguna. Hasilnya, sekali lagi, adalah lebih banyak data.

Apa yang baru saja kami jelaskan adalah siklus kebajikan AI: jika AI meningkatkan produk Anda, mendatangkan lebih banyak pengguna dan pemanfaatan, Anda akan menghasilkan lebih banyak data yang dapat Anda gunakan sebagai umpan balik ke dalam model, yang selanjutnya akan meningkatkannya. Ini adalah siklus yang saling memperkuat dan, berpotensi, tidak pernah berakhir. Anda dapat memvisualisasikan siklus ini pada Gambar 9.9.



Gambar 9.9 Siklus kebaikan AI. Seiring AI meningkatkan produk Anda, semakin banyak pengguna yang ingin menggunakannya. Hal ini menghasilkan lebih banyak data, yang meningkatkan AI Anda, yang selanjutnya meningkatkan produk Anda, dan mendatangkan lebih banyak pengguna.

Mari kita lihat bagaimana siklus AI yang baik memengaruhi produk di dunia nyata, menggunakan contoh platform real estat kami. Sebelum kami memulai proyek AI pertama, yang kami miliki hanyalah situs web standar, mungkin serupa dengan penawaran dari pesaing di bidang yang sama. Produk AI pertama yang kami putuskan untuk dibangun adalah prediktor harga rumah: sebuah algoritma pembelajaran mesin (ML) yang akan belajar dari penjualan properti sebelumnya dan secara otomatis menghitung harga yang paling mungkin untuk setiap properti di pasar.

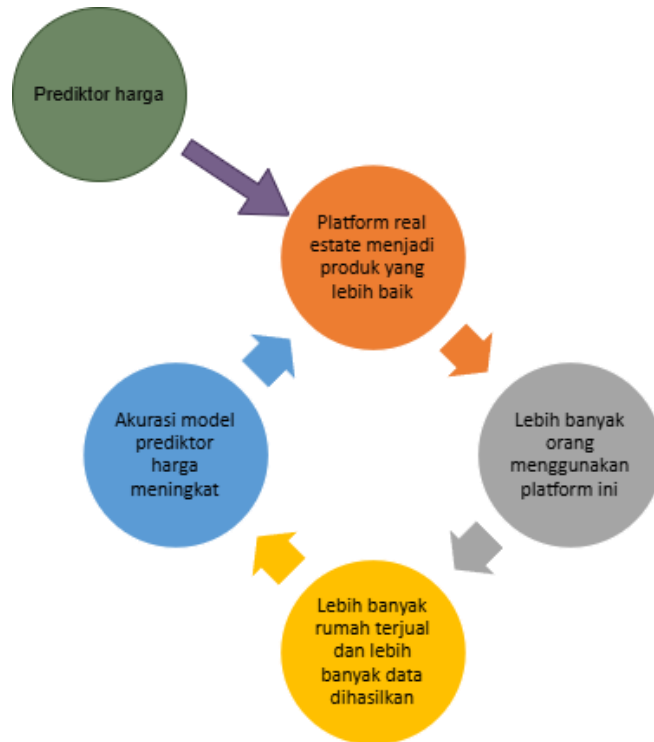
Fitur baru dan unik ini menghasilkan perbincangan di media dan menciptakan keunggulan baru bagi platform kami yang membedakan kami dari pesaing. Wajar saja, kami membayangkan bahwa pelanggan baru akan mulai menggunakan platform kami, tertarik dengan prediktor harga rumah baru kami. Pelanggan baru tidak hanya mendatangkan pendapatan tambahan, tetapi juga data baru.

Ketika pelanggan baru ini mulai menjual rumah mereka, data tambahan yang kami kumpulkan dapat digunakan untuk melatih ulang algoritma kami dengan lebih banyak data, sehingga meningkatkan kinerjanya. Secara wajar, algoritma yang lebih baik dapat menghasilkan lebih banyak data, yang akan menghasilkan lebih banyak pelanggan, dan seterusnya, seperti yang Anda lihat pada Gambar 9.10. Siklus AI yang baik dapat berlanjut selamanya, atau hingga Anda mulai merasakan hasil yang semakin berkurang dari algoritma Anda dan mencapai titik jenuh teknologi.

Anda mungkin bertanya, Bagaimana jika saya tidak memiliki banyak data, dan model awal saya tidak dapat memenuhi standar kualitas produk saya saat ini? Kami berpendapat bahwa dalam situasi ini, siklus AI yang baik sangatlah penting, dan Anda harus mengirimkan apa pun yang memberikan nilai bagi pelanggan Anda, meskipun jauh dari sempurna. Untuk memahami alasannya, mari kita bahas beberapa konsep dasar tentang bagaimana inovasi teknologi diadopsi dan dipasarkan. Pada tahun 1962, profesor studi komunikasi Everett Rogers menerbitkan *Diffusion of Innovations*, di mana ia menyajikan model sosiologis yang menggambarkan adopsi atau penerimaan produk atau inovasi baru. Wawasan utamanya adalah bahwa pasar tidak terdiri dari satu kelompok orang yang homogen; melainkan terbagi menjadi kelompok-kelompok individu yang memiliki karakteristik serupa:

- ❖ *Inovator*—Pengambil risiko ini memiliki status sosial yang tinggi, berjiwa sosial, dan memiliki likuiditas keuangan. Mereka adalah orang-orang yang mengantre di depan toko Apple sehari sebelum peluncuran iPhone baru.
- ❖ *Pengadopsi awal*—Mereka juga pengambil risiko, berjiwa sosial, dan memiliki daya beli yang tinggi. Pilihan mereka lebih konservatif daripada inovator, jadi mereka mungkin tidak mengantre sehari sebelum peluncuran iPhone baru, tetapi Anda dapat yakin bahwa mereka akan membelinya keesokan harinya.
- ❖ *Mayoritas awal*—Ini adalah kelompok besar orang dengan status sosial dan daya beli yang beragam, yang mengadopsi suatu inovasi setelah dirilis selama beberapa waktu. Mereka menghargai peningkatan yang dihadirkan oleh kemajuan teknologi dalam kehidupan mereka tetapi tidak ingin mengambil risiko. Mereka adalah orang-orang yang menunggu berbulan-bulan setelah peluncuran ponsel baru untuk melihat pendapat orang lain, atau hampir tidak dapat membedakannya dari model lama.
- ❖ *Mayoritas akhir*—Kelompok besar orang lainnya, umumnya skeptis dan menghindari risiko. Mereka lebih mengutamakan keamanan daripada inovasi, dan tidak ingin beralih ke teknologi baru sampai mereka benar-benar yakin dengan keuntungan yang akan mereka dapatkan. Mereka adalah orang-orang dengan ponsel pintar yang tertinggal dua hingga tiga generasi.

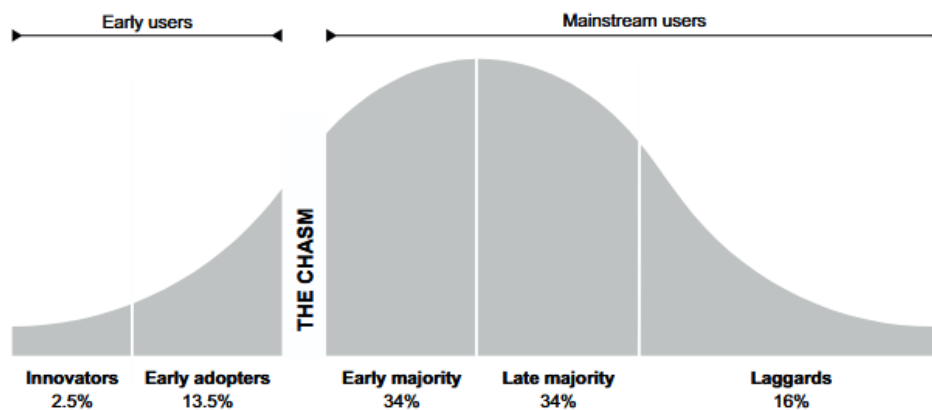
- ❖ *Lambat*—Mereka yang terakhir mengadopsi suatu inovasi. Mereka adalah orang-orang yang masih menggunakan ponsel keyboard mekanis, dan menggantinya dengan model termurah yang dijual ketika rusak.



Gambar 9.10 Siklus kebajikan AI yang diterapkan pada contoh real estat

Pada tahun 1991, Geoffrey A. Moore menerbitkan *Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customers*, yang dibangun di atas teori sebelumnya dan menambahkan wawasan sederhana namun penting: ketika memperkenalkan inovasi teknologi, beralih dari pengadopsi awal ke mayoritas awal bukanlah hal yang mudah, dan banyak inovasi gagal melakukannya. Moore memperkenalkan konsep jurang: periode ketidakpastian yang sulit ketika sebuah perusahaan telah mencapai semua inovator dan pengadopsi awal di pasarnya dan perlu mencapai mayoritas awal, seperti yang Anda lihat pada gambar 9.11.

Jurang ini ada karena pasar awal (inovator dan pengadopsi awal) dan pasar akhir (mayoritas awal, mayoritas akhir, dan lamban) didorong oleh tujuan yang sama sekali berbeda. Sementara yang pertama menghargai inovasi dan nilai yang dibawanya, yang terakhir menghargai keamanan dan menghindari risiko, dan sulit untuk membuat mereka melihat nilai produk inovatif Anda.



Gambar 9.11 Pasar untuk inovasi teknologi dibagi menjadi lima kelompok pelanggan. Setelah para pengguna awal, kita mencapai jurang pemisah: momen ketidakpastian yang tinggi ketika beralih dari pengguna awal ke pengguna umum.

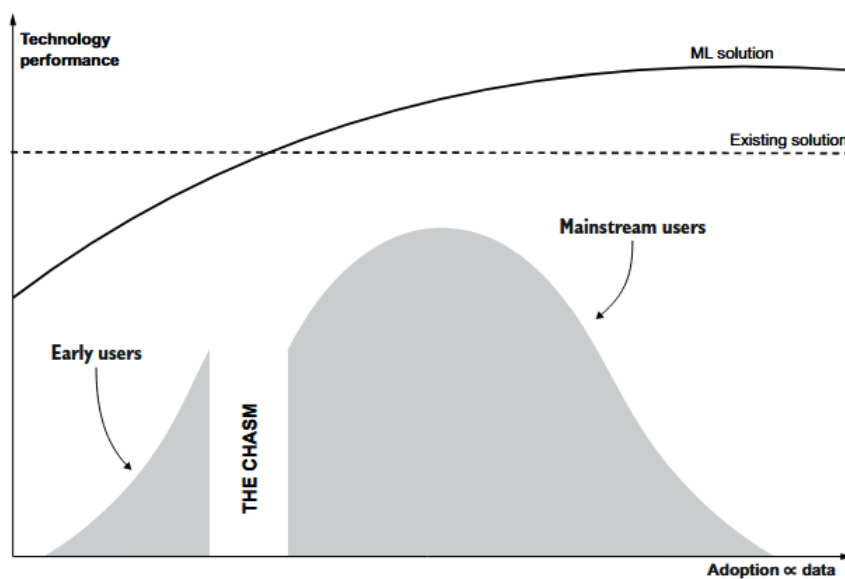
Sekarang, mari kita kembali ke platform real estat kita dan melihat bagaimana inovasi berbasis ML kita berhubungan dengan lima kelompok yang telah kita perkenalkan. Bayangkan algoritma pertama Anda lebih buruk daripada broker manusia: katakanlah prediksinya memiliki margin kesalahan 10%, sementara broker manusia hanya memiliki tingkat kesalahan 5%. Pemilik rumah yang terburu-buru tidak akan khawatir dengan peningkatan tingkat kesalahan, dan lebih tertarik untuk segera mendapatkan angka tanpa menunggu kunjungan broker manusia. Atau mungkin mereka adalah penggemar teknologi yang ingin segera mencoba fitur-fitur baru berbasis AI yang Anda tawarkan. Mereka adalah para inovator dan pengguna awal yang akan memulai siklus yang baik.

Setelah orang-orang ini mulai menggunakan produk Anda, Anda akan mulai mendapatkan lebih banyak data. Anda dapat menggunakan data baru ini untuk melatih kembali algoritma AI Anda. Mari kita asumsikan kinerja algoritma Anda meningkat dan mendekati broker sungguhan, katakanlah dengan margin kesalahan 7%. Setelah Anda mulai menutup kesenjangan akurasi dengan broker manusia, bahkan pengguna yang kurang berani pun akan memilih untuk mengorbankan akurasi 2% itu demi peningkatan kecepatan. Hasilnya, sekali lagi, akan berupa lebih banyak data yang dapat Anda gunakan untuk terus meningkatkan model.

Sekarang bayangkan bahwa berkat data baru yang Anda kumpulkan melalui model ML primitif, Anda akhirnya mencapai tonggak penting paritas manusia-mesin: algoritma Anda sama baiknya dengan broker ahli. Menggunakan produk Anda sudah jelas pada tahap ini; pengguna mendapatkan kinerja yang sama dengan broker ahli, tetapi hanya membutuhkan sedikit untuk mendapatkannya, tanpa perlu menjadwalkan kunjungan. Inilah tingkat kinerja yang dibutuhkan untuk menarik mayoritas pengguna awal: mereka tidak keberatan menggunakan teknologi baru selama tidak ada kekurangan dibandingkan dengan status quo. Seperti halnya usaha bisnis lainnya, sangat penting bagi Anda untuk dapat mengomunikasikan kinerja teknologi Anda dengan tepat kepada orang-orang, dan mereka dapat memahaminya. Ini bukan hal yang mudah, dan Anda tidak boleh menganggapnya remeh.

Jika Anda berhasil mengomunikasikan potensi teknologi Anda, penjual rumah akan ingin menggunakan model Anda untuk mendapatkan penawaran harga instan untuk rumah mereka. Bisakah Anda menebak hasilnya? Anda benar: lebih banyak data. Sekarang, algoritme Anda dapat memanfaatkan data tambahan ini untuk melampaui akurasi manusia. Selamat, Anda sekarang memiliki prediktor harga rumah AI yang super. Gambar 9.12 menunjukkan hubungan yang saling memperkuat antara kinerja teknologi dan adopsi.

Jika Anda tidak mengirimkan produk primitif pertama, Anda tidak akan pernah memulai proses yang membawa Anda ke prediktor harga AI yang super. Karena alasan ini, penting untuk memperhitungkan siklus positif AI dan mengirimkan produk AI meskipun produk tersebut tidak sempurna.



Gambar 9.12 Karena kekuatan data, kinerja masalah berbasis ML bergantung pada adopsi pasar. Pengguna awal mungkin bersedia berkompromi dalam hal kinerja, tetapi seiring bertambahnya penggunaan produk dan data, Anda akan meningkatkan teknologi hingga dapat menjangkau pasar arus utama.

Luncurkan produk lebih awal, dan rancang strategi Anda agar setiap iterasi produk berfungsi untuk meningkatkan algoritma Anda, hingga keunggulan Anda begitu besar sehingga tidak dapat diabaikan.

Jika Anda yakin bahwa perolehan lebih banyak data adalah kunci untuk memungkinkan Anda memperoleh keunggulan atas pesaing, Anda dapat mempertimbangkan cara-cara yang tidak lazim untuk mendapatkan pengguna awal pertama tersebut. Strategi ekstremnya adalah dengan memberikan diskon besar-besaran untuk layanan baru Anda atau bahkan menggratiskannya agar Anda dapat dengan cepat menarik pengguna awal. Memulai siklus AI yang baik sangatlah penting sehingga Anda bahkan mungkin berencana untuk membayar orang-orang agar mulai menggunakan produk Anda. Anda akan menutupi kerugian awal ini setelah siklus AI yang baik dimulai, teknologi Anda mulai berkembang pesat, dan segmen arus

utama mulai menggunakan layanan baru Anda dan membayar untuk nilai yang diberikannya kepada mereka.

9.4 MENGELOLA PROYEK AI

Strategi Lean AI membantu Anda memutuskan apakah akan membangun, meminjam, atau membeli solusi teknologi. Apa pun pilihannya, Anda tetap harus melakukan beberapa pekerjaan pengembangan. Misalnya, meskipun Anda mengadopsi solusi Pinjam untuk masalah prediksi harga rumah, tim tetap harus menulis kode pipa ledeng untuk menghubungkan keluaran model ke aplikasi web.

Salah satu rahasia terburuk yang dijaga di industri ini adalah bahwa mengelola insinyur itu seperti menggembala kucing. Bagian ini akan memberi Anda beberapa masukan untuk memimpin operasi sehari-hari selama fase implementasi proyek AI. Sekalipun Anda tidak ingin mengambil tanggung jawab ini secara pribadi, atau memutuskan untuk mengalihdayakan seluruh upaya kepada konsultan, tetap bermanfaat untuk mengetahui seperti apa praktik rekayasa perangkat lunak modern. Filosofi lean juga meluas ke tahap implementasi: yang terbaik adalah memecah pengembangan menjadi beberapa bagian yang dapat dikerjakan oleh masing-masing anggota tim dalam satu atau dua minggu. Kecepatan iterasi yang begitu cepat memungkinkan Anda untuk mengikuti perubahan persyaratan dari organisasi, dan bahkan umpan balik dari pelanggan dan pengguna produk.

Lean juga berarti mengirimkan perangkat lunak sesegera mungkin dalam banyak kasus, segera setelah MVP siap. Kami telah melihat banyak tim tanpa henti berjuang untuk kesempurnaan, sehingga menunda momen krusial ketika mereka bisa mendapatkan informasi yang berarti dari pengguna mereka.

Pendekatan lean memberi Anda kerangka kerja konseptual yang menekankan iterasi cepat, perbaikan berkelanjutan, dan perhatian terhadap umpan balik. Namun, bagaimana hal ini diterapkan dalam praktik di seluruh aktivitas harian gugus tugas AI Anda? Buku ini bukan tentang manajemen proyek, tetapi kami rasa bermanfaat bagi Anda untuk mendapatkan pengetahuan tentang bagaimana prosesnya.

Sebuah praktik yang selaras dengan prinsip-prinsip Strategi AI Lean dan digunakan oleh sebagian besar perusahaan Silicon Valley saat ini adalah Scrum, salah satu bentuk pengembangan Agile. Anggap Scrum sebagai buku implementasi dan aturan praktis yang membantu tim menerapkan konsep-konsep Strategi AI Lean. Iterasi cepat dicapai dengan membagi pekerjaan menjadi beberapa bagian (timeboxing) selama dua minggu, yang disebut sprint. Kemajuan menuju tujuan keseluruhan diuraikan menjadi cerita-cerita (story), yang merupakan item pekerjaan mandiri yang menghasilkan fitur lengkap untuk produk. Di awal setiap sprint, tim memutuskan cerita mana yang akan mereka kerjakan selama dua minggu ke depan.

Karena tugas diputuskan dan dialokasikan setiap dua minggu, terdapat banyak peluang untuk menyesuaikan diri dengan perubahan persyaratan, atau bahkan upaya yang gagal untuk mengimplementasikan model atau strategi pengumpulan data. Di banyak bidang AI, tidak jarang menghabiskan beberapa minggu untuk mengembangkan model, membersihkan data,

dan melakukan pelatihan, hanya untuk mengetahui bahwa pendekatan tersebut tidak berkinerja sebaik yang diharapkan. Dalam kasus tersebut, menjaga kecepatan iterasi yang cepat sangat penting untuk memastikan bahwa tim pada akhirnya akan mencapai solusi.

Lean adalah konsep umum yang sangat efektif untuk berbagai jenis proyek rekayasa perangkat lunak, terutama ketika melibatkan ketidakpastian teknis atau perubahan persyaratan. Hal ini juga berlaku untuk proyek AI, di mana dalam banyak kasus Anda tidak akan tahu seberapa baik kinerja model Anda sampai Anda mencobanya pada tugas Anda. Namun, untuk beberapa aspek, AI mengubah aturan mainnya, dan bagian ini menjelaskan caranya. Kami telah memutuskan untuk berfokus pada dua aspek utama yang menyebabkan proyek AI keluar jalur:

- Anda selalu mengkhawatirkan data.
- Proyek AI tidak pernah benar-benar "selesai".

Mari kita bahas yang pertama. Pengumpulan (dan pembersihan) data serta pengembangan model seringkali dilakukan oleh orang yang berbeda dalam tim Anda, hanya karena keahlian yang dibutuhkan berbeda. Hal ini menciptakan semacam masalah ayam dan telur: para insinyur perangkat lunak bergantung pada insinyur ML untuk mencapai kemajuan, dan para insinyur ML tidak memiliki data yang tersedia untuk eksperimen mereka.

Mengikuti prinsip iterasi inkremental, cara yang baik untuk memutus lingkaran setan ini adalah dengan membagi waktu beberapa siklus pengumpulan data dan pengembangan model. Segera setelah para insinyur perangkat lunak menyelesaikan, bahkan sebagian kecil dari set data pelatihan, mereka dapat menyerahkannya kepada para insinyur ML untuk dikembangkan dan diuji modelnya. Sementara itu, mereka dapat melanjutkan untuk meningkatkan dan memperluas skala kode mereka guna meningkatkan kinerja, sambil menerima umpan balik dari para ahli ML tentang area terpenting yang perlu ditingkatkan. Proses ini dimulai kembali pada siklus berikutnya, memberi Anda cara yang baik untuk mengukur kemajuan dari waktu ke waktu.

Mari kita bahas masalah kedua sekarang: Proyek AI tidak pernah benar-benar "selesai". Jika Anda mengerjakan proyek perangkat lunak konvensional, persyaratan yang jelas membantu Anda menentukan kapan produk "selesai". Jika Anda memimpin tim yang membuat sistem alarm rumah, Anda tentu perlu memastikan bahwa membuka jendela akan memicu sirene. Sebagai seorang penggiat AI, Anda tidak memiliki kemewahan ini. Ini karena AI modern didasarkan pada pembelajaran mesin, sehingga tidak akan pernah mencapai kesempurnaan 100% yang kita kaitkan dengan rekayasa konvensional.

Dalam banyak proyek, sulit untuk menghubungkan metrik di seluruh organisasi, seperti pertumbuhan teratas, dengan akurasi model. Anda akan tahu Anda berada dalam salah satu situasi tersebut ketika Anda kesulitan memberi tahu tim Anda tingkat kinerja yang Anda butuhkan dari model tersebut. Sekali lagi, iterasi inkremental adalah kunci untuk berhasil mengeluarkan Anda dari situasi ini. Semakin cepat Anda mengeluarkan prototipe proyek yang berfungsi, semakin cepat Anda dapat membuat organisasi menggunakannya, dan dengan demikian melihat bagaimana akurasi model berkorelasi dengan peningkatan dalam organisasi (atau kepuasan pengguna). Contoh bagus yang telah kita bahas adalah prediksi churn, karena

kita dapat langsung melihat bagaimana penyempurnaan model meningkatkan retensi. Bahkan prototipe dengan akurasi rendah pun dapat menjadi kemenangan besar bagi organisasi dan niscaya akan membantu Anda menyampaikan pesan.

Kami memastikan untuk menyertakan beberapa contoh proyek AI yang gagal dan poin-poin teknis yang disalahpahami di sepanjang buku ini. Namun, sebagai penulis yang teliti, kami tidak dapat mengakhiri bab ini tanpa diskusi yang lebih lengkap tentang apa yang harus dilakukan ketika segala sesuatunya mulai berjalan tidak sesuai rencana.

9.5 KETIKA AI GAGAL

Bekerja di bidang teknologi telah mengajarkan kita bahwa jika semuanya berjalan sesuai harapan, banyak orang akan kehilangan pekerjaan. Membangun dan memelihara perangkat lunak itu sulit, dan AI pun demikian. Sebaliknya, AI justru menambah tantangannya sendiri. Seluruh bagian ini dikhususkan untuk kisah-kisah kegagalan proyek AI. Kami memiliki dua alasan bagus untuk melakukan ini:

- Untuk memperingatkan bahwa AI bukanlah solusi instan
- Untuk membuktikan bahwa kesalahan dalam strategi AI bisa berakibat fatal

Tujuannya bukan untuk menakut-nakuti Anda agar tidak mengerjakan proyek Anda, melainkan untuk menemukan bagaimana beberapa strategi yang kami perkenalkan dalam buku ini mungkin telah menyelamatkan para tokoh utama dalam kisah-kisah ini. Kami memutuskan untuk menyimpannya di akhir bagian 2 agar Anda dapat melihat bagaimana semuanya berjalan dengan semestinya.

Anki

Contoh pertama adalah tentang Anki, sebuah perusahaan rintisan robotika yang tutup pada tahun 2019 setelah menerima pendanaan lebih dari Rp. 2 triliun. Pada tahun 2013, Anki begitu menjanjikan sehingga mendapatkan kesempatan langka untuk diundang ke panggung pada acara utama tahunan Apple. Di sana, CEO Apple, Tim Cook, menyampaikan kepada ribuan penggemar teknologi bahwa perusahaan yang sedang berkembang ini akan "membawa AI dan robotika ke dalam kehidupan kita sehari-hari."

Enam tahun kemudian, seorang juru bicara perusahaan mengatakan kepada Recode hal berikut:

Terlepas dari kesuksesan kami di masa lalu, kami mengupayakan setiap jalur finansial untuk mendanai pengembangan produk kami di masa mendatang dan memperluas platform kami. Kami kekurangan dana yang signifikan untuk mendukung bisnis perangkat keras dan perangkat lunak serta menjembatani peta jalan produk jangka panjang kami.

Saat perusahaan tersebut tutup, produk terancangannya di pasaran adalah Vector, mainan robot seharga Rp. 2.500.000 yang memelopori "kecerdasan emosional" khas Anki, yaitu kemampuan untuk memahami lingkungan dan mengekspresikan emosi kepada manusia di sekitarnya. Namun, Anki sudah mengiklankan bahwa mereka memiliki rencana yang lebih ambisius,

bahkan sampai pada pengembangan asisten rumah tangga humanoid. Kami tidak ingin terlibat dalam kritik yang asal-asalan, tetapi kami ingin menunjukkan potensi ketidakseimbangan yang kami lihat antara kemenangan jangka pendek dan strategi jangka panjang. Dalam buku ini, kami menekankan pentingnya menjalankan proyek AI dengan batasan waktu yang memberikan hasil cepat bagi organisasi. Sebaliknya, Anki mengalihkan sumber daya dari pengembangan produk ke "AI masa depan", rencana induk untuk menguasai dunia yang tidak pernah terwujud. Visi berbasis AI tentang "robot sehari-hari" sudah ada, tetapi belum pernah dipecah menjadi bagian-bagian kecil yang dapat diimplementasikan seiring waktu.

Lighthouse AI

Kisah kegagalan lainnya datang dari Lighthouse AI, sebuah perusahaan yang mengumpulkan dana Rp 170 miliar untuk membangun kamera keamanan rumah bertenaga AI. Kamera tersebut menggunakan AI untuk mengekstrak informasi bermanfaat dari rekaman video, dan memungkinkan Anda bertanya tentang apa yang terjadi saat Anda pergi, dalam bahasa alami. Misalnya, Anda dapat bertanya, "Jam berapa anak-anak pulang kemarin?" dan mendapatkan rekaman video sebagai jawaban.

Kedengarannya bermanfaat, bukan? Pasar tidak setuju. Seperti yang ditulis CEO Alex Teichman di situs web perusahaan:

Saya sangat bangga dengan pekerjaan inovatif yang dicapai tim Lighthouse—menghasilkan kecerdasan yang bermanfaat dan mudah diakses untuk rumah kita melalui AI canggih dan penginderaan 3-D. Sayangnya, kami tidak mencapai kesuksesan komersial yang kami harapkan dan akan segera menutup operasinya.

Meskipun produknya memang mengesankan dari segi teknis, tampaknya perusahaan tidak melakukan riset pasar dan eksperimen yang memadai sebelum terjun ke usaha ini. Anda dapat menggunakan berbagai strategi untuk menguji minat pasar sebelum menginvestasikan jutaan dolar untuk R&D; misalnya, menjalankan kampanye crowdfunding atau mengumpulkan pesanan awal. Sulit untuk mengatakan apakah strategi ini akan menjadi solusi yang pasti, tetapi Anda pasti dapat belajar dari contoh ini bahwa teknologi tercanggih sekalipun tidak akan cukup untuk menciptakan pasar. Sebaliknya, identifikasi ancaman bisnis terbesar bagi proyek Anda (pertama-tama, orang-orang yang tidak mau repot-repot membelinya) dan rancang strategi kreatif untuk mengelolanya.

IBM Watson dalam Onkologi

Sekarang mari kita bahas kasus kegagalan AI yang luar biasa lainnya: kegagalan IBM Watson dalam Onkologi. Pada tahun 2013, IBM mengeluarkan siaran pers yang membanggakan keterlibatannya dengan salah satu pusat medis terkemuka di dunia:

Pusat Kanker MD Anderson Universitas Texas dan IBM hari ini mengumumkan bahwa MD Anderson menggunakan sistem komputasi kognitif IBM Watson untuk misinya memberantas kanker.

Hanya lima tahun kemudian, STAT (jurnal yang berfokus pada industri kesehatan) meninjau dokumen internal IBM tentang proyek MD Anderson dan membagikan kutipan dari salah satu dokter yang berpartisipasi dalam uji coba tersebut:

Produk ini sampah. Kami membelinya untuk pemasaran dan dengan harapan Anda akan mencapai visinya. Kami tidak dapat menggunakannya untuk sebagian besar kasus.

Apa yang terjadi di sini? Kali ini, permintaan pasar jelas kuat: semua orang di bumi ingin melihat kanker diberantas. Masalah pertama yang kita lihat dalam kasus ini adalah fokus yang tidak beralasan pada visi AI jangka panjang. "Memberantas kanker" memang diinginkan, tetapi mustahil untuk diselesaikan sendirian oleh satu perusahaan teknologi hanya karena mereka telah mencapai kesepakatan dengan sebuah rumah sakit. Selain itu, Anda akan dapat menemukan serangkaian kesalahan teknis saat kami menyoroti cara kerja internal IBM Watson for Oncology. Teknologi medis ini dibangun di atas apa yang telah dikembangkan IBM untuk memainkan acara kuis AS Jeopardy dan menarik perhatian setelah menang melawan dua juara manusia pada tahun 2011.

Jeopardy adalah acara kuis di mana peserta menerima jawaban yang harus mereka tebak pertanyaannya (pada dasarnya, kuis terbalik). Bagaimana IBM mengubah model Jeopardy untuk membantu para ahli onkologi? Dimulai dengan menemukan kumpulan data yang sesuai. Salah satunya adalah serangkaian 5.000 pertanyaan medis dari *American College of Physicians* (ACP). Berikut ini adalah contoh pasangan pertanyaan/jawaban dalam kumpulan data tersebut:

Q: Tes skrining kanker kolorektal yang terkait dengan kepatuhan pasien tertinggi

A: Tes imunokimia feses

Q: Limfoma non-Hodgkin yang ditandai dengan keterlibatan ektranodal dan ekspresi berlebihan siklin D1

A: Limfoma sel mantel

Q: Mekanisme varikokel akut yang terkait dengan karsinoma ginjal

A: Obstruksi vena testis

Pada dasarnya, IBM Watson menggunakan data teks sebagai masukan dan keluaran model. Karena Anda telah mempelajari di bab 4 bahwa data teks masih merupakan sektor tersulit untuk AI, mari kita lihat masalah spesifik apa yang harus dihadapi IBM ketika mentransfer teknologinya ke rumah sakit. Di MD Anderson, baik "leukemia limfoblastik akut", sejenis kanker sel darah, maupun "alergi" sering disebut dengan akronim "ALL". Jelas, kanker dan alergi adalah dua kondisi medis yang sangat berbeda, tetapi Watson tidak dapat membedakan keduanya karena menggunakan akronim yang sama.

Masalah lainnya adalah Watson kesulitan mempertimbangkan berbagai aspek pasien, yang menyebabkan potensi bencana. Misalnya, penelitian ini menyarankan bahwa seorang

pria berusia 65 tahun yang didiagnosis kanker paru-paru dan menunjukkan bukti pendarahan hebat harus diobati dengan kombinasi kemoterapi dan obat yang disebut bevacizumab, yang dapat menyebabkan "perdarahan parah atau fatal" dan tidak boleh diberikan kepada pasien yang mengalami pendarahan serius.

Apa yang bisa kita pelajari dari ini? Kita sudah mengetahui beberapa hal yang muncul dari artikel ini: data teks sulit ditangani, terutama ketika Anda bekerja dalam konteks yang sangat rumit. Dalam kerangka kerja lebar-kedalaman yang kami perkenalkan untuk menggambarkan kompleksitas aplikasi AI (bab 4), Watson akan mendapatkan skor yang sangat tinggi baik dalam hal lebar maupun kedalaman. Lebar tinggi karena model perlu memahami spektrum kata yang luas yang berasal dari berbagai cabang bidang medis, dan semua potensi akronim dan singkatan juga. Kedalamannya juga sangat tinggi, karena kami meminta algoritme kami untuk menghasilkan kalimat yang menggambarkan kombinasi obat yang berpotensi kompleks.

Tantangan dalam bekerja dengan data semacam ini disadari oleh Dr. Amy Abernethy, kepala petugas medis di Flatiron Health dan mantan direktur penelitian kanker di Duke Cancer Institute. Dalam sebuah wawancara untuk sebuah makalah yang diterbitkan dalam *Journal of the National Cancer Institute* tahun 2017, ia menyatakan, "Pengalaman MD Anderson memberi tahu kita bahwa memecahkan masalah kualitas data dalam data yang tidak terstruktur merupakan tantangan yang jauh lebih besar bagi kecerdasan buatan daripada yang awalnya diantisipasi."

Salah satu cara untuk menyederhanakannya adalah dengan mengurangi kedalaman dengan mengajukan masalah sebagai tugas klasifikasi yang menangani data terstruktur. Cara lain adalah dengan mengurangi lebar dengan membatasi aplikasi pada jenis kanker tertentu. Dengan menerapkan kedua ukuran tersebut, alih-alih memberi algoritme kalimat bertele-tele yang penuh dengan jargon medis, mereka akan menggunakan tabel dengan informasi pasien yang dinyatakan dalam nilai (misalnya, "tekanan darah = 110 mmHg"). Algoritmenya akan merespons dengan memilih satu dari sejumlah terapi yang disarankan dan sudah diketahui sebelumnya. Proyek seperti ini tentu kurang keren dibandingkan AI serba tahu yang merespons dalam bahasa alami, tetapi kami berpendapat bahwa lebih baik memiliki sesuatu yang sederhana namun efektif daripada sesuatu yang rumit namun tidak efektif.

Buku Harian Emosional

Terakhir, mari kita bahas dua contoh dari pekerjaan konsultasi kami yang tidak akan Anda temukan di berita. Kami mengubah beberapa detail untuk melindungi privasi klien kami dan menghormati perjanjian kerahasiaan kami, tetapi pelajarannya tetap relevan.

Contoh pertama sebenarnya lebih merupakan bencana yang dapat dihindari. Sebuah perusahaan layanan kesehatan ingin membangun sebuah aplikasi untuk memberikan dukungan emosional kepada pasangan yang menjalani perawatan kesuburan. Pasangan-pasangan ini tidak hanya stres karena kesulitan untuk hamil, tetapi wanita tersebut juga diberi resep hormon yang dapat berdampak kuat pada suasana hatinya. Ide awalnya adalah membangun sebuah aplikasi "buku harian emosional" untuk pasangan tersebut. Setiap hari, aplikasi tersebut akan meminta mereka untuk mengunggah swafoto dengan wajah tersenyum

atau sedih. Sebuah algoritma akan mengenali suasana hati mereka dan menjawab dengan kalimat motivasi yang tepat. Di akhir terapi, aplikasi tersebut akan membuat sebuah "buku harian" tentang perjalanan emosional yang dilalui pasangan tersebut, yang diharapkan akan berakhir dengan anugerah indah berupa seorang anak.

Perusahaan telah mulai mengumpulkan kutipan dari berbagai penyedia teknologi dan telah menyusun rencana pemasaran untuk mempromosikan aplikasi tersebut di berbagai pusat perawatan. Suatu pagi, kami menemukan bahwa Microsoft menawarkan "API deteksi emosi": sebuah layanan sederhana yang memungkinkan Anda mengirim swafoto ke server Microsoft, yang menjalankan algoritma AI dan memberikan evaluasi mereka terhadap emosi yang diungkapkan oleh orang tersebut. Kami meminta untuk menghentikan sementara strategi tersebut selama sehari. Keesokan harinya, kami kembali dengan sebuah prototipe aplikasi yang terhubung ke layanan Microsoft, yang sudah menyertakan beberapa kalimat untuk menyemangati pasangan tersebut.

Kami menggunakan prototipe sederhana kami dan pergi ke pusat fertilitas untuk berbicara dengan beberapa pengguna. Keputusannya bulat: para wanita ingin melempar ponsel itu ke dinding. Apa yang kami pikir akan menjadi alat pendukung justru dianggap sebagai pelanggaran privasi dan keintiman yang tidak sopan. Mereka sudah stres dan ragu akan menjadi ibu; hal terakhir yang mereka inginkan adalah aplikasi yang menginginkan swafoto untuk mengirimkan kutipan motivasi yang nakal. Kami membatalkan program tersebut dan kembali ke rencana awal. Pembuatan prototipe yang cepat dan penggunaan API memungkinkan kami untuk mengetahui sejak awal proses bahwa kami akan langsung menabrak tembok.

Panggilan Telepon Marah

Contoh lain adalah sebuah perusahaan yang membutuhkan algoritma AI yang dapat mengidentifikasi emosi dari rekaman suara. Selama panggilan telepon, aplikasi harus menunjukkan kapan pembicara sedang marah, sedih, senang, atau netral dan memberikan saran untuk mengatasi perasaan negatif tersebut. Klien kami datang kepada kami dengan klaim bahwa mereka telah menemukan penyedia teknologi yang andal, dan kami langsung mengeksekusinya tanpa meragukan teknologi tersebut. Bersama-sama, kami menyusun strategi yang komprehensif, rencana bisnis, dan merancang produk minimum yang layak (MVP) yang mengintegrasikan teknologi pihak ketiga. Setelah kami mengirimkan MVP, kami menemukan bahwa kinerja penyedia teknologi tersebut jauh dari yang mereka klaim (dan kami harapkan). Tampaknya, perangkat lunak tersebut dirancang berdasarkan kumpulan data penutur bahasa Jerman. Kami perlu menerapkannya di pasar Italia, dan ternyata orang Jerman yang marah dan orang Italia yang marah terdengar sangat berbeda.

Setelah menyadari bahwa vendor tersebut berkinerja buruk, kami mempelajari literatur ilmiah untuk mempelajari performa terkini dalam tugas ini. Kami menemukan bahwa bahkan performa terkini pun jauh dari akurasi lebih dari 90% yang dicapai dalam tugas AI lainnya. Pelajaran yang dipetik? Uji, uji, uji. Selalu pertanyakan dan validasi apa yang Anda baca. Sekalipun penyedia teknologi mengklaim memiliki produk yang hebat, luangkan satu

atau dua hari untuk meneliti performa terkini dan menguji apakah produk tersebut dapat mencapai performa yang Anda butuhkan.

Penjualan yang Berkinerja Buruk

Dalam contoh lain, sebuah perusahaan besar meminta kami untuk mengelompokkan data penjualannya guna menemukan akun pelanggan yang berkinerja buruk. Permintaan tersebut datang dari departemen TI, yang mengelola data dalam jumlah besar dan ingin membuktikan nilai dari upaya pengumpulan data tersebut.

Kami menjalankan analisis dan menemukan beberapa toko berkinerja buruk yang sebenarnya dapat menghasilkan Rp. 85 miliar lebih banyak per tahun bagi perusahaan dengan mengubah strategi penjualan! Semua orang bersemangat, jadi kami menyusun presentasi yang indah untuk CEO. Ia menyela kami setelah beberapa menit dengan berkata: "Teman-teman, kalian tidak memberi tahu saya sesuatu yang baru. Kami membuat kesepakatan khusus 10 tahun yang lalu dengan toko-toko yang kalian sebut berkinerja buruk; saya dapat memberi tahu kalian bahwa keuntungannya sangat tinggi dengan mereka. Kalian melihat metrik yang salah."

Ternyata toko-toko yang dimaksud memiliki diskon volume khusus yang dipesan secara berbeda, sehingga tidak muncul dalam basis data yang kami gunakan. Apa yang kita pelajari dari kisah ini? Selalu sertakan perspektif bisnis saat merancang proyek AI. Jangan terjebak dalam merancang proyek AI dengan pendekatan yang mengutamakan teknologi: pastikan nilai bisnisnya ada sebelum memikirkan algoritma.

Anda mungkin menyadari bahwa tidak satu pun solusi untuk proyek-proyek baru ini adalah berinvestasi dalam teknologi yang lebih baik. Pahlawan yang mereka butuhkan bukanlah ilmuwan data yang sangat terampil, melainkan seorang pemimpin yang tercerahkan yang memahami prinsip-prinsip yang kita bahas dalam buku ini dan memiliki pengetahuan domain, semangat kritis, serta dorongan untuk membawa AI ke dalam sebuah organisasi.

Selamat! Anda pada dasarnya (hampir) berhasil mencapai akhir buku ini! Bab 2 membahas banyak materi dengan situasi dan kisah nyata yang akan membantu Anda menemukan dan menyelesaikan proyek AI yang berdampak di organisasi Anda. Bersama dengan dasar yang telah kita letakkan di bab 1, Anda sekarang siap untuk berpikir kritis tentang bagaimana AI dapat bermanfaat bagi organisasi Anda. Tarik napas dalam-dalam dan bersiaplah untuk bab terakhir, di mana kita akan mundur selangkah dan membahas bagaimana perangkat AI yang telah Anda pelajari akan membentuk masyarakat di masa depan.

BAB 10

APA YANG AKAN DATANG

Bab ini membahas

- Kemungkinan implikasi negatif AI terhadap masyarakat
- Cara-cara paling menjanjikan di mana AI dapat membentuk kembali umat manusia
- Potensi disrupsi oleh AI di industri-industri tertentu

Ini adalah bab terakhir buku ini. Sekarang, Anda memiliki semua alat untuk memahami apa itu AI dan apa saja keluarga algoritma terpenting yang dapat bermanfaat. Kami juga berbagi dengan Anda pelajaran yang telah kami dan para pemimpin industri pelajari melalui pembangunan proyek-proyek AI. Namun, sebagaimana pendakian gunung tidak akan lengkap tanpa berhenti sejenak untuk melihat lanskapnya, perjalanan kita tidak akan lengkap tanpa pandangan yang lebih luas tentang masa depan AI. Berkat pengetahuan baru Anda tentang aspek teknis dan organisasional AI, Anda sekarang memiliki pola pikir yang Anda butuhkan untuk menembus gambar-gambar dan berpikir kritis tentang bagaimana teknologi ini akan membentuk masyarakat dalam 5 hingga 10 tahun ke depan.

Buku ini berfokus pada apa yang dapat dilakukan AI dalam batasan-batasan organisasi. Untuk bab terakhir ini, izinkan kami untuk sedikit lebih filosofis sembari memperluas cakrawala diskusi kami ke masyarakat secara keseluruhan. Untuk memberikan pandangan yang seimbang, kami memutuskan untuk membingkai diskusi kami dalam konteks ancaman dan peluang. Kami juga menyertakan tinjauan beberapa tren spesifik industri yang perlu Anda ketahui. Perlu diingat bahwa ide-ide yang akan kami diskusikan adalah ide kami sendiri: kami mendorong Anda untuk menantanginya dan berpikir kritis.

10.1 BAGAIMANA AI MENGANCAM MASYARAKAT

Setiap kali kita, manusia, mengembangkan teknologi inovatif, butuh waktu sebelum kita menyadari potensi risiko dan kerugiannya. Pernahkah Anda melihat salah satu iklan era pasca-Perang Dunia II yang terlalu optimis tentang energi nuklir? Konon, kita semua akan menggunakan bola lampu atom dan sarapan yang diperkaya Uranium. Pesawat jet pertama di tahun 40-an memiliki jendela persegi karena lebih mudah diproduksi. Butuh beberapa kecelakaan sebelum para insinyur menyadari bahwa sudut-sudut tajam sering kali menyebabkan retakan fatal pada badan pesawat. Tampaknya ada bias fisiologis yang mendorong kita untuk mengabaikan sisi negatif teknologi baru.

Kita seharusnya tidak membuat kesalahan yang sama dengan AI. Salah satu sifat terpenting dari produk perangkat lunak saat ini adalah sangat mudah untuk ditingkatkan dan melayani jutaan atau bahkan miliaran orang. Bayangkan apa yang akan terjadi jika aplikasi AI yang menyesatkan, cacat, atau benar-benar berbahaya didistribusikan kepada miliaran orang. Hasilnya akan sangat buruk.

Kita sudah memiliki beberapa contoh (untungnya) terbatas tentang hal ini. Pada tahun 2016, Microsoft merilis chatbot yang dapat belajar berkicau berdasarkan kicauan yang

dikirimkan orang. Setelah kurang dari sehari, chatbot tersebut berkembang menjadi rasis. Jika ada sesuatu yang dapat diajarkan oleh eksperimen ini kepada kita, itu adalah bahwa AI dapat menjadi sangat salah dengan sangat cepat, yang merupakan alasan mengapa penting untuk mengetahui kelemahan teknologi dan merancang produk dengan mempertimbangkannya.

Di bagian ini, kami akan menyoroti apa yang kami yakini sebagai ancaman utama yang harus dihadapi masyarakat dan organisasi ketika produk dan layanan berbasis AI mulai mencapai puncak popularitas. Sebagai teknolog dan optimis, kami tidak bermaksud membohongi diri sendiri: tujuan kami adalah memberi Anda pengetahuan dan perspektif untuk membentuk opini Anda sendiri.

Bias dan Keadilan

Di bab 8, kami telah memperingatkan Anda tentang bahaya data yang bias (atau tidak seimbang), dan bagaimana hal itu sering kali mengarah pada pengembangan model pembelajaran mesin yang bias serupa. Bagian ini melengkapi pembahasan teknis tersebut dengan diskusi yang lebih mendalam tentang implikasi sosial dan etika dari masalah ini.

Secara umum, Anda perlu memahami dua jenis bias:

- Bias distribusi pelatihan
- Bias pelabelan

Anda telah melihat contoh bias distribusi pelatihan di bab 8, di mana kami membahas model visi komputer yang kesulitan membedakan domba dari ladang hijau. Karena sebagian besar contoh dombanya memiliki latar belakang rumput, model tersebut belajar berhalusinasi domba bahkan dalam gambar ladang yang tidak berlatar rumput. Anda mungkin berpikir bahwa bias semacam ini bukanlah masalah besar. Hal terburuk yang dapat ditimbulkannya adalah mengganggu pengguna dan mengurangi akurasi model, bukan? Pikirkan lagi.

Amazon mengembangkan model AI untuk membantu departemen SDM-nya memilih kandidat pekerjaan secara otomatis. Set pelatihan diperoleh dengan mencocokkan resume karyawan saat ini dengan hasil tinjauan kinerja di tempat kerja. Sayangnya, sudah diketahui umum bahwa perusahaan teknologi memiliki masalah kesenjangan gender. Menurut sebuah studi tahun 2018 oleh McKinsey & Company dan Pivotal Ventures, “Mem-boot Ulang Representasi: Menggunakan CSR dan Filantropi untuk Menutup Kesenjangan Gender dalam Teknologi,” hanya 26% tenaga kerja teknologi adalah perempuan. Jumlah tersebut turun menjadi 11% untuk peran eksekutif senior.

Sekarang setelah Anda mengetahui lebih banyak tentang ML, Anda mungkin dapat menebak apa yang terjadi ketika data SDM dimasukkan ke dalam algoritma ML. Karena sebagian besar contoh “positif” dalam tenaga kerja saat ini (dan dengan demikian dalam set data pelatihan) adalah laki-laki, model akan mengambil bahwa menjadi laki-laki adalah prediktor kinerja kerja yang baik. Bahkan mengosongkan nama depan dengan harapan menyeimbangkan gender dalam algoritma tidak akan membantu memecahkan masalah, karena model AI sangat pandai menangkap petunjuk halus. Misalnya, hanya 5,6% anggota klub sepak bola di Inggris adalah perempuan. Ini berarti model dapat dengan mudah menyadari fakta bahwa siapa pun yang mencantumkan sepak bola di resume mereka kemungkinan besar adalah laki-laki.

Pengenalan ucapan adalah area lain di mana bias pelatihan dapat merusak hari. Film dengan teks terjemahan merupakan sumber data yang bagus untuk melatih algoritma dalam mentranskripsikan dialog lisan menjadi kata-kata tertulis. Namun, ternyata sebagian besar aktor Hollywood memiliki pengucapan yang sempurna. Imigran dan orang-orang yang menggunakan bahasa Inggris sebagai bahasa kedua akan mendapati bahwa model kesulitan mengenali suara mereka karena aksen mereka tidak terwakili dengan baik dalam kumpulan data pelatihan.

Selama sistem ini diterapkan pada produk yang berlebihan, seperti asisten suara Siri, hal itu bukanlah masalah besar. Namun apa yang terjadi ketika mereka mulai menjadi bagian penting dalam hidup kita, seperti mobil atau kantor pemerintahan? Jika algoritma ini gagal memahami aksen bahasa Inggris imigran, mereka pasti akan didiskriminasi dan kehilangan akses ke informasi.

Jenis bias kedua yang dapat memengaruhi sistem AI adalah bias pelabelan. Bagian dari narasi sehari-hari tentang AI adalah bahwa mesin akan lebih tidak memihak dan karenanya kurang diskriminatif dibandingkan pengambil keputusan manusia. Sebaliknya, yang terjadi adalah model pembelajaran mesin (ML) hanya belajar mereproduksi bias yang sama yang muncul dalam set data pelatihan mereka.

Contoh mencolok dari masalah ini adalah sistem *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS) yang digunakan oleh beberapa pengadilan AS untuk membantu hakim memutuskan hukuman pidana. Sistem berbasis ML ini konon memperkirakan "risiko" terdakwa akan melakukan kejahatan lagi di masa mendatang dan merekomendasikan hukuman yang lebih berat bagi mereka yang lebih mungkin melakukannya. Pada prinsipnya, COMPAS adalah ide bagus yang dapat mengurangi bias rasial dan sosial dalam sistem peradilan. Namun, "Machine Bias", sebuah artikel tahun 2016 oleh lembaga nirlaba independen ProPublica, melaporkan bahwa sistem tersebut sangat bias terhadap orang kulit berwarna.

Meskipun kami tidak memiliki wawasan langsung tentang bagaimana COMPAS dirancang atau diimplementasikan, tampaknya kumpulan data yang digunakan untuk melatih model itu sendiri bias terhadap orang kulit berwarna karena didasarkan pada keputusan pengadilan sebelumnya. Dengan kata lain, model yang dibangun menggunakan data pelatihan yang berasal dari keputusan manusia sama biasnya dengan manusia yang membuat keputusan tersebut sejak awal. Bias dan kewajaran sering muncul bersamaan dalam diskusi tentang implikasi etis AI, dan oleh karena itu mudah tertukar. Anda dapat menganggap bias sebagai masalah matematika. Kumpulan data atau algoritma dapat bias terhadap fitur tertentu dan cenderung mendukungnya, tetapi ini belum tentu menjadi masalah. Kewajaran, di sisi lain, adalah konstruksi sosial.

Bayangkan kita sedang membangun sebuah algoritma untuk membantu tim sepak bola memilih pemainnya berdasarkan pemain terbaik musim sebelumnya. Kemungkinan besar algoritma tersebut akan lebih mengutamakan pemain muda dan berotot daripada pemain yang lebih tua dan kelebihan berat badan. Perilaku ini pasti bias terhadap pria muda yang bugar. Namun, apakah kita akan menganggapnya tidak adil? Kebanyakan orang tidak akan

menganggapnya demikian. Masyarakat menerima bahwa pemain sepak bola harus muda dan bugar, dan algoritma dengan perilaku ini akan dianggap efektif dan tidak terlalu tidak adil. Sekarang bayangkan sebuah algoritma serupa yang bertugas membuat keputusan perekrutan untuk departemen SDM sebuah perusahaan teknologi. Kita pasti tidak akan menganggapnya adil jika terus-menerus memilih pria muda dan bugar daripada demografi lainnya.

Namun, tidak semua harapan hilang. Dengan memilih cara memberi label, kita secara tidak langsung memutuskan jenis perilaku yang ingin kita dorong dalam algoritma kita. Jika kita membangun algoritma untuk SDM dan kita memberi label berdasarkan keputusan SDM sebelumnya, kita mendorong otomatisasi sederhana dari keputusan manusia, yang mungkin bias oleh manajer SDM sebelumnya. Di sisi lain, jika kita menemukan cara untuk memberi label berdasarkan kinerja karyawan, kita mendorong pemilihan karyawan berkinerja tinggi. Pendekatan yang sama dapat direplikasi untuk keputusan sensitif dan berpotensi mengubah hidup lainnya:

- ❖ Membuat keputusan kelayakan pinjaman berdasarkan bagaimana orang benar-benar membayar kembali pinjaman mereka, alih-alih keputusan sebelumnya dari petugas bank
- ❖ Membuat keputusan asuransi kesehatan berdasarkan kondisi kesehatan aktual seseorang, alih-alih keputusan masa lalu oleh perusahaan asuransi
- ❖ Menilai risiko seorang narapidana melakukan kejahatan setelah dibebaskan dari penjara berdasarkan perilaku historis narapidana yang sebelumnya dibebaskan, alih-alih keputusan pembebasan bersyarat sebelumnya oleh hakim

Perlu diingat bahwa pilihan desain yang lebih cerdas dan pelabelan yang sadar bias mungkin bukan solusi yang lengkap. Banyak bentuk diskriminasi lain yang dapat memengaruhi kehidupan seseorang dan dengan demikian menyuntikkan bias ke dalam data secara halus. Misalnya, meskipun kita memperhitungkan kinerja karyawan yang sebenarnya, tetapi perempuan tetap mendapatkan lebih sedikit promosi karena budaya yang bias, bias ini akan bocor ke dalam data dan algoritma akan mempelajarinya.

Solusi untuk masalah ini, mungkin tidak mengherankan, adalah menganalisis data secara akurat sebelum menggunakannya untuk membangun dan menerapkan aplikasi AI. Para peneliti sedang bekerja keras untuk membangun model mekanis kompleks yang dapat secara otomatis mendeteksi bias data. Sementara itu, kita perlu menjadi orang yang mengajukan pertanyaan yang tepat dan mengembangkan kepekaan terhadap potensi bias sosial yang dapat disembunyikan oleh data kita.

Bias dan keadilan adalah salah satu alasan terpenting kami berupaya keras untuk membantu Anda memahami bagaimana data memengaruhi algoritma ML. Karena kami berharap Anda akan menjadi protagonis era AI, penting bagi Anda untuk menyadari ancaman ini dan berupaya membangun teknologi yang melayani semua orang.

AI dan Pekerjaan

Dampak AI terhadap pekerjaan adalah topik yang telah memenuhi seluruh buku, namun masih menyisakan hipotesis yang belum teruji dan pertanyaan yang belum terjawab. Bagian singkat ini memang jauh dari jawaban yang lengkap, tetapi kami tetap ingin

menawarkan beberapa bahan renungan untuk membantu Anda mengembangkan pemikiran Anda sendiri seputar topik yang kompleks ini.

Pertama-tama, perhatikan bahwa di sepanjang buku ini, kita telah membahas tentang tugas. Seperti yang telah Anda pelajari, pembelajaran mesin modern sangat cocok untuk memecahkan masalah yang sangat terdefinisi dengan baik, dengan masukan dan keluaran yang sangat spesifik. Ini berarti kita harus mengubah narasi kita dari "AI akan menggantikan pekerjaan manusia" menjadi "AI akan menggantikan tugas manusia."

Sekarang, beberapa pekerjaan begitu sederhana sehingga dapat disederhanakan menjadi satu tugas: inilah pekerjaan yang berisiko hilang sepenuhnya. Salah satu contohnya adalah pengemudi truk, yang menghabiskan sebagian besar waktu mereka hanya untuk mengemudi. Seperti yang Anda ketahui, teknologi self-driving telah mengalami beberapa perkembangan luar biasa selama beberapa tahun terakhir, dan kemungkinan besar akan segera dapat sepenuhnya menggantikan manusia di belakang kemudi. Artinya, suatu hari nanti, mesin akan mampu melakukan tugas yang 90% waktunya dihabiskan pengemudi truk. Tak perlu dikatakan lagi apa implikasinya bagi pekerjaan mereka.

Pertanyaan yang patut kita ajukan kepada diri sendiri adalah: Jika suatu pekerjaan dapat disederhanakan menjadi satu tugas saja dan oleh karena itu dapat ditangani oleh AI masa kini, apakah kita yakin pekerjaan itu cocok untuk manusia? Kita tidak ragu menggunakan mesin untuk mengambil alih pekerjaan berbahaya, seperti memeriksa pembangkit listrik tenaga nuklir atau memantau anjungan minyak lepas pantai. Namun, beberapa pekerjaan menghancurkan jiwa manusia tanpa menempatkan mereka pada risiko fisik. Pada tahun 2010, 14 pekerja melompat keluar jendela di sebuah pabrik elektronik Foxconn di Shenzhen, Tiongkok. Kami berpendapat bahwa tidak ada pekerjaan yang seharusnya begitu menghancurkan jiwa hingga mendorong manusia untuk bunuh diri, dan kita mungkin akan lebih baik jika pekerjaan-pekerjaan ini diambil alih oleh mesin yang tidak memiliki jiwa untuk dihancurkan.

Lagipula, sejarah penuh dengan kisah-kisah tentang pekerjaan manual yang digantikan oleh mesin. Kita bisa dengan mudah berargumen bahwa revolusi teknologi di masa lalu, seperti mesin uap atau mesin fotokopi, pada akhirnya memicu pertumbuhan ekonomi dan meningkatkan standar hidup banyak orang di planet ini. Namun, argumen-argumen yang tidak masuk akal ini tampaknya tidak meyakinkan para pekerja yang pekerjaannya dipertaruhkan. AI konon akan menggantikan beberapa pekerjaan tetapi menciptakan beberapa pekerjaan baru. Hal ini benar, dan bukan hanya untuk posisi berkeahlian tinggi seperti ilmuwan data dan insinyur ML. Bahkan, tren baru yang sedang berkembang adalah melatih orang untuk melakukan tugas pelabelan data untuk tugas ML. Misalnya, untuk setiap perusahaan mobil swakemudi dengan ratusan insinyur di Silicon Valley, ribuan pelabel manusia memeriksa rekaman dan membuat set data pelatihan. Namun, kita harus jujur: kecil kemungkinan pekerjaan baru ini akan sebanyak atau bergaji tinggi seperti yang telah tergantikan oleh AI.

Pendapat kami adalah kita perlu melihat industri lain yang dapat menyerap pekerjaan yang hilang di industri lain yang lebih rentan terhadap otomatisasi. Secara spesifik, perawatan kesehatan dan pendidikan adalah dua industri penting yang biayanya terus meningkat dalam

dekade terakhir, bahkan ketika menghadapi peningkatan digitalisasi dan proses yang lebih efisien. Misalnya, menurut Federal Reserve Bank of St. Louis, biaya kuliah di universitas meningkat hampir delapan kali lebih cepat daripada upah.

Situasinya tampaknya tidak lebih baik di sisi perawatan kesehatan: *Journal of the American Medical Association* (JAMA) melaporkan bahwa pengeluaran perawatan kesehatan di Amerika Serikat meningkat hampir satu triliun dolar dari tahun 1996 hingga 2015. Harapan kami adalah bahwa pekerjaan yang hilang karena otomatisasi di industri seperti transportasi dan manufaktur dapat diserap oleh mereka yang membutuhkan "sentuhan manusia" dan tidak akan pernah cocok untuk otomatisasi penuh. Namun, harapan kami hanyalah harapan yang terpelajar karena dinamika di balik fenomena ini sangat kompleks.

Kami memulai percakapan ini dengan menekankan bahwa AI akan menggantikan tugas, dan bukan pekerjaan. Sebagian besar pekerjaan akan terdampak secara lebih halus daripada tergantikan sepenuhnya. Kami menantikan pertumbuhan pekerjaan yang didukung AI atau dibantu AI: profesi di mana AI membantu orang melakukan pekerjaan mereka lebih cepat, lebih baik, atau lebih menyenangkan.

Kami melihat dua contoh hal ini dalam studi kasus Translated di bab 5. Perusahaan ini menggunakan AI untuk membantu manajer proyek memilih penerjemah terbaik untuk tugas penerjemahan dengan memeringkat ribuan resume berdasarkan persyaratan pekerjaan tertentu. Perusahaan ini juga menawarkan kesempatan pertama bagi penerjemah profesional untuk mencoba menerjemahkan, sehingga mereka bebas untuk mengoreksinya dan memfokuskan upaya mereka untuk memilih kata yang tepat. Menurut Marco Trombetti, salah satu pendiri Translated, teknologi ini sangat dihargai oleh para penerjemah:

Teknologi kami menciptakan banyak peluang, tidak hanya bagi kami, tetapi juga bagi para penerjemah, yang akhirnya tidak lagi bekerja mengoreksi hal-hal yang sama dan membosankan, tetapi menghabiskan waktu mereka untuk menjadi lebih manusiawi, mengomunikasikan makna teks yang sebenarnya, dan menjadi lebih persuasif dan kreatif.

Kenyataannya, saat ini, AI sedang meningkatkan cara kita bekerja, alih-alih membuang manusia ke jalanan. Mudah membayangkan hal terakhir akan terjadi dalam waktu dekat, tetapi kami telah memberikan lebih dari satu argumen untuk mendukung gagasan bahwa kenyataan tidaklah sejelas itu.

Gelembung Filter AI

Platform daring seperti YouTube dan Facebook telah menjadi media utama untuk menerima berita dan komentar tentang dunia. Mengingat jumlah konten yang tersedia di masing-masing platform ini pada dasarnya tak terbatas, algoritma merekalah yang menentukan apa yang kita lihat dan apa yang tidak kita lihat. Dalam arti tertentu, sistem rekomendasi telah mengambil peran sebagai penjaga utama informasi. Bayangkan saja: Seberapa besar kemungkinan Anda akan melihat video di halaman ke-10 hasil YouTube? Baik atau buruk, algoritmalah yang menentukan konten mana yang boleh kita konsumsi.

Faktanya, personalisasi bisa menjadi begitu ekstrem sehingga konsumen media terjebak dalam gelembung filter, di mana semua konten yang mereka konsumsi dipilih sesuai dengan keyakinan dan pandangan dunia mereka. Istilah ini dicetuskan oleh aktivis internet Eli Pariser pada tahun 2010 ketika ia menyadari bahwa umpan media sosial teman-temannya menampilkan konten pro-Demokrat atau pro-Republik, tetapi jarang keduanya. Secara umum, gelembung filter menggambarkan sistem rekomendasi yang telah disetel secara akurat dengan preferensi pengguna sehingga mereka berhenti menampilkan konten kontrafaktual sama sekali.

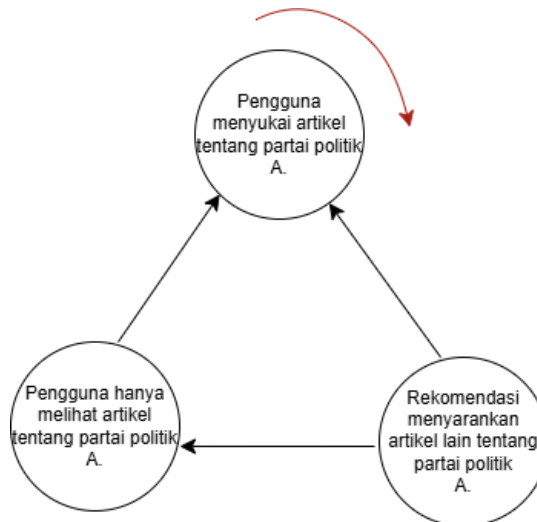
Bayangkan bergabung dengan beberapa grup Bumi Datar di Facebook, dan melihat umpan berita Anda terisi dengan konten terkait Bumi Datar dan komentar-komentar yang antusias. Tidak akan lama sebelum otak kita yang rentan bias mulai membesar-besarkan jumlah penganut Bumi Datar di komunitas tersebut. Faktanya, idiom umum lainnya untuk gelembung filter adalah ruang gema, karena konten dan interaksi sosial yang direkomendasikan secara otomatis cenderung memperkuat pola pikir yang ada, alih-alih menantang Anda dengan pandangan yang berlawanan.

Hal serupa terjadi dengan YouTube, yang mesin rekomendasinya dituduh memupuk ekstremisme dengan berfokus pada video yang semakin ekstrem. *The New York Times* terkenal melaporkan bahwa mulai menonton video tentang vegetarianisme akan menghasilkan rekomendasi tentang veganisme, seperti halnya video tentang jogging akan menghasilkan rekomendasi tentang lari maraton. Yang lebih mengkhawatirkan, dinamika yang sama juga berlaku untuk ranah yang lebih sensitif, seperti misinformasi atau berita palsu tentang politik, hak asasi manusia, atau kekerasan. Dan memang, banyak komunitas telah mulai menyuarakan keprihatinan mereka kepada YouTube (dan pemerintah), menuntut pengawasan yang lebih ketat terhadap apa yang terjadi di situs tersebut.

Dari sudut pandang teknis, Anda harus menyadari bahwa tidak satu pun dari masalah ini yang intrinsik terhadap teknologi itu sendiri. Banyak algoritma rekomendasi didasarkan pada kesamaan, baik di berbagai item maupun pengguna lain dalam komunitas. Tidak ada tombol "ujaran kebencian" yang dimodifikasi oleh para insinyur YouTube untuk meningkatkan keterlibatan di situs. Sebaliknya, dinamika ekstremis ini merupakan konsekuensi langsung dari tujuan bisnis yang mendorong adopsi sistem rekomendasi sejak awal. Algoritma ini dirancang untuk menyediakan konten yang kemungkinan besar akan kita sukai, dan suka atau tidak, kita ingin melihat hal-hal yang kita setujui.

Ketika minat kita mulai bergeser ke satu topik, rekomendasi selanjutnya akan mulai berfokus pada satu topik tersebut, yang mendorong kita untuk mengonsumsi lebih banyak konten tersebut, yang mengarah pada rekomendasi yang lebih spesifik, dan seterusnya. Gambar 10.1 menggambarkan lingkaran setan ini. Sistem rekomendasi tidak memiliki kemampuan abstraksi untuk memahami bahwa merekomendasikan gigi palsu kepada pelanggan berusia dua puluhan tahun adalah ide bisnis yang bodoh, tetapi menunjukkan kedua sisi debat politik sangat penting bagi masyarakat yang sehat. Sebaliknya, mereka memperlakukan yang terakhir seperti yang pertama, menarik orang ke dalam gelembung filter yang semakin tebal.

Setelah pemilihan presiden 2016 di Amerika Serikat dan referendum Brexit di Inggris (juga pada tahun 2016), platform media sosial utama telah memperkuat upaya mereka melawan radikalisasi dan ekstremisme. Dan solusinya mungkin menambahkan lebih banyak AI ke dalam campuran. Mengandalkan manusia untuk mengawasi perilaku kasar dan tidak etis tidak benar-benar layak pada skala operasi besar-besaran jaringan media sosial utama saat ini. Sebaliknya, pemrosesan bahasa alami dan model AI visual dapat digunakan untuk memproses dan memblokir konten terlarang secara otomatis.



Gambar 10.1 Sistem rekomendasi memiliki begitu banyak kekuatan dalam menentukan konsumsi media harian kita sehingga dapat menciptakan gelembung ekstremisme di media sosial.

Masalah inti masih tetap sama: untuk menyediakan informasi yang sehat bagi publik, Anda perlu orang-orang membaca apa yang mereka setuju dan apa yang sebaliknya. Sayangnya, menampilkan konten yang tidak mereka setuju tidak baik untuk bisnis, jadi masalahnya masih terbuka.

Ketika AI Gagal: Kasus-Kasus Kecil dan Serangan Musuh

Ketika maskapai penerbangan pertama dibuka untuk transportasi sipil setelah Perang Dunia II, pesawat terbang tidak dapat diandalkan dan, sejujurnya, merupakan pesawat yang berbahaya. Dengan pengalaman terbatas dengan pesawat militer (yang masa pakainya tidak terlalu lama karena alasan yang jelas), para insinyur tidak memiliki pemahaman yang baik tentang bagaimana pesawat akan gagal selama operasi: mesin yang tidak berfungsi dengan baik, retakan di badan pesawat, dan sebagainya.

Selama beberapa dekade berikutnya, mereka mengumpulkan cukup pengetahuan, teori, dan pengalaman untuk menjadikan perjalanan udara sebagai bentuk transportasi yang sangat aman yang kita nikmati saat ini. Seiring umat manusia mulai mengandalkan AI untuk situasi hidup-mati, seperti perawatan kesehatan atau transportasi otonom, pemahaman kita tentang batasannya perlu berkembang. Bagian ini akan mengajarkan Anda tentang dua mode kegagalan penting AI: kasus-kasus kritis dan serangan adversarial.

AI masa kini, yang berbasis pembelajaran mesin, mendapatkan kekuatannya dari kemampuannya mempelajari korelasi statistik yang dangkal dari sejumlah besar data. Sayangnya, tidak ada algoritma atau sistem yang dikenal saat ini yang menunjukkan kemampuan generalisasi dan inferensi logis yang dimiliki manusia. Dengan kata lain, dapat dikatakan bahwa algoritma AI sangat memperhatikan detail, tetapi "bodoh" dalam hal menarik kesimpulan yang jelas dari pengetahuan mereka atau menangani situasi yang tidak lazim.

Ambil contoh autopilot Tesla, yang dapat berkendara dengan mulus sejauh ratusan mil di jalan raya, namun mengalami kecelakaan pada tahun 2017 karena mengira truk yang dicat tipis sebagai sepetak langit biru dan langsung menerobosnya. Tidak ada manusia yang akan salah mengira gambar langit di atas truk sebagai langit yang sebenarnya, namun salah satu teknologi tercanggih saat ini masih melakukannya. Ini adalah kasus-kasus khusus: situasi langka yang sulit ditemukan dalam data pelatihan dan yang dapat diinterpretasikan oleh model dengan cara yang bodoh.

Sebenarnya tidak banyak yang bisa kita lakukan untuk kasus-kasus khusus selain mengumpulkan dataset yang lebih luas dan lengkap, serta mengembangkan model yang lebih robust. Namun, dalam hal bencana yang disebabkan oleh model AI, kasus-kasus khusus bukanlah satu-satunya masalah yang harus kita waspadai. Ada risiko lain yang lebih halus: serangan adversarial.

Ternyata para peneliti yang giat telah menemukan cara untuk "membingungkan" jaringan saraf, dan menghasilkan gangguan visual (lihat gambar 10.2) yang hampir 100% salah diklasifikasikan oleh jaringan saraf. Dengan kata lain, mereka telah menemukan cara untuk menghasilkan percikan warna yang tampak acak yang secara konsisten diklasifikasikan oleh model sebagai objek nyata, seperti mobil atau pizza. Kita bisa membayangkan menempelkan cetakan gambar 10.2 untuk melumpuhkan mobil-mobil self-driving di area tersebut. Lebih halus lagi, gambar-gambar yang tampak tidak berbahaya dapat dimasukkan ke dalam konten daring, sehingga memaksanya diklasifikasikan sebagai konten berbahaya.

Penelitian tentang cara membuat dan mempertahankan diri dari serangan yang bersifat merugikan masih terus berlangsung. Kasus-kasus kecil dan serangan yang bersifat merugikan adalah dua sisi mata uang yang sama, di mana kita tidak dapat mempercayai keluaran model pembelajaran mesin (ML) dan perlu menemukan cara untuk melindungi masyarakat dari kesalahan mereka.

Solusi untuk melindungi diri dari masalah ini adalah dengan menggunakan sistem redundan yang tidak berbasis AI. Misalnya, dalam konteks mobil swakemudi, banyak perusahaan menggunakan radar tradisional untuk memeriksa penghalang fisik di jalur mobil. Tidak masalah apakah AI mobil menafsirkan sinyal dari kamera sebagai langit biru: jika radar mendeteksi suatu objek, Anda sebaiknya berhenti. Meskipun solusi ini dapat memberikan sedikit kepastian, solusi ini tetaplah solusi sementara. Pada intinya, beberapa aplikasi AI masih belum cukup tangguh untuk diterapkan pada aplikasi yang berpotensi berbahaya.

Ketika yang Buatan Terlihat Nyata: Konten Palsu yang Dihasilkan AI

Bab 4 dan 5 dimulai dengan model AI yang dapat memahami media dan bahasa, tetapi juga menunjukkan algoritma yang dapat menciptakan keduanya misalnya, dengan mengubah

foto zebra menjadi foto kuda. Hal ini menarik, karena menyempurnakan kemampuan AI, memungkinkan kita membangun aplikasi yang kreatif, dan bukan hanya analitis. Namun, kreasi-kreasi ini kini begitu realistis dan meyakinkan sehingga menjadi senjata baru yang ampuh bagi para pelaku kejahatan yang ingin merekayasa dan menyebarkan realitas mereka sendiri.

Rezim Stalin di Uni Soviet pada tahun 1930an terkenal karena mengubah foto-foto untuk menyingkirkan lawan politik dan umumnya mendukung pandangan pemerintah yang berlaku. Pada saat itu, teknik-teknik tersebut efektif justru karena sangat sedikit orang yang tahu bahwa mengubah foto itu mungkin. Seiring Photoshop dan efek khusus dalam film menyadarkan masyarakat umum akan teknik-teknik manipulasi ini, teknik-teknik tersebut kehilangan sebagian kekuatannya. Mungkin saja satu-satunya alasan kita menganggap gambar atau video itu autentik adalah karena kita tahu bahwa dibutuhkan sejumlah upaya dan kehati-hatian untuk mengubahnya. Dengan munculnya jaringan generatif, hambatan masuk ini dengan cepat turun menjadi nol.

Dengan kata lain, ini adalah situasi lain di mana AI membawa keuntungan skala: kini begitu mudah dan murah untuk memproduksi konten palsu sehingga masyarakat perlu menghadapi risiko penggunaannya untuk tujuan yang tidak etis. Yang tidak kami sampaikan di bab 5 adalah bahwa GPT-2, model OpenAI tercanggih untuk pembuatan teks, awalnya tidak dirilis ke publik. Kekhawatiran para peneliti adalah teks yang dihasilkan oleh model tersebut begitu meyakinkan sehingga mereka khawatir akan digunakan oleh oknum jahat untuk menyebarkan konten yang tidak bermakna namun menarik (click bait) di internet. Mereka memutuskan untuk merilis versi model yang kurang canggih dan menunggu komunitas NLP untuk mengejar dan mengembangkan model yang mampu membedakan konten palsu ini.

Pemain lain tidak seserius itu. Memang, salah satu klaim awal ketenaran jaringan generatif muncul pada tahun 2017, ketika pengguna Reddit anonim Deepfakes mengunggah beberapa video porno yang menampilkan selebritas terkenal. Ternyata para bintang tersebut tidak pernah membintangi video-video tersebut. Sebaliknya, Deepfakes menggunakan jaringan generatif untuk menerapkan ciri-ciri wajah aktris terkenal ke dalam materi yang sudah ada, dengan cara menyalin dan menempel wajah dan ekspresi mereka. Deepfake berfokus pada selebritas karena ketersediaan data pelatihan (mudah menemukan ribuan foto selebritas), tetapi teknologi yang sama dapat digunakan untuk orang-orang yang kita cintai atau bahkan politisi.



Gambar 10.2 Model deteksi objek dapat mengidentifikasi Gianluca di sebelah kiri dengan benar. Namun, gambar berwarna yang dipegang Nico direkayasa secara khusus untuk membingungkan model sehingga tidak dapat mendeteksinya.

Apa solusi untuk menghindari penyebaran konten palsu? Nah, karena skala adalah masalahnya, skala juga bisa menjadi solusinya. Seiring algoritma semakin baik dalam menghasilkan teks dan gambar palsu, begitu pula algoritma yang dapat mendeteksinya. Ini adalah perlombaan senjata antara siapa yang ingin menipu kita dan siapa yang ingin melindungi kita, dan senjata dalam pertarungan ini adalah algoritma yang canggih. Ketika indra kita berhenti membedakan mana yang nyata dan mana yang tidak, satu-satunya hal yang dapat kita lakukan adalah meminta bantuan algoritma.

10.2 PELUANG AI DALAM MASYARAKAT

Dalam arti tertentu, seluruh buku ini didedikasikan untuk kekuatan dan potensi AI, jadi kami tidak akan mengulanginya di bagian ini. Sebaliknya, kami akan mengambil pandangan yang lebih luas dan berfokus pada dampak AI yang lebih luas terhadap masyarakat, alih-alih pada organisasi individual.

Demokratisasi Teknologi

Peningkatan teknologi AI baru-baru ini sebagian besar didorong oleh demokratisasi teknologi dan pengetahuan. Bab 1 telah menyajikan faktor-faktor yang mendorong perkembangan AI dalam 20 tahun terakhir: ketersediaan data dalam jumlah besar dan daya komputasi yang murah.

Selain itu, penting untuk menyebutkan peran yang dimainkan oleh riset terbuka dan perangkat lunak sumber terbuka dalam beberapa tahun terakhir. Komunitas AI sangat terbuka, dan seperti yang telah Anda lihat di bab 8, dimungkinkan untuk menemukan riset mutakhir yang dipublikasikan secara gratis di platform seperti arXiv (<https://arxiv.org>). Riset terbuka merupakan katalisator yang luar biasa bagi perkembangan AI. Bayangkan diri Anda sebagai peneliti yang perlu memecahkan masalah ML yang rumit: mereka bisa mendapatkan akses gratis ke sejumlah besar makalah penelitian yang diterbitkan oleh tim AI terkemuka. Seringkali,

makalah ini dilengkapi dengan kode dan data sehingga setiap orang dapat mencoba algoritmanya sendiri dan mengadaptasinya (atau menyempurnakannya) agar sesuai dengan kebutuhan mereka.

Sebagaimana penelitian terbuka memungkinkan setiap orang memiliki akses gratis ke pengetahuan, perangkat sumber terbuka memungkinkan akses gratis ke perangkat canggih. Perusahaan teknologi besar seperti Google dan Facebook berlomba-lomba membangun kerangka kerja ML mereka sendiri: pustaka perangkat lunak yang memungkinkan pengembang membangun algoritma ML dengan cepat dan efisien.

Organisasi-organisasi ini sangat tertarik untuk menjadikan kerangka kerja mereka sendiri sebagai standar, agar lebih mudah menarik bakat. Google, misalnya, telah berinvestasi besar dalam memasarkan kerangka kerja ML-nya sendiri yang disebut TensorFlow, sedemikian rupa sehingga kita sering mendengar para eksekutif bisnis menyebutnya sebagai teknologi yang "wajib dimiliki" dalam tumpukan mereka. Jika ketersediaan data, daya komputasi yang murah, serta riset dan perangkat sumber terbuka telah menjadi faktor pendorong AI dalam 20 tahun terakhir, apa yang akan mendorong AI dalam 20 tahun mendatang?

Lonjakan ketersediaan data dan daya komputasi belum sepenuhnya hilang: keduanya akan terus mendorong inovasi AI di tahun-tahun mendatang. Namun, kita sudah mulai melihat contoh-contoh awal perusahaan dan kelompok riset yang mundur dari keterbukaan mereka.

Di bagian sebelumnya, kami menyebutkan OpenAI menunda rilis model pembangkit bahasanya karena kekhawatiran tentang aktor jahat. Sekalipun rilis riset terbuka dapat melambat, faktor-faktor lain ikut berperan. Salah satu hambatan AI selalu adalah kelangkaan talenta: sulit menemukan orang dengan keahlian yang tepat untuk bekerja di bidang AI. Untungnya, semakin banyak kursus AI daring berkualitas tinggi yang tersedia secara gratis telah membantu puluhan ribu orang menguasai AI. Tren ini akan terus berlanjut dan menyuntikkan talenta yang sangat dibutuhkan ke dalam dunia AI.

Kami juga antusias dengan perkembangan perangkat ML untuk orang-orang nonteknis. Misalnya, Google telah memperkenalkan alat bernama Cloud AutoML untuk memperluas jangkauan pelanggan yang dapat memanfaatkan algoritma ML. Alat ini menawarkan pengalaman pengguna yang lebih sederhana berkat otomatisasi tugas-tugas tersulit dalam pengembangan model ML. Bayangkan seperti ini: sebelum Microsoft Excel hadir, satu-satunya cara Anda dapat menggunakan komputer adalah dengan mempelajari bahasa pemrograman dan menulis kode. Dengan Excel, tiba-tiba lebih banyak orang memiliki keterampilan untuk membuat buku kerja untuk akuntansi, pelacakan waktu, atau ribuan kegunaan lainnya. Hal yang sama terjadi dengan AI dan produk seperti Cloud AutoML.

Tidak sulit membayangkan masa depan ketika Anda dapat membangun seluruh produk ML tanpa harus menulis baris kode apa pun. Inti dari buku ini adalah bahwa tugas Anda adalah membayangkan produk-produk baru berbasis AI dan membiarkan ilmuwan data mengambil alih, tetapi Anda mungkin akan segera diberdayakan untuk mengurus seluruh proses dengan alat yang lebih baik dan lebih sederhana. Jika kita sampai di sana, hambatan untuk masuk ke proyek AI pada dasarnya akan menjadi nol, dan satu-satunya hambatan adalah kreativitas dan inspirasi Anda untuk proyek-proyek baru.

Skala Masif

Salah satu janji AI yang paling menarik adalah bahwa ia akan mendorong biaya marjinal pengetahuan dan keahlian mendekati nol, membuka produk dan layanan bagi khalayak yang sebelumnya tidak mampu membelinya. Harganya begitu murah sehingga Google bahkan menerjemahkan halaman web secara gratis kapan pun Anda mau. Sebagaimana ponsel pintar seharga Rp. 100.000 memungkinkan satu miliar orang mengakses internet untuk pertama kalinya, bayangkan apa yang dapat dilakukan oleh dokter, guru, atau pengacara berbasis AI seharga Rp. 10.000 untuk dunia. Seperti yang telah Anda lihat di bagian kedua buku ini, ketika investasi dalam pengumpulan data dan pengembangan model telah dilakukan, biaya variabel untuk terus menjalankannya sangat rendah (pada dasarnya hanya daya komputasi, yang semakin murah).

Jika Google mendemokratisasi pengetahuan dan pencarian, dan menyediakannya secara gratis bagi setiap manusia yang terhubung internet di planet ini, teknologi AI yang lebih maju menjanjikan hal yang sama dengan wawasan dan kecerdasan. Banyak perekonomian dunia terhambat oleh kurangnya pengetahuan dan tenaga kerja terampil, yang tidak dapat dengan mudah diekspor atau dilatih. Bayangkan anak-anak prasekolah di negara-negara miskin memiliki akses ke perangkat berbasis AI yang dapat mengajari mereka membaca dan menulis ketika tidak ada cukup guru terampil untuk semua orang. Bagaimana dengan algoritma kontrol industri yang dapat beradaptasi sendiri yang memungkinkan industrialisasi di daerah tertinggal dengan kelangkaan pekerja lini terampil?

Potensi skalabilitas juga tidak terbatas pada negara-negara berkembang. Di sebagian besar negara di dunia, banyak produk dan layanan berbasis pengetahuan memiliki jangkauan terbatas karena struktur harganya. Berapa banyak pasien kanker yang tidak didiagnosis dan diobati cukup dini karena mereka hanya mengunjungi dokter setahun sekali? Bagaimana dengan memberi mereka pemeriksaan awal berbasis AI setiap hari (atau bahkan setiap jam, menggunakan sensor yang dapat dikenakan)? Berapa banyak lagi nyawa yang bisa kita selamatkan jika setiap orang memiliki akses ke layanan semacam itu?

Siklus kebajikan AI berkaitan erat dengan konsep skalabilitas. Produk berbasis AI lebih murah, sehingga dapat diterapkan dan dijual dalam skala yang lebih besar, sehingga menghasilkan akumulasi lebih banyak data yang dapat meningkatkan kualitas produk, sehingga lebih menarik bagi pengguna.

10.3 PELUANG AI DALAM INDUSTRI

Di bagian ini, kami menyoroti industri-industri yang kami yakini memiliki potensi paling kuat untuk perubahan dan disrupsi oleh AI dalam waktu dekat. Tentu saja, daftar ini tidak lengkap: daftar ini dirancang untuk memberi Anda gambaran tentang apa yang terjadi di berbagai sektor, dan beberapa elemen untuk mengembangkan pemikiran kritis Anda sendiri. Kami mendorong Anda untuk tidak menganggap bagian ini sebagai kebenaran mutlak, tetapi menggunakannya sebagai titik awal yang dapat Anda kembangkan berkat latar belakang dan pengetahuan AI baru Anda.

Di bab 7, kami mencoba menemukan kriteria objektif untuk menilai potensi sebuah proyek AI. Kami menyimpulkan bahwa potensi tersebut berkorelasi langsung dengan seberapa besar potensi tersebut memanfaatkan dua kekuatan utama AI: skala dan akurasi. Kita dapat dengan mudah mengadaptasi kriteria yang sama untuk seluruh industri. Bagaimanapun, disrupsi industri dapat dilihat sebagai efek gabungan dari banyak proyek AI dari satu atau lebih perusahaan.

Dalam konteks proyek AI, akurasi adalah ukuran seberapa besar manfaat yang akan diperoleh bisnis Anda dari peningkatan akurasi tugas tertentu. Untuk menerapkan faktor akurasi ke seluruh industri, perluas maknanya ke rangkaian proses inti yang membentuk industri tersebut. Misalnya, sektor yang sebagian besar telah terdigitalisasi dan sudah efisien kemungkinan besar hanya akan memperoleh sedikit manfaat dari peningkatan akurasi yang dihasilkan oleh AI. Di sisi lain, industri yang masih dicirikan oleh teknologi yang tidak efisien dan tidak efektif dapat sangat dipengaruhi oleh AI.

Hal yang sama berlaku untuk skala. Ketika berbicara tentang proyek AI individual, faktor skala mengukur dampak positif yang dihasilkan ketika suatu tugas diotomatisasi melalui AI. Hal yang sama berlaku sekarang karena kita menganalisis seluruh industri. Jika suatu industri sering beroperasi melalui proses "kasar" yang dilakukan secara manual, terdapat potensi tinggi terjadinya disrupsi oleh AI.

Dalam konteks proyek AI, kami juga mengevaluasi dan memprioritaskan proyek berdasarkan kesiapan organisasi. Konsep yang sama berlaku untuk seluruh industri. Seperti yang Anda ketahui, AI modern sangat bergantung pada keberadaan kumpulan data besar untuk melatih model ML. Jika suatu industri masih didukung oleh proses analog (alias pena dan kertas), industri tersebut harus melalui proses digitalisasi yang sulit sebelum AI dapat memberikan dampak positif. Di sisi lain, industri yang telah berinvestasi dalam digitalisasi dan manajemen data dapat segera mulai menerapkan AI. Hati-hati: ini bukan berarti industri dengan kesiapan teknologi yang rendah memiliki potensi AI yang rendah. Kesiapan teknologi yang rendah menunjukkan bahwa disrupsi akan membutuhkan waktu lebih lama untuk terjadi.

Jejaring Media Sosial

Bisa dibilang, kita dapat mengatakan bahwa jejaring media sosial adalah Big Bang di balik lonjakan investasi AI saat ini. Dengan menghubungkan miliaran orang ke internet, jejaring ini memungkinkan para peneliti mengumpulkan kumpulan data besar yang diperlukan untuk pengembangan model AI canggih. Populasi pengguna yang besar ini juga menyediakan sumber daya yang besar untuk eksperimen berisiko rendah dengan produk-produk berbasis AI baru, seperti pengenalan wajah otomatis dan sistem rekomendasi.

Pengguna media sosial menghasilkan sebagian besar dari semua jenis data yang kita bahas dalam buku ini: mulai dari data bisnis inti tentang iklan, churn pengguna, dan aktivasi, hingga kumpulan data teks dan gambar yang sangat besar. Kita dapat memisahkan dua kategori utama aplikasi:

- ❖ Untuk memenuhi tujuan utama platform media sosial: memungkinkan pengiklan menjangkau pengguna yang tepat
- ❖ Untuk menyediakan layanan

Bab 5 telah membahas analisis sentimen, salah satu alat paling dasar dalam kotak peralatan NLP. Media sosial menyediakan platform yang ideal untuk menerapkan analisis sentimen dalam skala besar, karena interaksi dengan masyarakat umum memiliki nilai yang sangat besar bagi perusahaan. Namun, ini sudah lama. Platform sosial masa kini menggunakan algoritma AI untuk memodelkan perilaku dan minat pengguna guna mempersonalisasi pesan pemasaran dan mengoptimalkan kampanye iklan. Dalam konteks ini, akurasi AI akan terus mendorong efektivitas rekomendasi konten dan penempatan iklan.

Kita telah membahas ancaman personalisasi dalam skala besar sebelumnya di bab ini. Dalam konteks ini, kita ingin berfokus pada peluang yang ditawarkan oleh potensi skala AI. Aplikasi pertama adalah pencegahan dan deteksi penyalahgunaan otomatis: Facebook dan raksasa media sosial lainnya telah menerapkan model yang dapat mendeteksi konten yang tidak etis dan kasar serta mencegahnya dibagikan di jaringan mereka.

Misalnya, setiap jaringan sosial besar menggunakan AI untuk secara otomatis mendeteksi materi kekerasan atau pornografi sebelum diunggah ke platform dan disajikan kepada miliaran penggunanya. Bayangkan dunia tanpa AI, di mana setiap kali Anda mengunggah gambar atau video, seseorang yang duduk di kantor harus memeriksa apakah itu pantas. Jaringan sosial tidak akan ada (atau mungkin ada, tetapi akan dipenuhi dengan segala macam kemerosotan manusia). Algoritma visi komputer memungkinkan pemeriksaan semacam ini dalam skala besar, dan mereka bekerja sangat baik dalam menjaga keamanan komunitas; misalnya, sangat jarang menemukan materi kekerasan atau pornografi di Facebook.

Meskipun personalisasi dan moderasi konten sudah ketinggalan zaman, kami beralih ke seorang pakar untuk membayangkan hubungan masa depan antara AI dan media sosial. Kami bertanya kepada Luca La Mesa, seorang pemasar media sosial di Italia yang telah bekerja dengan banyak organisasi terkemuka, tentang tren yang ia perkirakan akan datang. Ia menyoroti kisah positif, yaitu fitur Pemeriksaan Keamanan Facebook yang menggunakan model NLP untuk memindai pesan dan mendeteksi tanda-tanda awal kecenderungan bunuh diri. Para pakar manusia secara manual meninjau kasus-kasus yang telah ditandai oleh algoritma, dan memberi tahu pihak berwenang jika diperlukan.

Namun, ia juga mengutip episode yang lebih menyeramkan, seperti yang dilaporkan oleh Gillian Brockell dari Washington Post. Saat hamil, ia mengunggah postingan di media sosial menggunakan tagar seperti #30week-spregnant atau #babybump. Algoritme mendeteksi postingannya dan menargetkannya dengan iklan produk bayi. Sayangnya, Brockell kehilangan anaknya, tetapi bahkan saat ia mencari di Google dengan istilah seperti "kontraksi Braxton Hicks" atau "bayi tidak bergerak", ia terus menjadi target iklan produk bayi yang hanya menimbulkan kesedihan. Satu-satunya cara untuk menghentikannya adalah dengan menulis surat terbuka ke Twitter.

Dari apa yang dikatakan La Mesa kepada kami, tampaknya AI selalu membawa keunggulan skalanya lebih dalam ke dalam interaksi manusia. Jika memblokir materi kekerasan adalah fitur "wajib" yang kita semua harapkan dari platform ini, Pemeriksaan Keamanan Facebook adalah tambahan yang indah yang mengejutkan banyak orang. Tidak seorang pun

menyanka platform media sosial mampu menyediakan layanan seperti itu, dan satu-satunya alasan hal itu mungkin adalah karena AI secara dramatis meningkatkan kemampuannya dalam memproses data tekstual. Contoh negatif yang disebutkan La Mesa sekali lagi menjadi bukti bahwa AI tidak cocok untuk kasus-kasus khusus. Semoga, kedalaman pemahaman algoritma ini terhadap teks akan terus meningkat, dan akan selalu ada semakin sedikit kasus seperti yang dialami Brockell.

Layanan Kesehatan

Layanan kesehatan merupakan sektor yang sangat kompleks dengan banyak tantangan spesifik, beragam pemangku kepentingan, dan lingkungan regulasi yang kompleks. Dalam industri yang begitu luas, akan sangat membantu jika aplikasi-aplikasi tersebut dipecah menjadi beberapa bagian. AI akan memengaruhi layanan kesehatan dalam tiga bidang utama:

- Diagnostik
- Penemuan Obat
- Efisiensi Operasional dan Model Perawatan Baru

Tabel 10.1 mencantumkan perbandingan ketiga bidang tersebut dalam hal akurasi, skala, dan kesiapan teknologi.

Tabel 10.1 Analisis Potensi dan Kesiapan Teknologi untuk Aplikasi AI dalam Layanan Kesehatan

	Faktor akurasi	Faktor skala	Kesiapan teknologi
Diagnosa	Tinggi	Tinggi	Rendah
Penemuan obat	Tinggi	Tinggi	Meningkat secara eksponensial
Perawatan berbasis nilai	Tinggi	Tinggi	Rendah

Menggunakan AI untuk meningkatkan cara dokter mendiagnosis penyakit mungkin merupakan salah satu aplikasi AI yang paling diminati. Dunia memiliki persediaan dokter yang terbatas, dan mereka semua manusia. Seperti semua manusia, dokter mengandalkan contoh terbatas yang telah mereka lihat dalam hidup mereka, data yang samar, dan banyak firasat.

Coba pikirkan ini: setiap kali Netflix merekomendasikan film kepada Anda, ia menggunakan jutaan titik data dan teknologi canggih yang telah disempurnakan selama bertahun-tahun oleh ilmuwan data kelas dunia. Setiap kali dokter Anda meresepkan terapi, keputusan mereka didasarkan pada pelatihan bertahun-tahun, mungkin ratusan orang yang telah mereka tangani dengan gejala serupa, dan kumpulan data yang sangat terbatas (dengan asumsi Anda tidak menghabiskan waktu berminggu-minggu dikurung di klinik untuk melakukan tes medis). Netflix jauh lebih canggih dalam merekomendasikan film daripada dokter Anda dalam merekomendasikan obat. Jelas bahwa pendekatan yang lebih berbasis data untuk diagnosis dan resep dapat sangat membantu.

Menggunakan AI untuk meningkatkan diagnosis jelas memiliki faktor akurasi yang tinggi. Kita semua sepakat bahwa jika kita memiliki kesempatan untuk meningkatkan akurasinya, umat manusia akan mendapatkan manfaatnya. Aplikasi spesifik ini juga dapat

memiliki faktor skala besar. Meskipun AI kemungkinan besar akan digunakan untuk mendukung dokter, bukan untuk menggantikan mereka, beberapa wilayah di dunia tidak memiliki dokter yang siap membantu. Dalam kasus ini, dokter AI dapat sangat membantu dalam menyediakan layanan kesehatan sederhana dalam skala besar.

Hal ini juga berlaku untuk negara-negara terkaya di dunia, tetapi untuk penyakit yang umumnya tidak ditangani oleh dokter. Contohnya adalah kondisi dermatologis. Mari kita ambil kasus jerawat: seperti yang dilaporkan oleh perusahaan rintisan MD Algorithms, 90% orang dengan jerawat tidak pernah menemui dokter kulit. Ini adalah peluang yang dimanfaatkan perusahaan dengan membangun algoritma visi komputer yang dapat mendiagnosis jenis jerawat yang dimiliki pasien berdasarkan swafoto.

Sayangnya, akses ke data mungkin merupakan faktor utama yang menghambat aplikasi ini. Sebagian besar proses perawatan kesehatan masih analog (berapa kali Anda melihat seorang dokter masuk ke rumah sakit sambil memegang laporan kertas dan pena?). Meskipun dorongan baru-baru ini menuju rekam medis elektronik (RME) telah sedikit memperbaiki situasi, industri masih harus menghadapi dampak dari jalur digitalisasi.

Selain itu, data yang tersedia masih terfragmentasi, karena berbagai pemangku kepentingan, termasuk rumah sakit, pembayar, layanan kesehatan nasional, dan perusahaan farmasi, masing-masing hanya memiliki sebagian kecil dari riwayat klinis pasien. Dengan kata lain, terdapat potensi yang sangat besar, tetapi akan sulit untuk mengajak semua pihak berkolaborasi.

Area besar kedua untuk disrupsi dalam perawatan kesehatan adalah penemuan obat. Pemicu utama disrupsi dalam kasus ini adalah pengamatan bahwa biaya untuk menemukan obat baru meningkat secara eksponensial. Biaya litbang industri farmasi telah meroket hampir seratus kali lipat antara tahun 1950 dan 2010.

Di sisi lain, biaya untuk pengurutan DNA menurun secara eksponensial. Pengurutan genom manusia utuh pertama menghabiskan biaya sekitar Rp. 27 triliun pada tahun 2003. Pada tahun 2006, biayanya turun menjadi Rp. 3 miliar. Biaya pada tahun 2016 sekitar Rp. 10 juta, dan masih belum menunjukkan tanda-tanda akan berhenti menurun. Ini berarti kini dimungkinkan untuk menghasilkan data genom dalam jumlah besar dengan biaya murah.

Ketersediaan data biologis yang luas merupakan langkah pertama menuju kemampuan AI untuk mengintip genom manusia dan membantu merancang obat-obatan baru. Baik perusahaan rintisan kecil maupun organisasi besar ikut serta dalam hal ini, seringkali berkolaborasi bersama. Pada tahun 2018, Pfizer menjalin kemitraan strategis dengan perusahaan rintisan AI yang didukung oleh raksasa teknologi seperti Tencent dan Google bernama XtalPi. Tujuannya adalah untuk memprediksi sifat farmasi molekul kecil dan mengembangkan "desain obat rasional berbasis komputasi." Pfizer tidak sendirian: organisasi besar lainnya seperti Novartis, GSK, Sanofi, Amgen, dan Merck telah mengumumkan kemitraan mereka sendiri dengan perusahaan rintisan AI. Tujuannya adalah untuk menggabungkan pengetahuan, data, dan sumber daya mereka dengan keterampilan AI dari perusahaan rintisan yang lebih kecil untuk menemukan kandidat obat baru untuk berbagai penyakit.

Penemuan obat yang dibantu AI masih dalam tahap awal, tetapi hasilnya sangat menjanjikan. Akurasi dan faktor skalanya sangat tinggi, karena kemampuan untuk mengembangkan obat yang lebih efektif dengan biaya lebih murah dan lebih cepat akan berdampak besar pada kesehatan masyarakat. Kesiapan teknologi juga meningkat pesat, sebagaimana ditunjukkan oleh penurunan biaya pengurutan DNA yang konstan.

Terakhir, AI juga dapat menangani sisi bisnis layanan kesehatan, bukan hanya sisi pengobatan. Layanan kesehatan terkenal sebagai sektor yang tidak efisien, dengan banyak pemangku kepentingan yang memiliki insentif dan praktik berbeda yang tidak selalu menguntungkan pasien.

Dengan populasi yang menua dan meningkatnya biaya, upaya pemerintah beralih ke model layanan berbasis nilai: sebuah sistem yang berfokus pada pasien, di mana penyedia layanan kesehatan diberi insentif untuk memberikan perawatan terbaik dengan biaya terendah. Hal ini berbeda dengan model biaya per layanan, di mana penyedia layanan dibayar secara proporsional dengan jumlah layanan yang dilakukan, sehingga menciptakan insentif yang menyebabkan pemborosan dan tidak selalu memberikan hasil terbaik bagi pasien.

Coba bayangkan seperti ini: ketika Anda menemui dokter, sering kali mereka meresepkan sejumlah tes yang akan ditanggung oleh Anda, asuransi Anda, atau layanan kesehatan nasional Anda. Seringkali, hasil tes ini tidak lebih dari sekadar mengarahkan Anda ke... lebih banyak tes. Sekarang, setiap kali Anda menemui dokter dan menjelaskan gejala Anda, kemungkinan besar orang lain di dunia ini memiliki riwayat klinis dan gejala yang sangat mirip, serta telah menjalani serangkaian tes. Jika Anda ingat pembahasan kita tentang sistem rekomendasi di bab 6, sepertinya ini merupakan aplikasi potensial untuk AI, yang dapat membantu dokter mengarahkan ke jalur diagnosis atau tes terbaik, menghemat uang, waktu, dan bisa dibilang meningkatkan kualitas perawatan pasien.

Jelas, beberapa perubahan sistem yang cukup mendalam diperlukan sebelum kita dapat menempuh jalur ini. Pergeseran ini mungkin akan terjadi di kemudian hari, bukan karena keterbatasan teknologi, melainkan karena kesiapan keseluruhan sistem. Namun, potensinya begitu tinggi sehingga kami yakin pergeseran ini akan terjadi.

Energi

Produksi dan distribusi energi sedang mengalami transformasi substansial yang didorong oleh konvergensi faktor teknologi, ekonomi, dan lingkungan. Energi berkelanjutan akhirnya mulai masuk akal secara bisnis, karena skala ekonomi menurunkan harga tanpa perlu subsidi. Pada saat yang sama, kendaraan listrik semakin meluas di banyak negara, dan sensor pintar semakin murah dan semakin sering digunakan di gedung dan pabrik.

Tabel 10.2 menyajikan faktor skala, akurasi, dan kesiapan teknologi untuk dua aplikasi utama AI di sektor energi: penyeimbangan beban dan optimalisasi konsumsi energi.

Tabel 10.2 Analisis potensi AI dan kesiapan teknologi untuk AI di sektor energi

	Perkiraan produksi energi terbarukan		
	Faktor akurasi	Faktor skala	Kesiapan teknologi
Penyeimbangan beban	Tinggi	Tinggi	Tinggi

Optimalisasi konsumsi energi	Tinggi	Tinggi	Tinggi
------------------------------	--------	--------	--------

Jika Anda belum familiar dengan cara kerja pasar energi, berikut panduan singkatnya. Energi listrik masih sulit dan mahal untuk disimpan dan ditimbun untuk penggunaan selanjutnya (meskipun teknologi baterai baru telah mulai mengubah hal ini). Ini berarti perusahaan energi harus memproduksi energi sesuai kebutuhan jaringan listrik setiap saat. Jika Anda menyalakan AC di kamar, pembangkit listrik di suatu tempat akan mulai membakar lebih banyak gas untuk memenuhi permintaan energi tambahan dari rumah Anda. Jika perusahaan energi tidak memproduksi cukup banyak, pemadaman listrik dapat terjadi. Produksi berlebih juga dapat merusak jaringan listrik. Penyeimbangan beban adalah kunci untuk menjalankan masyarakat kita: setiap saat, permintaan dan pasokan energi perlu diseimbangkan.

Perusahaan energi telah menjadi cukup andal dalam memprediksi permintaan energi, memodelkan bagaimana keluarga menggunakan TV dan microwave mereka di malam hari, atau menyalakan AC di musim panas. Mengendalikan produksi bukanlah masalah secara historis karena pembangkit listrik berbahan bakar fosil dapat diatur naik atau turun sesuai kebutuhan. Namun, energi yang dihasilkan dari sumber terbarukan jauh lebih sulit dikendalikan, atau bahkan diprediksi, karena bergantung pada faktor-faktor yang tidak menentu seperti angin atau matahari. Di saat yang sama, deregulasi pasar mendorong perusahaan energi untuk mengatasi tantangan ini tanpa berinvestasi dalam infrastruktur tambahan.

Untungnya, seiring dengan tantangan tersebut, peluang juga dihadirkan oleh teknologi baru seperti AI. Investasi dalam teknologi digital oleh perusahaan energi telah meningkat tajam selama beberapa tahun terakhir, dengan investasi global dalam infrastruktur dan perangkat lunak kelistrikan digital meningkat 20% per tahun sejak 2014, mencapai Rp. 470 triliun pada tahun 2016. Dorongan ini memungkinkan kita untuk membangun aset dasar yang dibutuhkan untuk mendukung AI apa pun: data.

Perhatikan bahwa semua yang kita bahas sebelumnya dapat dilihat sebagai tantangan prediksi/optimalisasi, alih-alih tantangan rekayasa. Misalnya, kita dapat menggunakan data cuaca untuk memprediksi energi yang akan dihasilkan oleh sumber daya terbarukan. Karena kita dapat memperoleh data dari turbin angin, panel surya, dan sebagainya bersama dengan data cuaca, kita dapat menggunakan informasi ini untuk memperkirakan produksi dari energi terbarukan dengan akurasi yang lebih tinggi dan mempermudah pekerjaan utilitas. Mengoptimalkan konsumsi energi adalah kasus penggunaan lain yang umumnya sangat cocok untuk kemampuan AI.

Di Bab 2, Anda melihat bagaimana Google mengotomatiskan operasi pusat datanya untuk menghemat energi. Hal ini dimungkinkan sebagian besar karena Google telah memasang banyak sensor dan berinvestasi dalam pengumpulan data serta otomatisasi. Seiring menyebarnya teknologi pendukung ini ke gedung, pabrik, dan infrastruktur lain, pengoptimalan yang didukung AI juga akan semakin lazim. Teknologi serupa juga dapat diterapkan untuk perumahan, yang dapat membantu keluarga mengurangi konsumsi energi.

Bidang pengoptimalan energi tidak terbatas pada pengurangan jumlah total energi yang digunakan. Perusahaan energi, dan dengan demikian lingkungan, juga diuntungkan ketika beban didistribusikan sepanjang hari sesuai dengan pola produksi sumber energi terbarukan. Misalnya, ini mungkin berarti menyiapkan kendaraan listrik untuk mengisi daya di siang hari, ketika tenaga surya berlimpah.

Di dunia di mana perubahan iklim semakin dan memang seharusnya mengkhawatirkan publik, AI dapat dan harus menjadi salah satu senjata dalam gudang senjata perusahaan energi. Dengan pola pikir yang tepat, teknologi ini dapat berfungsi sebagai lapisan yang mengendalikan pembangkitan dan produksi energi, mencapai tingkat efisiensi baru dan berkontribusi pada perjuangan kita melawan perubahan iklim.

Manufaktur

Sejak revolusi lean manufacturing di tahun 80-an, sektor manufaktur sangat bergantung pada data. Dunia kapitalis kita mendorong peningkatan dan optimalisasi yang konstan, dan Anda tidak dapat meningkatkan apa yang tidak dapat Anda ukur. Karena alasan ini, pabrik menjadi salah satu lokasi pertama di mana data dalam format digital mulai dikumpulkan untuk dianalisis oleh komputer.

Seperti yang telah Anda pelajari di buku ini, setiap kali ada banyak data, potensi proyek AI mengikutinya, karena kesiapannya tinggi dan hambatan untuk masuknya rendah. Contoh yang paling umum berkisar pada peningkatan efisiensi produksi, seperti yang Anda lihat dalam studi kasus pusat data Google. Sebuah mesin lini produksi tunggal mungkin memiliki puluhan atau ratusan sensor suhu, tekanan, dan getaran yang menghasilkan aliran data yang konstan.

Salah satu aplikasi ML tertua dan paling mapan dalam manufaktur adalah prediksi kesalahan: menggunakan data sensor untuk mendapatkan peringatan dini ketika suatu peralatan akan rusak. Upaya yang lebih maju telah membantu pengguna memahami apakah mungkin untuk memodelkan pengaruh parameter produksi utama (seperti tekanan atau kecepatan siklus) terhadap kualitas produksi. Sebagaimana dalam studi kasus pusat data Google, kembaran digital semacam itu memungkinkan para insinyur untuk menyempurnakan dan mengoptimalkan proses industri secara digital, mentransfer pengalaman yang terakumulasi ke pabrik fisik setelah mereka yakin dengan hasilnya.

Semua aplikasi yang telah kita bahas sejauh ini menggunakan data terstruktur, seperti deret waktu sensor. Namun, model AI berbasis gambar seperti yang kita bahas di bab 4 berpotensi membawa otomatisasi rantai pabrik ke tingkat berikutnya. Kandidat utama untuk evolusi ini adalah kendali mutu, sebagaimana dalam studi kasus klasifikasi mentimun di bab 4. Model klasifikasi dan deteksi objek berdasarkan jaringan dalam bersifat serbaguna dan dapat digunakan untuk menemukan oklusi atau cacat pada produk jadi hanya dengan menambahkan kamera ke lini produksi.

Algoritma AI yang lebih canggih juga akan meningkatkan robotika dan otomatisasi di lantai pabrik. Keterbatasan besar yang menghambat peningkatan penggunaan robotika dalam manufaktur adalah besarnya waktu dan upaya yang dibutuhkan untuk memprogramnya guna membangun produk baru. Jika Anda mengingat pembahasan kita di bab 7, Anda akan

menyadari bahwa tugas-tugas yang berulang dan sempit seperti ini merupakan kandidat ideal untuk solusi berbasis pembelajaran mesin.

Jelas bahwa manufaktur sangat membutuhkan teknologi yang skalabel yang memungkinkannya memproduksi barang dengan lebih murah dan lebih cepat. AI adalah alat utama untuk tujuan ini, baik kita mencoba memprediksi kesalahan maupun mengotomatiskan kendali mutu. Dari segi akurasi, kita tahu bahwa AI sangat sempit, dan tidak mungkin mencapai kemampuan manusia dalam menangani pengecualian dalam waktu dekat. Ini berarti manusia akan tetap menguasai dunia barang mewah dan barang pesanan untuk beberapa waktu. Lagipula, barang buatan tangan sudah menjadi jalan bebas menuju harga tiket yang lebih tinggi, dan kesenjangan tersebut kemungkinan akan melebar seiring produksi massal menjadi lebih terstandarisasi dan lebih murah.

Keuangan

Dalam penerapan AI di bidang keuangan, kita perlu membedakan dua dunia yang berbeda: perbankan komersial dan perdagangan. Bank komersial adalah bank yang kita semua hadapi dalam kehidupan sehari-hari: mereka mencairkan cek, menggadaikan rumah, dan sebagainya. Perusahaan perdagangan membeli dan menjual saham dan obligasi di pasar keuangan, mencoba mengalahkan investor lain untuk menghasilkan uang. Karena perbankan dan keuangan sangat bergantung pada data dan angka, keduanya merupakan pengadopsi awal teknologi pembelajaran mesin. Bab 2 menyajikan contohnya ketika kita membahas pinjaman pedagang Square.

Bank komersial telah banyak menggunakan perangkat AI untuk beberapa kompetensi inti mereka, seperti pencegahan penipuan dan penilaian risiko. Di tahun-tahun mendatang, penyempurnaan model ini akan diterapkan di area AI lainnya untuk menciptakan pasar keuangan yang lebih sehat dan dengan demikian mengurangi biaya bagi peminjam dan perusahaan. Dalam banyak aktivitas inti industri perbankan, laba berkorelasi langsung dengan akurasi pengambilan keputusan, dan oleh karena itu kesesuaian untuk AI cukup jelas. Di sisi lain, terdapat pula potensi besar untuk menggunakan AI dengan cara-cara yang benar-benar baru, memanfaatkan skalanya.

Mengotomatiskan proses internal dengan perangkat AI seperti pengenalan visual dan NLP bukanlah sesuatu yang mewah atau ide baru, tetapi akan penting untuk mengurangi biaya di seluruh organisasi. Sebagaimana halnya organisasi yang berhadapan langsung dengan pelanggan lainnya, AI memiliki potensi besar untuk mendisrupsi dan meningkatkan hubungan layanan pelanggan yang sebagian besar ditangani oleh karyawan langsung saat ini. Hal yang sama berlaku untuk penawaran pemasaran yang dikurasi dan dipersonalisasi yang dapat diciptakan oleh perangkat seperti sistem rekomendasi.

Pertimbangan yang sama juga berlaku untuk industri terkait: asuransi. Seiring model AI meningkatkan kemampuannya untuk memahami dan menyusun informasi tentang dunia, perusahaan asuransi dapat memiliki gambaran yang lebih baik tentang risiko apa yang mereka asuransikan kepada pelanggan mereka, dan dengan demikian menawarkan tarif yang lebih memadai dan personal. Bayangkan analisis otomatis data kesehatan untuk menyediakan asuransi kesehatan yang dipersonalisasi, atau intelijen waktu nyata tentang jenis jalan yang

biasa Anda lalui. Meskipun semua ide ini mungkin terwujud saat ini, AI akan menyediakan kunci rahasia penskalaan yang memungkinkan perusahaan asuransi untuk memperluasnya lebih luas ke basis pelanggan mereka.

Terakhir, perdagangan keuangan agak aneh, dan kita tidak akan menghabiskan banyak waktu untuk membicarakannya. Wall Street adalah dunia yang penuh rahasia dan terspesialisasi di mana beberapa pemikir paling kutu buku mencoba yang terbaik untuk mengekstrak informasi kepemilikan yang dapat mereka gunakan untuk menginformasikan keputusan perdagangan mereka dan "mengalahkan pasar" dengan pengembalian yang lebih unggul untuk investasi mereka. Ini berarti bahwa banyak alat AI yang telah kami jelaskan dalam buku ini dapat dengan satu atau lain cara digunakan untuk menghasilkan uang di pasar keuangan.

Contoh yang umum adalah menggunakan model klasifikasi objek visual pada data satelit untuk menghitung jumlah mobil yang meninggalkan pabrik Ford, dan dengan demikian memprediksi bagaimana kinerja produsen mobil tersebut sebelum merilis laporannya. Informasi kepemilikan ini kemudian dapat digunakan untuk bertaruh pada harga saham yang naik atau turun, mengantongi keuntungan apa pun. Ide yang sama juga berlaku untuk banyak model lain yang telah kami sajikan dalam buku ini; Misalnya, analisis sentimen di media sosial untuk memahami bagaimana persepsi pelanggan terhadap suatu perusahaan, dan dengan demikian, apakah prospek masa depannya baik.

Wall Street telah menggunakan alat-alat ini selama satu dekade atau lebih, jadi apa yang akan terjadi dalam beberapa tahun ke depan? Kita tidak tahu! Ada begitu banyak uang yang dipertaruhkan sehingga perusahaan-perusahaan perdagangan ini merahasiakan semua kemajuan AI mereka. Dua hal yang pasti: semakin akurat Anda dalam mengambil keputusan dan semakin cepat Anda dapat membuatnya, semakin banyak uang yang akan Anda hasilkan. Karena alasan-alasan ini, AI dan perdagangan akan selalu menjadi sahabat karib.

Pendidikan

Dari semua industri yang telah kita analisis sejauh ini di bagian ini, pendidikan mungkin yang paling tertinggal dalam penggunaan AI. Menurut pendapat kami, potensi paling menarik untuk persimpangan keduanya terletak pada personalisasi: kemampuan untuk menyesuaikan rencana pembelajaran setiap siswa berdasarkan kekuatan dan kelemahan mereka. Beberapa kursus daring terbuka masif (MOOC) telah menjadi pelopor di bidang ini justru karena tantangan luar biasa yang muncul dari kebutuhan untuk membimbing dan menilai puluhan ribu siswa dengan anggaran terbatas.

Ternyata umpan balik yang konstan dan personal merupakan komponen terpenting dari strategi pembelajaran yang efektif. Untungnya, banyak model AI yang telah kita bahas dalam buku ini sebenarnya sangat baik dalam mengklasifikasikan masukan sebagai "baik" atau "buruk." Melampaui sistem penilaian otomatis klasik, kita dapat membayangkan model berbasis penglihatan yang dapat secara otomatis menilai teknik melukis berdasarkan kemiripannya dengan karya seniman-seniman hebat di masa lalu. Atau bahkan model berbasis NLP yang dapat menilai kosakata para penulis muda dengan membandingkan pekerjaan rumah mereka dengan mahakarya sastra yang hebat.

Tentu saja, masih banyak pekerjaan yang harus diselesaikan sebelum model berbasis AI dapat menjadi tulang punggung sistem sekolah seperti yang kita kenal sekarang. Namun, beberapa segmen pendidikan tertentu telah memanfaatkan perangkat AI. Kita sudah memiliki guru musik otomatis yang dapat memperingatkan siswa ketika mereka bermain sumbang, atau perangkat pembelajaran bahasa yang dapat membantu menghilangkan aksen. Sektor-sektor khusus ini memanfaatkan kekuatan algoritma AI yang sangat canggih, seperti yang ada di media, dan mendorong pendidikan agar lebih ramah bagi pembelajar "coba-coba". Dengan kata lain, personalisasi utama menjadi gamifikasi, yang memberikan umpan balik instan kepada setiap siswa. Salah satu janji paling menarik tentang AI dalam pendidikan adalah bahwa AI akan membantu menyediakan satu miliar siswa berikutnya di negara berkembang dengan tingkat pendidikan yang sama dengan yang dinikmati negara maju saat ini.

10.4 AI UMUM ADALAH TOPIK YANG PENTING

Sejak istilah kecerdasan buatan muncul, orang-orang senang bertaruh tentang berapa lama waktu yang dibutuhkan sebelum AI dapat menyamai kinerja otak manusia. Seperti yang Anda ingat dari pelajaran sejarah singkat kita di bab 1, tujuan agung ini disebut AI umum, tepatnya karena ia dapat menangani tugas apa pun. Tidak mengherankan, serangkaian keberhasilan pembelajaran mendalam baru-baru ini telah memicu tebakan optimis, dengan tokoh-tokoh terkemuka di komunitas tersebut memperkirakan kita akan mencapai paritas manusia-mesin paling cepat pada tahun 2030.

Banyak buku telah ditulis tentang kecerdasan umum buatan (AGI), jadi kami tidak berani mencoba membahas topik ini secara mendalam. (Kami merekomendasikan *Superintelligence* karya Nick Bostrom, Oxford University Press, 2014.) Tujuan kami untuk bagian ini adalah untuk membuat Anda berpikir tentang peluang AGI yang paling realistis dan langsung.

Setelah membaca buku ini, kami harap Anda menyadari bahwa kita masih jauh dari keajaiban kecerdasan manusia yang serba guna dan cepat belajar. AI modern didasarkan pada pembelajaran mesin, dan pembelajaran mesin modern pada dasarnya masih berupa pencocokan pola yang luar biasa, meskipun memungkinkan kita untuk membangun produk yang berharga. Banyak orang percaya (dan kami cenderung setuju) bahwa teknologi yang digunakan untuk membangun aplikasi AI sempit yang dijelaskan dalam buku ini tidak cocok untuk AGI.

Pendapat ini bahkan dianut oleh beberapa pelopor AI. Misalnya, Geoffrey Hinton, salah satu pelopor pembelajaran mendalam, mengaku dalam sebuah wawancara bahwa ia "sangat curiga" terhadap prinsip-prinsip dasar yang saat ini memungkinkan jaringan saraf untuk belajar, dan tidak yakin prinsip-prinsip tersebut akan pernah mencapai skala kecerdasan sejati. Judea Pearl, pelopor AI lainnya, berpendapat bahwa pembelajaran mesin saat ini terjebak dalam kurva yang pas, tanpa memahami arti dunia di sekitar kita. Dengan kata lain, kita masih sibuk mengajari komputer "apa" yang terjadi di dunia, tetapi kita bahkan belum mulai membahas "mengapa".

Pengetahuan yang Anda terima dari buku ini kemungkinan besar membawa Anda pada kesimpulan yang sama tentang masa depan AGI. Bagaimanapun, kami rasa Anda tidak perlu putus asa. Kami tentu saja tidak. Banyak sekali aplikasi dunia nyata yang dapat dibangun dengan alat ML dan AI yang sudah kita ketahui cara penggunaannya.

Kami menghubungi Denis Rothman, pakar AI dan penulis tiga buku sukses, dan menanyakan pendapatnya tentang AGI. Kami sangat berterima kasih:

Kecerdasan buatan sebagai semacam kesadaran super bukan hanya fatamorgana, untuk saat ini, tetapi yang terpenting, AGI tidak berguna.

Otomatisasi, dengan agen AI atau tanpanya, akan terus meningkat secara eksponensial di tahun-tahun mendatang. Agen AI yang berspesialisasi dalam berbagai domain akan memberikan peningkatan kekuatan pada IoT.

Jika kita menggabungkan ini sekarang, kita mendapatkan jaringan agen otomatis yang terhubung dan disempurnakan dengan AI yang akan mengubah dunia seperti sekarang ini. Ini tidak lebih baik atau lebih buruk daripada mobil, kereta api, pesawat terbang, dan telepon pertama.

Sekarang, beberapa nama terkemuka masih memperingatkan kita tentang bahaya AI superintelen. Salah satunya adalah Elon Musk, yang sering memperingatkan kita tentang bahaya AI super-manusia, bahkan berspekulasi tentang potensinya untuk mengakhiri peradaban. Kita masih sangat awal dalam pengembangan AI sehingga sulit untuk meramalkan apakah skenario apokaliptik ini akan benar-benar terjadi. Untuk saat ini, bahayanya tampaknya terfokus pada perubahan struktural masyarakat dan hilangnya pekerjaan, seperti yang telah kita bahas sebelumnya. Anda bisa bersantai sejenak: Terminator masih belum terlihat.

10.5 PIKIRAN PENUTUP

Tujuan kami menulis buku ini adalah memberi Anda pengetahuan dan perangkat yang Anda butuhkan untuk menghadirkan manfaat AI ke dalam organisasi Anda, bahkan tanpa keahlian teknis. Kami bekerja keras untuk membimbing Anda menjadi seorang pemimpin yang dapat memulai dan mengarahkan diskusi tentang menemukan area paling menjanjikan untuk penggunaan AI di organisasi Anda. Di bagian pertama buku ini, Anda mempelajari informasi yang lebih teknis tentang perangkat AI: berbagai jenis algoritma untuk bekerja dengan jenis data yang mungkin Anda temui di organisasi saat ini. Studi kasus dan contoh industri yang tersebar di seluruh materi telah menunjukkan kepada Anda bagaimana para pemimpin teknologi berpikir dan merencanakan tujuan AI mereka. Di bagian kedua buku ini, Anda belajar cara bekerja keras. Kami menunjukkan kepada Anda cara menemukan dan memahami peluang paling menjanjikan untuk menggunakan AI dalam bisnis Anda, merekrut tim yang tepat, dan memulai implementasinya.



Beberapa dari Anda sudah memiliki rencana untuk sebuah proyek AI saat membaca buku ini. Bagi Anda, kami berharap buku ini telah memberikan substansi pada rencana Anda dengan mengajarkan kemampuan AI masa kini secara nyata tanpa perlu dibesar-besarkan. Bagi Anda yang belum memiliki proyek, kami harap kami telah menginspirasi Anda dengan informasi terbaru dari industri dan akademisi, dan rencana Anda kini lebih konkret. Apa pun hasilnya, harapan terbaik kami adalah Anda merasa percaya diri dan bersemangat untuk kembali bergabung dengan tim, organisasi, atau startup Anda, dan mulai mengerjakan proyek AI pertama Anda. Semoga sukses selalu.

DAFTAR PUSTAKA

- Agarwal, A. (2023). AI adoption by human resource management: a study of its antecedents and impact on HR system effectiveness. *foresight*, 25(1), 67-81.
- Ahmad, S., Priyadharshini, S. L., Hosen, M. S., Ng, A., Islam, S., & Manik, J. A. (2024). The Impact Of Artificial Intelligence On Business & Social Values: Benefits, Challenges, And Future Directions. *Educational Administration: Theory and Practice*, 30(6), 3174-3180.
- Alhosani, K., & Alhashmi, S. M. (2024). Opportunities, challenges, and benefits of AI innovation in government services: a review. *Discover Artificial Intelligence*, 4(1), 18.
- Ali, M. H., Hamdan, A., & Alareeni, B. (2020, November). The implementation of artificial intelligence in organizations' systems: opportunities and challenges. In *International Conference on Business and Technology* (pp. 153-163). Cham: Springer International Publishing.
- Amado, D. P. A., Diaz, F. A. C., Pantoja, R. D. P. C., & Sanchez, L. M. B. (2023). Benefits of Artificial Intelligence and its innovation in organizations. *Multidisciplinar (Montevideo)*, (1), 8.
- Anaya, L., Braizat, A., & Al-Ani, R. (2024). Implementing AI-based Chatbot: Benefits and Challenges. *Procedia Computer Science*, 239, 1173-1179.
- Balidemaj, V. (2024). *Artificial Intelligence Integration, Concerns, Benefits, and the Need for Ethical Policies for Community Foundations and Nonprofit Organizations* (Doctoral dissertation, Clemson University).
- Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2024). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. *Journal of organizational behavior*, 45(2), 159-182.
- Calderon, R. (2019). The benefits of artificial intelligence in cybersecurity.
- Chowdhury, R. H. (2024). AI-driven business analytics for operational efficiency. *World Journal of Advanced Engineering Technology and Sciences*, 12(2), 535-543.
- Chukwunweike, J. N., Yussuf, M., Okusi, O., & Oluwatobi, T. (2024). The role of deep learning in ensuring privacy integrity and security: Applications in AI-driven cybersecurity solutions. *World Journal of Advanced Research and Reviews*, 23(2), 2550.
- Detyna, B., & Sylwestrzak, P. (2025). Different aspects of using artificial intelligence in organizations—challenges, benefits and good practices. *SZTUCZNA INTELIGENCJA I AUTOMATYZACJA PROCESÓW BIZNESOWYCH*, 127.
- Elufioye, O. A., Ike, C. U., Odeyemi, O., Usman, F. O., & Mhlongo, N. Z. (2024). Ai-Driven predictive analytics in agricultural supply chains: a review: assessing the benefits and

- challenges of ai in forecasting demand and optimizing supply in agriculture. *Computer Science & IT Research Journal*, 5(2), 473-497.
- Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-powered organization. *Harvard business review*, 97(4), 62-73.
- Frenette, J. (2022). Benefits of the rise of human-centric AI in enterprise decision making.
- Ganapathy, V. (2024). AI-Based Risk Assessments in Forensic Auditing: Benefits, Challenges and Future Implications. *Shodh Sari-An International Multidisciplinary Journal*, 3(04), 100-128.
- Gökçearslan, S., Tosun, C., & Erdemir, Z. G. (2024). Benefits, challenges, and methods of artificial intelligence (AI) chatbots in education: A systematic literature review. *International Journal of Technology in Education*, 7(1), 19-39.
- Granjeiro, J. M., Cury, A. A. D. B., Cury, J. A., Bueno, M., Sousa-Neto, M. D., & Estrela, C. (2025). The future of scientific writing: AI tools, benefits, and ethical implications. *Brazilian Dental Journal*, 36, e25-6471.
- Hanson, B., Stall, S., Cutcher-Gershenfeld, J., Vrouwenvelder, K., Wirz, C., Rao, Y., & Peng, G. (2023). Garbage in, garbage out: mitigating risks and maximizing benefits of AI in research. *Nature*, 623(7985), 28-31.
- iu Zaman, B. U. (2024). Transforming education through AI benefits risks and ethical considerations.
- Jang, K., & Landuyt, N. G. (2023). Limited benefits of technological advances in human service organizations: Going beyond the hype using sociotechnical knowledge management system. *Journal of Social Service Research*, 49(4), 426-446.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, 1(9), 389-399.
- Johnson, R. D., Stone, D. L., & Lukaszewski, K. M. (2021). The benefits of eHRM and AI for talent acquisition. *Journal of Tourism Futures*, 7(1), 40-52.
- Kim, J. K., Xiong, Y., & Hunt, D. S. (2025). Communicating the benefits and risks of AI technology in hiring: implications for organization–public relationships. *Corporate Communications: An International Journal*.
- Kirova, M., & Boneva, M. (2024). Artificial intelligence: challenges and benefits for business.
- Köchling, A., & Wehner, M. C. (2023). Better explaining the benefits why AI? Analyzing the impact of explaining the benefits of AI-supported selection on applicant responses. *International Journal of Selection and Assessment*, 31(1), 45-62.
- Kruhse-Lehtonen, U., & Hofmann, D. (2020). How to define and execute your data and AI strategy. *Harvard Data Science Review*, 2(3).

- Kumar, A., Bhattacharyya, S. S., & Krishnamoorthy, B. (2023). Automation-augmentation paradox in organizational artificial intelligence technology deployment capabilities; an empirical investigation for achieving simultaneous economic and social benefits. *Journal of Enterprise Information Management*, 36(6), 1556-1582.
- Lazić, L. J. U. B. O. M. I. R. (2019, January). Benefit from Ai in cybersecurity. In *Proc. 11th Int. Conf. Bus. Inf. Secur.(BISEC)* (pp. 103-119).
- Manthena, S. R. L. (2021). Impact of Artificial Intelligence on Recruitment and its Benefits. *International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences*, 9(5), 1-1.
- Mari, A. (2019). The Rise of machine learning in marketing: goal, process, and benefit of AI-driven marketing.
- Mazingue, C. (2023). Perceived challenges and benefits of AI implementation in customer relationship management systems. *Journal of Digitovation and information system*, 3(1), 72-98.
- Mellouli, S., Janssen, M., & Ojo, A. (2024). Introduction to the issue on artificial intelligence in the public sector: risks and benefits of AI for governments. *Digital Government: Research and Practice*, 5(1), 1-6.
- Mir, A. I. (2024). Application of AI in talent management: a systematic review of benefits, challenges, and prospects. *Academy of Education and Social Sciences Review*, 4(4), 612-630.
- Mohammed, A. (2023). Elevating Cybersecurity Audits: How AI is Shaping Compliance and Threat Detection. *Aitoz Multidisciplinary Review*, 2(1), 35-43.
- Mori, Y., East, J. E., Hassan, C., Halvorsen, N., Berzin, T. M., Byrne, M., ... & Rex, D. K. (2023). Benefits and challenges in implementation of artificial intelligence in colonoscopy: World Endoscopy Organization position statement. *Digestive Endoscopy*, 35(4), 422-429.
- Mukherjee, A. N. (2022). Application of artificial intelligence: benefits and limitations for human potential and labor-intensive economy—an empirical investigation into pandemic ridden Indian industry. *Management Matters*, 19(2), 149-166.
- Mwita, K. M., & Kitole, F. A. (2025). Potential benefits and challenges of artificial intelligence in human resource management in public institutions. *Discover Global Society*, 3(1), 35.
- Painoli, G. K., Rao, D. V. M., Chowdary, N., & Kalgi, N. (2021). Artificial Intelligence In Business- Benefits And Challenges. *Turkish Online Journal of Qualitative Inquiry*, 12(5).
- Patel, K. (2024). Ethical reflections on data-centric AI: balancing benefits and risks. *Available at SSRN 4993089*.
- Pi, Y. (2021). Machine learning in governments: Benefits, challenges and future directions. *JeDEM-eJournal of eDemocracy and Open Government*, 13(1), 203-219.

- Ransbotham, S., Candelon, F., Kiron, D., LaFountain, B., & Khodabandeh, S. (2021). The cultural benefits of artificial intelligence in the Enterprise. *MIT Sloan Management Review and Boston Consulting Group*, 1.
- Rosário, C. (2025). Benefits and risks of using AI in Human Resource Management: Literature review. *Educational Research (IJMCER)*, 7(2), 146-154.
- Sakib, M. N., Salehin, M., Younus, M., Al-Omari, M. A., Sahabuddin, M., & Tabash, M. I. (2024). The ChatGPT and the future of HR: A critical review on the benefits and challenges of AI chatbots in human resource management. *Multidisciplinary Reviews*, 7(8), 2024136-2024136.
- Šakytė-Statnickė, G., & Budrytė-Ausiejienė, L. (2025). Application of artificial intelligence in the tourism sector: Benefits and challenges of AI-based digital tools in tourism organizations of Lithuania, Latvia, and Sweden. *Tourism and Hospitality*, 6(2), 67.
- Sanyaolu, E., & Atsaboghena, R. (2022). Role of Artificial Intelligence in Human Resource Management: Overview of its benefits and challenges. *ResearchGate*, (December), 1-8.
- Sharma, R., & Reddy, B. (2024, August). Strategic Approaches to AI Integration in Organizations: Managing Change and Maximizing Benefits. In *2024 4th Asian Conference on Innovation in Technology (ASIANCON)* (pp. 1-5). IEEE.
- Taherdoost, H., & Madanchian, M. (2023). Artificial intelligence and knowledge management: Impacts, benefits, and implementation. *Computers*, 12(4), 72.
- Tajabadi, M., Grabenhenrich, L., Ribeiro, A., Leyer, M., & Heider, D. (2023). Sharing data with shared benefits: artificial intelligence perspective. *Journal of Medical Internet Research*, 25, e47540.
- Talaei, J., Yang, A., Takishova, T., & Masialeti, M. (2024). How Does Cost Leadership Strategy Suppress the Performance Benefits of Explainability of AI Applications in Organizations?. *Journal of Global Information Management (JGIM)*, 32(1), 1-23.
- Toll, D., Lindgren, I., Melin, U., & Madsen, C. Ø. (2020). Values, benefits, considerations and risks of AI in government: A study of AI policies in Sweden. *JeDEM-EJournal of EDemocracy and Open Government*, 12(1), 40-60.
- Tosun, A., Bener, A., & Kale, R. (2010, July). Ai-based software defect predictors: Applications and benefits in a case study. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 24, No. 2, pp. 1748-1755).
- Tveita, L. J., & Hustad, E. (2025). Benefits and Challenges of Artificial Intelligence in Public sector: A Literature Review. *Procedia Computer Science*, 256, 222-229.
- Van Esch, P., Black, J. S., & Ferolie, J. (2019). Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior*, 90, 215-222.
- Wamba-Taguimdje, S. L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence (AI) on firm performance: the business value



of AI-based transformation projects. *Business process management journal*, 26(7), 1893-1924.

World Health Organization. (2024). *Benefits and risks of using artificial intelligence for pharmaceutical development and delivery*. World Health Organization.

Manfaat AI dalam Organisasi

Dr. Ir. 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 978-634-7227-57-7 (PDF)



9

786347

227577