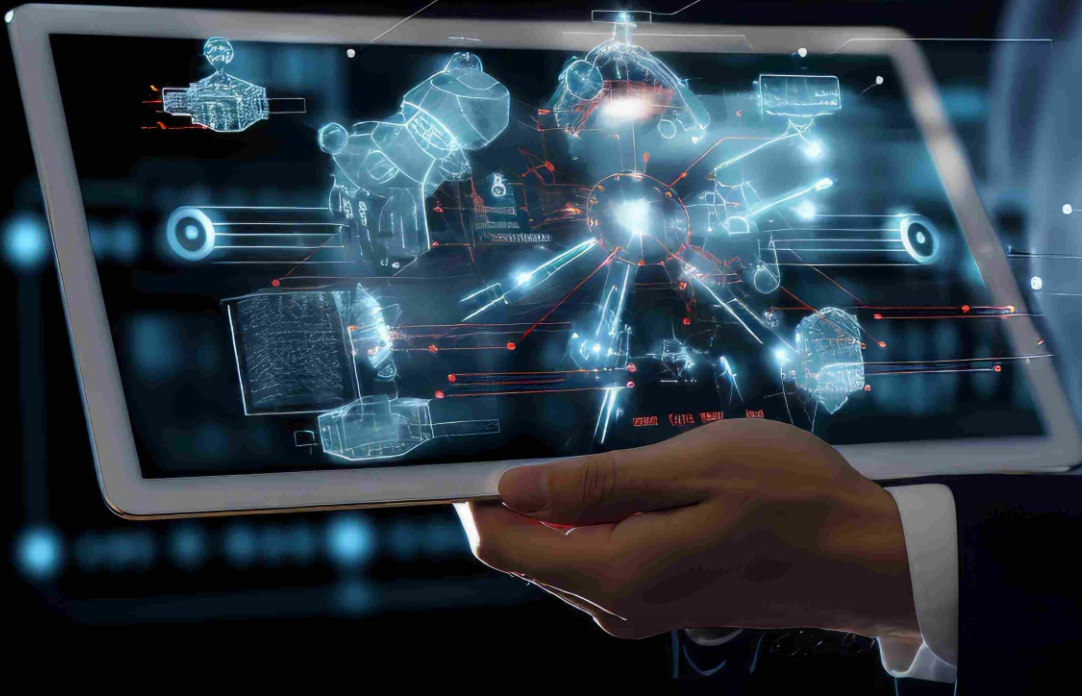




YAYASAN PRIMA AGUS TEKNIK



# TRANSFORMASI AI (Artificial Intelligence)

PADA

# PERUSAHAAN

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





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

# TRANSFORMASI AI (Artificial Intelligence) PADA PERUSAHAAN



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 978-634-7227-41-6 (PDF)



9

786347

227416

# **TRANSFORMASI AI (Artificial Intelligence) PADA PERUSAHAAN**

**Penulis :**

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

**ISBN : 978-634-7227-41-6 (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 Yuniato, 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](mailto:penerbit_ypat@stekom.ac.id)

**Distributor Tunggal :**

**Universitas STEKOM**

Jl. Majapahit no 605 Semarang

Telp. 08122925000

Fax. 024-6710144

Email : [info@stekom.ac.id](mailto: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 kami panjatkan kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga buku berjudul "**TRANSFORMASI AI (Artificial Intelligence) PADA PERUSAHAAN**" ini dapat diselesaikan dengan baik. Buku ini disusun untuk memberikan wawasan mendalam mengenai perkembangan, penerapan, serta dampak kecerdasan buatan (AI) dalam transformasi bisnis di berbagai sektor perusahaan.

Perkembangan teknologi AI yang pesat telah membawa perubahan signifikan dalam cara perusahaan beroperasi, berinteraksi dengan pelanggan, dan mengambil keputusan strategis. Buku ini mengulas secara komprehensif mulai dari dasar-dasar AI, siklus hidup AI, hingga aplikasi implementasi AI di bidang e-commerce, layanan keuangan, manufaktur, energi, dan kesehatan. Selain itu, buku ini membahas strategi pengembangan AI serta tantangan yang dihadapi perusahaan dalam mengadopsi teknologi ini.

Bab 1 Menguraikan dampak empat revolusi industri dan mengklarifikasi mitos serta realitas seputar AI. Membahas siklus kebajikan data dan algoritma serta alasan mengapa revolusi AI saat ini sangat berkelanjutan dan penting sebagai keunggulan kompetitif perusahaan. Bab 2 Memperkenalkan perkembangan AI sempit dengan fokus pada jaringan saraf awal dan pembelajaran mesin. Mengulas metode pembelajaran terawasi, tanpa pengawasan, dan semi-terawasi serta bagaimana data bisa dimanfaatkan optimal melalui penalaran semantik dan penerapan AI di berbagai bidang. Bab 3 Membahas penerapan AI dalam periklanan digital, pemasaran, akuisisi pelanggan, serta strategi cross-selling dan up-selling. Memaparkan keterlibatan pelanggan dengan asisten digital dan optimalisasi penetapan harga serta rantai pasokan.

Bab 4 Menguraikan pemanfaatan AI dalam anti pencucian uang, penilaian risiko kredit, layanan prediktif, perdagangan algoritmik, riset investasi, dan otomatisasi operasi bisnis di sektor keuangan. Bab 5 Menggambarkan AI untuk optimasi operasi pabrik, pemeliharaan aset, siklus hidup produksi, manajemen rantai pasokan, peramalan energi, serta produksi minyak dan perdagangan energi.

Bab 6 Membahas aplikasi AI dalam penemuan obat farmasi, uji klinis, diagnosis penyakit, persiapan perawatan paliatif, dan perawatan di rumah sakit. Bab 7 Menjelaskan tujuan sistem kecerdasan terhubung dan tantangan implementasi AI. Menguraikan komponen strategi AI dan langkah untuk mengembangkannya termasuk perakitan elemen strategis dan bagaimana melangkah maju.

Bab 8 Memaparkan definisi kasus penggunaan, pengumpulan serta penilaian data, rekayasa fitur, pelatihan model, manajemen model, sampai tahap pengujian, penyebaran, dan aktivasi model AI. Bab 9 Menjelaskan perbedaan antara platform AI dan aplikasi AI, arsitektur platform AI, dan pertimbangan penting dalam merancang arsitektur tersebut. Bab 10 Mengulas potensi kesalahan algoritma, mitigasi risiko model, dan peran kantor risiko model dalam menjaga keandalan AI. Bab 11 Membicarakan penyelarasan pemangku kepentingan, pengorganisasian untuk skala, pembentukan pusat keunggulan AI, penyusunan tim proyek, dan pengelolaan talenta serta literasi data.

Bab 12 Menguraikan detail arsitektur platform AI, pola teknis yang digunakan, serta kesimpulan teknis terkait pengembangan AI. Bab 13 Menjelaskan tahapan mendefinisikan kasus penggunaan AI, pemilihan dan persiapan data, analisis data eksploratori, rekayasa fitur, dan pemilihan model optimal. Bab 14 Membahas dampak AI terhadap masyarakat dan

pekerjaan, isu regulasi data dan AI, perkembangan teknologi AI ke depan, serta penegasan bahwa ini adalah awal dari era AI yang lebih luas.

Setiap bab dirancang untuk memberikan pemahaman holistik dan mendalam mengenai berbagai aspek transformasi AI yang relevan bagi perusahaan agar dapat memaksimalkan potensi dan menghadapi tantangan di era digital ini.

Diharapkan buku ini dapat menjadi referensi berguna bagi para praktisi, akademisi, dan siapa saja yang tertarik memahami bagaimana AI mampu menjadi keunggulan kompetitif dalam menghadapi era digital yang terus berkembang. Kami menyadari bahwa proses penulisan buku ini tidak lepas dari dukungan banyak pihak, oleh karenanya kami ucapkan terima kasih kepada semua yang telah membantu baik secara langsung maupun tidak langsung.

Semoga buku ini dapat memberikan inspirasi dan pemahaman yang bermanfaat untuk mendorong transformasi digital berbasis AI di perusahaan-perusahaan di Indonesia dan seluruh dunia.

*Selamat dan Semangat Membaca...!!!*

Jakarta, September 2025  
Penulis

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

*Melampaui Batas Manusia, Menyongsong Era Kecerdasan Buatan —  
Mengubah Tantangan Menjadi Peluang Melalui Transformasi AI.  
Buku ini kami dedikasikan bagi para pemimpi dan pelaku transformasi  
yang berani menghadapi perubahan dengan hati terbuka dan pikiran  
inovatif. Semoga setiap halaman membuka cakrawala baru,  
membangkitkan semangat, dan memberi keberanian untuk  
menjadikan kecerdasan buatan bukan hanya teknologi, tapi kekuatan  
penggerak perubahan positif dalam dunia usaha dan kehidupan.  
Menginspirasi Perubahan, Menggerakkan Inovasi, dan Mewujudkan  
Visi Perusahaan Masa Depan.*

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

# DAFTAR ISI

<b>KATA PENGANTAR.....</b>	<b>ii</b>
<b>DAFTAR ISI.....</b>	<b>v</b>
<b>BAB 1 SEBUAH REVOLUSI YANG SEDANG TERJADI .....</b>	<b>1</b>
1.1 Dampak Empat Revolusi.....	1
1.2 Mitos Dan Realitas AI.....	3
1.3 Siklus Kebajikan Data Dan Algoritma.....	3
1.4 Revolusi Yang Berkelanjutan – Mengapa Sekarang? .....	4
1.5 AI: Keunggulan Kompetitif Anda .....	8
<b>BAB 2 APA ITU AI DAN BAGAIMANA CARA KERJANYA? .....</b>	<b>10</b>
2.1 Perkembangan AI Sempit .....	10
2.2 Jaringan Saraf Pertama .....	12
2.3 Pembelajaran Mesin.....	12
2.4 Pembelajaran Terawasi, Tanpa Pengawasan, Dan Semi-Terawasi.....	18
2.5 Membuat Data Lebih Bermanfaat.....	21
2.6 Penalaran Semantik.....	23
2.7 Penerapan AI .....	26
<b>BAB 3 AI DALAM E-COMMERCE DAN RITEL .....</b>	<b>28</b>
3.1 Periklanan Digital.....	28
3.2 Pemasaran Dan Akuisisi Pelanggan .....	29
3.3 Cross-Selling, Up-Selling, Dan Loyalitas.....	33
3.4 Kecerdasan Pelanggan Bisnis-Ke-Bisnis .....	35
3.5 Penetapan Harga Dinamis Dan Optimalisasi Rantai Pasokan.....	36
3.6 Asisten Digital Dan Keterlibatan Pelanggan .....	38
<b>BAB 4 AI DALAM LAYANAN KEUANGAN .....</b>	<b>42</b>
4.1 Anti Pencucian Uang.....	43
4.2 Pinjaman Dan Risiko Kredit.....	44
4.3 Layanan Dan Saran Prediktif.....	45
4.4 Perdagangan Algoritmik Dan Otonom.....	47
4.5 Riset Investasi Dan Wawasan Pasar.....	49
4.6 Operasi Bisnis Otomatis.....	52
<b>BAB 5 AI DALAM MANUFAKTUR DAN ENERGI .....</b>	<b>54</b>
5.1 Operasi Pabrik Dan Pemeliharaan Aset Yang Dioptimalkan .....	56
5.2 Siklus Hidup Produksi Otomatis .....	58
5.3 Optimalisasi Rantai Pasokan.....	58
5.4 Manajemen Inventaris Dan Logistik Distribusi .....	60
5.5 Peramalan Tenaga Listrik Dan Respons Permintaan .....	61
5.6 Produksi Minyak .....	62

5.7	Perdagangan Energi .....	64
<b>BAB 6</b>	<b>AI DALAM LAYANAN KESEHATAN .....</b>	<b>67</b>
6.1	Penemuan Obat Farmasi .....	67
6.2	Uji Klinis .....	68
6.3	Diagnosis Penyakit .....	69
6.4	Persiapan Untuk Perawatan Paliatif .....	71
6.5	Perawatan Di Rumah Sakit .....	72
<b>BAB 7</b>	<b>MENGEMBANGKAN STRATEGI AI .....</b>	<b>75</b>
7.1	Tujuan Sistem Kecerdasan Terhubung .....	75
7.2	Tantangan Implementasi AI .....	77
7.3	Komponen Strategi AI .....	80
7.4	Langkah-Langkah Untuk Mengembangkan Strategi AI .....	81
7.5	Diperlukan Beberapa Perakitan .....	82
7.6	Melangkah Maju .....	86
<b>BAB 8</b>	<b>SIKLUS HIDUP AI .....</b>	<b>87</b>
8.1	Mendefinisikan Kasus Penggunaan .....	88
8.2	Mengumpulkan, Menilai, Dan Memulihkan Data .....	92
8.3	Rekayasa Alam .....	95
8.4	Memilih Dan Melatih Model .....	98
8.5	Mengelola Model .....	104
8.6	Menguji, Menyebarkan, Dan Mengaktifkan Model .....	107
8.7	Kesimpulan .....	110
<b>BAB 9</b>	<b>MEMBANGUN MESIN AI YANG SEMPURNA .....</b>	<b>111</b>
9.1	Platform Ai Versus Aplikasi Ai .....	111
9.2	Apa Yang Seharusnya Dilakukan Arsitektur Platform Ai .....	111
9.3	Beberapa Pertimbangan Penting .....	116
9.4	Arsitektur Platform Ai .....	122
<b>BAB 10</b>	<b>MENGELOLA RISIKO MODEL .....</b>	<b>126</b>
10.1	Ketika Algoritma Berjalan Salah .....	127
10.2	Mitigasi Risiko Model .....	128
10.3	Kantor Risiko Model .....	137
<b>BAB 11</b>	<b>MENGAKTIFKAN KAPABILITAS ORGANISASI .....</b>	<b>139</b>
11.1	Menyelaraskan Pemangku Kepentingan .....	139
11.2	Mengorganisir Untuk Skala .....	140
11.3	Pusat Keunggulan Ai .....	142
11.4	Menyusun Tim Untuk Eksekusi Proyek .....	145
11.5	Mengelola Bakat Dan Perekrutan .....	147
11.6	Literasi Data, Eksperimen, Dan Keputusan Berbasis Data .....	150
11.7	Kesimpulan .....	151
<b>BAB 12</b>	<b>ARSITEKTUR DAN POLA TEKNIS .....</b>	<b>152</b>
12.1	Arsitektur Platform Ai .....	152

12.2	Pola Teknis .....	157
12.3	Kesimpulan .....	166
<b>BAB 13</b>	<b>PROSES PEMODELAN AI .....</b>	<b>167</b>
13.1	Mendefinisikan Kasus Penggunaan Dan Tugas AI .....	167
13.2	Memilih Data Yang Dibutuhkan .....	169
13.3	Menyiapkan Lingkungan Notebook Dan Mengimpor Data .....	170
13.4	Membersihkan Dan Mempersiapkan Data .....	171
13.5	Memahami Data Menggunakan Analisis Data Eksploratori .....	173
13.6	Rekayasa Fitur .....	177
13.7	Membuat Dan Memilih Model Yang Optimal .....	179
<b>BAB 14</b>	<b>MASA DEPAN MASYARAKAT, PEKERJAAN, DAN AI .....</b>	<b>187</b>
14.1	AI Dan Masa Depan Masyarakat .....	188
14.2	AI Dan Masa Depan Pekerjaan .....	189
14.3	Mengatur Data Dan Kecerdasan Buatan .....	190
14.4	Masa Depan AI: Meningkatkan Teknologi AI .....	193
14.5	Dan Ini Baru Permulaan .....	198
<b>DAFTAR PUSTAKA</b>	<b>.....</b>	<b>199</b>

# **BAB 1**

## **SEBUAH REVOLUSI YANG SEDANG TERJADI**

Sejak tahun 1940-an, terobosan teknologi yang dramatis tidak hanya menjadikan komputer bagian penting dan ada di mana-mana dalam kehidupan kita, tetapi juga memungkinkan pengembangan AI modern – bahkan, tak terelakkan. Di sekitar kita, AI digunakan dengan cara yang secara fundamental memengaruhi cara kita berfungsi. AI memiliki kekuatan untuk menghemat banyak uang, waktu, dan bahkan nyawa. AI kemungkinan besar akan berdampak besar pada interaksi setiap perusahaan dengan pelanggannya. Strategi AI yang efektif telah menjadi prioritas utama bagi sebagian besar bisnis di seluruh dunia.

Asisten pribadi digital yang sukses seperti Siri dan Alexa telah mendorong perusahaan untuk menghadirkan asisten yang diaktifkan suara ke semua aspek kehidupan kita, mulai dari lampu jalan hingga lemari es. Perusahaan telah membangun aplikasi AI dengan beragam variasi dan dampak, mulai dari alat yang membantu mengatur foto secara otomatis hingga terobosan penelitian genomik berbasis AI yang telah menghasilkan terapi gen individual. AI menjadi begitu penting sehingga Forum Ekonomi Dunia<sup>1</sup> menyebutnya sebagai revolusi industri keempat.

### **1.1 DAMPAK EMPAT REVOLUSI**

Tiga revolusi industri pertama memiliki dampak yang jauh melampaui lingkungan kerja. Revolusi-revolusi tersebut membentuk kembali di mana dan bagaimana kita hidup, bagaimana kita bekerja, dan sebagian besar, bagaimana kita berpikir. Forum Ekonomi Dunia telah mengusulkan bahwa revolusi keempat tidak akan kalah berdampak.

Selama revolusi industri pertama pada abad kedelapan belas dan kesembilan belas, pabrik menggantikan produsen rumahan untuk segala hal mulai dari pakaian hingga kereta, menciptakan awal mula hierarki organisasi. Mesin uap digunakan untuk memperluas pabrik-pabrik ini, memulai proses urbanisasi massal, yang menyebabkan sebagian besar orang beralih dari gaya hidup agraris dan pedesaan ke gaya hidup industrial dan perkotaan.

Dari akhir abad kesembilan belas hingga awal abad kedua puluh, revolusi industri kedua adalah periode di mana industri-industri yang sudah ada sebelumnya tumbuh secara dramatis, dengan pabrik-pabrik beralih ke tenaga listrik untuk meningkatkan produksi massal. Kebangkitan industri baja dan minyak pada masa ini juga membantu meningkatkan urbanisasi dan transportasi, dengan minyak menggantikan batu bara untuk angkatan laut dunia dan pelayaran global.

Revolusi industri ketiga, juga disebut sebagai revolusi digital, lahir ketika teknologi beralih dari analog dan mekanis ke digital dan elektronik. Transisi ini dimulai pada tahun 1950-an dan masih berlangsung. Teknologi baru mencakup mainframe dan komputer pribadi, internet, dan telepon pintar. Revolusi digital mendorong otomatisasi manufaktur, penciptaan komunikasi massa, dan peningkatan skala industri jasa global.

Pergeseran penekanan dari teknologi informasi (TI) standar ke kecerdasan buatan kemungkinan akan memiliki dampak yang lebih signifikan terhadap masyarakat. Revolusi keempat ini mencakup perpaduan teknologi yang mengaburkan batas antara ranah fisik, digital, dan biologis dan ditandai dengan terobosan di bidang-bidang seperti robotika, AI, blockchain, nanoteknologi, komputasi kuantum, bioteknologi, Internet of Things (IoT), pencetakan 3D, dan kendaraan otonom, serta inovasi kombinatorial yang menggabungkan berbagai teknologi ini menjadi solusi bisnis yang canggih. Seperti listrik dan TI, AI dianggap sebagai teknologi serba guna – teknologi yang dapat diterapkan secara luas dalam berbagai situasi yang pada akhirnya akan memengaruhi seluruh perekonomian.

Dalam bukunya *The Fourth Industrial Revolution*, pendiri dan ketua eksekutif World Economic Forum, Klaus Schwab, mengatakan, “Dari sekian banyak tantangan beragam dan menarik yang kita hadapi saat ini, yang paling intens dan penting adalah bagaimana memahami dan membentuk revolusi teknologi baru, yang tidak lain adalah transformasi umat manusia. Dalam skala, cakupan, dan kompleksitasnya, apa yang saya anggap sebagai revolusi industri keempat tidak seperti apa pun yang pernah dialami umat manusia sebelumnya.” Revolusi keempat ini menciptakan paradigma baru yang siap mengubah cara kita hidup dan bekerja secara dramatis, mengubah segalanya, mulai dari reservasi restoran hingga menjelajahi tepian alam semesta.

Hal ini juga menyebabkan perubahan signifikan dalam cara kita berbisnis. Perubahan selama 10 tahun terakhir telah membuat perubahan ini tak terelakkan. Perusahaan perlu proaktif agar tetap kompetitif; perusahaan yang tidak proaktif akan menghadapi rintangan yang lebih besar daripada sebelumnya. Dan segala sesuatunya terjadi lebih cepat daripada yang disadari banyak orang. Laju setiap revolusi industri telah meningkat drastis dibandingkan revolusi sebelumnya, dan revolusi AI pun tak terkecuali. Bahkan perusahaan seperti Google, yang telah memimpin dunia yang mengutamakan perangkat seluler, telah mengubah strategi secara substansial untuk tetap menjadi yang terdepan. Seperti yang dijanjikan CEO Google, Sundar Pichai, “Kita akan beralih dari dunia yang mengutamakan perangkat seluler ke dunia yang mengutamakan AI.”

Richard Foster, dari Sekolah Manajemen Yale, mengatakan bahwa karena teknologi baru, sebuah perusahaan S&P kini digantikan hampir setiap dua minggu, dan rata-rata umur perusahaan S&P telah turun 75% menjadi 15 tahun selama setengah abad terakhir. Yang lebih menarik lagi adalah bahwa terlepas dari seberapa baik kinerja sebuah perusahaan, kesuksesan sebelumnya tidak memberikan perlindungan kecuali jika perusahaan tersebut memanfaatkan inovasi teknologi terkini.

Sejalan dengan itu, McKinsey menemukan bahwa perusahaan B2B dengan pertumbuhan tercepat “menggunakan analitik canggih untuk meningkatkan produktivitas penjualan mereka secara radikal dan mendorong pertumbuhan penjualan dua digit dengan penambahan minimal pada tim penjualan dan basis biaya mereka.” Dalam makalah lain, mereka memperkirakan bahwa pada tahun 2016, Rp. 260 Triliun hingga Rp. 390 Triliun diinvestasikan dalam AI, dan jumlah itu terus bertambah. McKinsey mengajukan alasannya: “Bukti awal menunjukkan bahwa AI dapat memberikan nilai nyata bagi pengadopsi serius dan

dapat menjadi kekuatan yang kuat untuk disrupti.” Pengadopsi AI awal, lanjut studi tersebut, memiliki margin keuntungan yang lebih tinggi, dan kesenjangan antara mereka dan perusahaan yang tidak mengadopsi AI di seluruh perusahaan diperkirakan akan melebar di masa depan.

Semua ini adalah kabar baik bagi bisnis yang merangkul inovasi. Peralihan ke lingkungan bisnis yang digerakkan oleh AI akan menciptakan pemenang besar di antara mereka yang bersedia merangkul revolusi AI.

## 1.2 MITOS DAN REALITAS AI

Bagi kebanyakan orang, AI mungkin tampak hampir supranatural. Namun setidaknya untuk saat ini, terlepas dari kemampuannya yang luas, AI lebih terbatas dari itu. Saat ini, ilmuwan komputer mengelompokkan AI ke dalam dua kategori: AI lemah atau *sempit* dan AI kuat, yang juga dikenal sebagai *kecerdasan umum buatan* (AGI). AGI didefinisikan sebagai AI yang dapat mereplikasi seluruh kemampuan kognitif manusia dan dapat menerapkan kecerdasan pada setiap masalah, bukan hanya satu. AI sempit hanya dapat berfokus pada tugas yang spesifik dan sempit.

Ketika Steven Spielberg menciptakan film AI, ia memvisualisasikan robot humanoid yang dapat melakukan hampir semua hal yang dapat dilakukan manusia. Dalam beberapa kasus, mereka menggantikan manusia sepenuhnya. AGI jenis ini masih bersifat hipotetis saat ini, dan belum jelas apakah atau kapan kita akan mengembangkannya. Para ilmuwan bahkan memperdebatkan apakah AGI benar-benar dapat dicapai dan apakah kesenjangan antara kecerdasan mesin dan manusia dapat ditutup. Penalaran, perencanaan, kesadaran diri: ini adalah karakteristik yang dikembangkan oleh manusia ketika mereka berusia dua atau tiga tahun; tetapi mereka tetap menjadi tujuan yang sulit dipahami oleh komputer modern mana pun.

Tidak ada komputer yang ada saat ini yang dapat berpikir seperti manusia, dan mungkin tidak ada komputer yang akan melakukannya dalam waktu dekat. Terlepas dari perhatian media, tidak ada alasan untuk khawatir bahwa simulacrum HAL, dari film Stanley Kubrick 2001, akan mengubah kehidupan perusahaan Anda. Di sisi lain, kecerdasan buatan bukan lagi bagian dari fiksi ilmiah, dan sudah ada berbagai macam aplikasi yang sukses dan pragmatis, beberapa di antaranya dibahas dalam Bagian II. Mayoritasnya adalah AI sempit, dan beberapa, paling banter, adalah AI luas. Kami mendefinisikan *AI luas* sebagai kombinasi dari sejumlah solusi AI sempit yang bersama-sama memberikan kemampuan yang lebih kuat seperti kendaraan otonom. Tidak satu pun dari ini adalah aplikasi AGI.

Jadi bagaimana perusahaan menggunakan AI untuk berhasil di dunia yang terus berubah ini?

## 1.3 SIKLUS KEBAJIKAN DATA DAN ALGORITMA

Semakin banyak perusahaan yang menyadari bahwa dalam iklim bisnis yang terus berkembang saat ini, mereka akan segera dihargai bukan hanya karena bisnis yang sudah ada, tetapi juga karena data yang mereka miliki dan penggunaan algoritmiknya. Algoritma

memberikan nilai ekstrinsik pada data, dan terkadang bahkan nilai intrinsiknya – misalnya, data IoT seringkali sangat banyak sehingga tanpa algoritma yang kompleks, data tersebut tidak memiliki nilai inheren.

Manusia telah menganalisis data sejak petani pertama menjual atau menukar berkas gandum pertama dengan pelanggan pertamanya. Individu, dan kemudian perusahaan, terus menghasilkan analitik pada data mereka selama tiga revolusi industri pertama. Analisis data untuk meningkatkan bisnis menjadi semakin penting mulai sekitar tahun 1980, ketika perusahaan mulai menggunakan data mereka untuk meningkatkan proses bisnis sehari-hari. Pada akhir 1980-an, organisasi mulai mengukur sebagian besar proses bisnis dan rekayasa. Hal ini menginspirasi insinyur Motorola, Bill Smith, untuk menciptakan teknik formal untuk pengukuran pada tahun 1986. Tekniknya kemudian dikenal sebagai Six Sigma.

Perusahaan menggunakan Six Sigma untuk mengidentifikasi dan mengoptimalkan variabel dalam manufaktur dan bisnis untuk meningkatkan kualitas output suatu proses. Data yang relevan tentang operasi dikumpulkan, dianalisis untuk menentukan hubungan sebab-akibat, dan kemudian proses ditingkatkan berdasarkan analisis data. Menggunakan Six Sigma berarti mengumpulkan data dalam jumlah besar, tetapi itu tidak menghentikan sejumlah besar perusahaan untuk melakukannya. Pada tahun 1990-an, manajemen GE menjadikan Six Sigma sebagai pusat strategi bisnisnya, dan dalam beberapa tahun, dua pertiga dari perusahaan Fortune 500 telah menerapkan strategi Six Sigma.

Semakin banyak data, semakin banyak orang ingin menggunakannya untuk meningkatkan proses bisnis mereka. Semakin banyak data membantu, semakin mereka bersedia mengumpulkan data. Lingkaran umpan balik ini menciptakan siklus yang baik. Siklus baik ini adalah cara kerja AI dalam bisnis yang digerakkan oleh data—mengumpulkan data, membuat model yang memberikan wawasan, dan kemudian menggunakan wawasan ini untuk mengoptimalkan bisnis. Perusahaan yang ditingkatkan memungkinkan pengumpulan data lebih banyak – misalnya, dari pelanggan tambahan atau transaksi yang dimungkinkan oleh bisnis yang lebih optimal – memungkinkan model AI yang lebih canggih dan lebih akurat, yang selanjutnya mengoptimalkan bisnis.

#### **1.4 REVOLUSI YANG BERKELANJUTAN – MENGAPA SEKARANG?**

Meskipun AI telah ada sejak tahun 1950-an, baru dalam beberapa tahun terakhir ini AI mulai memberikan dampak bisnis yang signifikan. Hal ini disebabkan oleh perpaduan antara data berbasis internet, perangkat keras komputasi khusus, dan algoritma yang semakin matang.

Ide menghubungkan komputer melalui jaringan area luas, atau Internet, telah lahir pada tahun 1950-an, bersamaan dengan komputer elektronik itu sendiri. Pada tahun 1960-an, salah satu jaringan area luas ini didanai dan dikembangkan oleh Departemen Pertahanan AS dan disempurnakan di laboratorium ilmu komputer yang berlokasi di berbagai universitas di seluruh negeri. Pesan pertama di salah satu jaringan ini dikirim melalui apa yang saat itu dikenal sebagai ARPANET pada tahun 1969, dari Universitas California, Los Angeles, ke Universitas Stanford. Penyedia layanan internet (ISP) komersial mulai bermunculan pada akhir

tahun 1980-an. Protokol untuk apa yang kelak menjadi World Wide Web dikembangkan pada tahun 1980-an dan 1990-an. Pada tahun 1995, World Wide Web melejit, dan perdagangan daring pun muncul. Perusahaan-perusahaan daring mulai mengumpulkan lebih banyak data daripada yang mereka ketahui cara memanfaatkannya.

Bisnis selalu menggunakan data internal untuk analisis data. Namun, sejak awal mula internet, adopsi broadband di rumah, dan kemunculan media sosial serta ponsel pintar, interaksi digital kita tumbuh secara eksponensial, menciptakan era data yang dihasilkan pengguna. Proliferasi sensor, seperti yang dapat mengukur getaran mesin di lingkungan industri, atau mengukur suhu produk konsumen, seperti mesin pembuat kopi, menambah kekayaan data ini. Diperkirakan saat ini terdapat lebih dari 100 sensor per orang, yang semuanya mampu mengumpulkan data. Data ini menjadi apa yang kita sebut sebagai big data.

Big data mencakup informasi digital dalam jumlah luar biasa, yang dikumpulkan dalam bentuk yang dapat digunakan oleh komputer: data seperti gambar, video, riwayat belanja, informasi jejaring sosial, profil penelusuran, serta berkas suara dan musik. Kumpulan data yang sangat besar ini merupakan hasil digitalisasi proses-proses tambahan, seperti interaksi media sosial dan pemasaran digital. Paradigma baru harus dikembangkan untuk menangani data berskala internet ini: MapReduce pertama kali digunakan oleh Google pada tahun 2004 dan Hadoop oleh Yahoo pada tahun 2006 untuk menyimpan dan memproses kumpulan data besar ini. Penggunaan data ini untuk melatih model AI telah memungkinkan kita mendapatkan wawasan yang lebih signifikan dengan kecepatan yang lebih cepat, sehingga meningkatkan potensi solusi AI secara signifikan.

Meskipun volume data yang tersedia melonjak, biaya penyimpanan anjlok, menyediakan AI dengan semua bahan baku yang dibutuhkannya untuk membuat prediksi yang canggih. Pada awal tahun 2000-an, Amazon menghadirkan komputasi dan penyimpanan berbasis cloud, yang memungkinkan komputasi berkinerja tinggi pada kumpulan data besar tersedia bagi departemen TI di banyak bisnis. Pada tahun 2005, harga penyimpanan telah turun 300 kali lipat dalam 10 tahun, dari sekitar Rp. 3.000.000 menjadi sekitar Rp. 10.000 per gigabyte. Pada tahun 2010, Microsoft dan Google membantu memperluas kapasitas penyimpanan lebih lanjut dengan merilis produk penyimpanan dan komputasi cloud mereka: Microsoft Azure dan Google Cloud Platform.

Pada tahun 1960-an, salah satu pendiri Intel, Gordon Moore, meramalkan bahwa daya pemrosesan chip komputer akan berlipat ganda kira-kira setiap tahun. Dikenal sebagai Hukum Moore, hukum ini merujuk pada pertumbuhan eksponensial daya komputasi di komputer-komputer ini. Pada tahun 1990-an, terobosan perangkat keras seperti pengembangan unit pemrosesan grafis (GPU) meningkatkan daya pemrosesan komputasi lebih dari satu juta kali lipat, dengan kemampuan untuk menjalankan pemrosesan komputasi paralel. Awalnya digunakan untuk rendering grafis, GPU kemudian memungkinkan untuk melatih dan menjalankan algoritma AI canggih yang membutuhkan kumpulan data yang sangat besar. Baru-baru ini, Google telah memperkenalkan unit pemrosesan tensor (TPU), sebuah chip yang dipercepat AI untuk komputasi pembelajaran mendalam.

Selain perangkat keras, kemajuan dalam komputasi paralel dimanfaatkan untuk memparalelkan pelatihan model AI. Akses ke layanan-layanan ini di cloud dari Amazon, Microsoft, dan Google untuk perusahaan mana pun yang menginginkannya memudahkan banyak perusahaan untuk menjelajah ke ruang ini, di mana mereka akan lebih ragu-ragu jika setiap perusahaan harus membangun infrastruktur pemrosesan paralelnya sendiri yang besar dan skalabel.

Teknik-teknik terobosan dalam kecerdasan buatan telah terjadi sejak tahun 1950-an, ketika pekerjaan awal pada AI mulai dipercepat. Model-model yang didasarkan pada gagasan teoretis tentang cara kerja otak manusia, yang dikenal sebagai *jaringan saraf tiruan*, dikembangkan, diikuti oleh berbagai upaya lain untuk mengajarkan komputer agar belajar sendiri. Algoritma *pembelajaran mesin* (ML) ini memungkinkan komputer mengenali pola dari data dan membuat prediksi berdasarkan pola tersebut, seperti halnya jaringan saraf tiruan berlapis-lapis yang semakin kompleks yang digunakan dalam jenis pembelajaran mesin yang dikenal sebagai *pembelajaran mendalam*. Terobosan lain datang pada tahun 1980-an ketika metode *back-propagation* digunakan untuk melatih jaringan saraf tiruan, yang memungkinkan jaringan untuk mengoptimalkan dirinya sendiri tanpa campur tangan manusia. Sepanjang tahun 1990-an dan awal 2000-an, para ilmuwan mengembangkan lebih banyak pendekatan untuk membangun jaringan saraf tiruan guna memecahkan berbagai jenis masalah seperti pengenalan gambar, ucapan ke teks, peramalan, dan lainnya.

Pada tahun 2009, ilmuwan Amerika Andrew Ng, yang saat itu bekerja di Google dan Universitas Stanford, melatih jaringan saraf dengan 100 juta parameter pada unit pemrosesan grafis (GPU), menunjukkan bahwa apa yang mungkin membutuhkan waktu berminggu-minggu pada CPU kini dapat dihitung hanya dalam beberapa hari. Implementasi ini menunjukkan bahwa algoritma yang canggih dapat memanfaatkan kumpulan data besar yang tersedia dan memprosesnya pada perangkat keras khusus untuk melatih algoritma pembelajaran mesin dan pembelajaran mendalam yang kompleks.

Kemajuan dalam algoritma dan teknologi terus berlanjut, menghasilkan kemajuan yang mengejutkan dalam kemampuan komputer untuk melakukan tugas-tugas kompleks, yang ditunjukkan dengan jelas ketika program AlphaGo mengalahkan pemain Go manusia terbaik dunia pada tahun 2016. Permainan Go memiliki aturan yang sangat sederhana, tetapi lebih rumit untuk dimainkan daripada catur, dengan lebih banyak kemungkinan posisi papan daripada atom di alam semesta. Kompleksitas ini membuat mustahil untuk memprogram AlphaGo dengan pohon keputusan atau aturan tentang langkah mana yang harus diambil ketika berada di posisi papan tertentu. Untuk menang, AlphaGo harus belajar dari mengamati permainan profesional dan bermain melawan dirinya sendiri.

Masalah pelik pengenalan suara merupakan kebutuhan lain yang sulit dipecahkan. Keragaman aksen dan warna suara manusia yang tak terbatas sebelumnya menenggelamkan serangkaian upaya untuk membuat suara dapat dipahami oleh komputer. Namun, alih-alih memprogram untuk setiap skenario yang mungkin, para insinyur memasukkan terabyte data (seperti sampel suara) ke jaringan di balik algoritma pembelajaran pengenalan suara yang canggih. Mesin-mesin tersebut kemudian dapat menggunakan contoh-contoh ini untuk

mentranskripsikan suara. Pendekatan ini telah memungkinkan terobosan seperti Google, yang aplikasi Terjemahannya saat ini dapat menerjemahkan lebih dari 100 bahasa. Google juga telah merilis headphone yang dapat menerjemahkan 40 bahasa secara real-time.

Lebih dari sekadar pengenalan suara, perusahaan kini telah "mengajari" komputer cara memastikan dengan tepat apa yang diinginkan seseorang dan memenuhi kebutuhan itu, semua itu agar Alexa dapat memahami bahwa Anda ingin mendengarkan Bryan Adams, bukan Ryan Adams, atau membedakan antara dua band Australia, Dead Letter Circus dan Dead Letter Chorus. Asisten virtual seperti ini bisa jauh lebih berguna, karena bisa melakukan segalanya, mulai dari mencatat untuk dokter saat berinteraksi dengan pasien, hingga memilah sejumlah besar data penelitian dan merekomendasikan pilihan untuk terapi.

Bahkan ketika teknologi melaju pesat, teknik AI yang ada terus memberikan nilai yang luar biasa, memungkinkan cara-cara baru dan menarik untuk melakukan tugas-tugas seperti menganalisis gambar. Dengan kamera digital dan ponsel pintar, mengunggah gambar ke jejaring sosial seperti Facebook, Pinterest, dan Instagram menjadi lebih mudah dari sebelumnya. Gambar-gambar ini menjadi bagian yang semakin besar dari big data. Kekuatannya dapat diilustrasikan oleh penelitian yang dilakukan oleh Fei-Fei Li, profesor ilmu komputer di Universitas Stanford, dan kepala pembelajaran mesin di Google Cloud hingga saat ini.

Li, yang berspesialisasi dalam visi komputer dan pembelajaran mesin, berperan penting dalam menciptakan basis data berlabel ImageNet. Pada tahun 2017, ia menggunakan data berlabel untuk memprediksi secara akurat bagaimana lingkungan yang berbeda akan memilih hanya berdasarkan mobil yang diparkir di jalan mereka. Untuk melakukannya, ia mengambil gambar mobil berlabel dari situs web penjualan mobil Edmunds.com, dan menggunakan Google Street View, mengajarkan komputer untuk mengidentifikasi mobil mana yang diparkir di jalan mana. Dengan membandingkannya dengan data berlabel dari Survei Komunitas Amerika dan data pemungutan suara pemilihan presiden, ia dan rekan-rekannya dapat menemukan korelasi prediktif antara mobil, demografi, dan keyakinan politik.

Penelitian tentang AI dan penerapannya berkembang secara eksponensial. Universitas dan perusahaan teknologi besar melakukan penelitian dalam jumlah yang lebih besar untuk memajukan kapabilitas AI dan untuk lebih memahami mengapa AI bekerja dengan baik. Populasi mahasiswa yang mempelajari teknologi AI telah tumbuh secara proporsional, dan bahkan berbagai bisnis mendirikan kelompok penelitian AI dan program magang multi-tahun, seperti program residensi AI di Shell. Semua investasi ini terus mendorong evolusi AI.

Revolusi ini belum melambat. Dalam lima tahun terakhir, telah terjadi peningkatan 300.000 kali lipat dalam daya komputasi model AI. Pertumbuhan ini secara eksponensial lebih cepat daripada Hukum Moore, yang sendiri bersifat eksponensial. Namun, revolusi ini tidak lagi hanya berada di tangan akademisi dan sekelompok perusahaan teknologi besar. Transisi dari penelitian ke aplikasi sedang berlangsung dengan baik. Kombinasi kekuatan komputasi saat ini; gudang data yang sangat besar yaitu Internet; dan berbagai kerangka kerja pemrograman sumber terbuka yang gratis, serta tersedianya perangkat lunak yang mudah

digunakan dari Google, Microsoft, Amazon, dan lainnya mendorong semakin banyak bisnis untuk mengeksplorasi AI.

### 1.5 AI: KEUNGGULAN KOMPETITIF ANDA

Mendapatkan nilai dari AI bukan hanya tentang model mutakhir atau algoritma yang canggih: ini tentang menerapkan algoritma ini secara efektif dan mendapatkan adopsi bisnis untuk penggunaannya. AI belum menjadi teknologi plug-and-play. Meskipun data merupakan sumber daya yang melimpah, mengekstraksi nilai darinya bisa menjadi proposisi yang mahal. Bisnis harus membayar untuk pengumpulan, hosting, pembersihan, dan pemeliharannya. Untuk memanfaatkan data, perusahaan perlu membayar gaji insinyur data, ilmuwan AI, analis, pengacara, dan pakar keamanan untuk menangani masalah seperti risiko pelanggaran. Namun, keuntungannya bisa sangat besar.

Sebelum AI, perusahaan telepon biasa melihat metrik seperti berapa lama waktu yang dibutuhkan untuk memasang saluran pribadi. Rumah sakit memperkirakan berapa banyak uang yang akan mereka tagih yang tidak akan pernah terkumpul. Setiap perusahaan yang menjual sesuatu mempelajari siklus penjualannya – misalnya, berapa lama waktu yang dibutuhkan setiap tenaga penjualan mereka untuk menutup transaksi? Dengan menggunakan AI, perusahaan dapat melihat data secara berbeda. Perusahaan yang dulu bertanya "Berapa siklus penjualan rata-rata kami?" kini dapat bertanya, "Apa saja karakteristik pelanggan atau tenaga penjualan yang memiliki siklus penjualan lebih pendek? Apa yang dapat kita prediksi tentang siklus penjualan untuk pelanggan tertentu?" Kedalaman pengetahuan ini membawa keuntungan bisnis yang sangat besar.

Tidak diragukan lagi, ada potensi kerugian dalam mencoba menggunakan aplikasi AI secara luas. Membangun aplikasi AI itu rumit, dan sebagian besar digunakan tanpa benar-benar memahami bagaimana ia sampai pada keputusannya. Mengingat kurangnya transparansi ini (sering disebut masalah *kotak hitam*), sulit untuk mengetahui apakah mesin AI membuat penilaian yang benar dan tidak bias. Saat ini, masalah kotak hitam terutama melibatkan keputusan operasional berbasis AI yang tampaknya menangani faktor-faktor seperti ras atau gender secara tidak adil.

Sebuah studi oleh ProPublica<sup>21</sup> tentang algoritma yang dirancang untuk memprediksi residivisme (pelanggaran berulang) dalam populasi penjara menemukan bahwa narapidana kulit hitam jauh lebih mungkin ditandai memiliki tingkat kekambuhan yang lebih tinggi daripada narapidana kulit putih. Namun, ketika angka-angka ini dibandingkan dengan tingkat aktual yang terjadi selama dua tahun di Broward County, Florida, ternyata algoritmanya salah. Perbedaan ini menunjukkan masalah nyata: bukan hanya algoritma yang dapat membuat prediksi yang salah, tetapi kurangnya transparansi algoritma dapat membuatnya mustahil untuk menentukan alasannya. Akuntabilitas juga bisa menjadi masalah. Terlalu mudah bagi orang untuk berasumsi bahwa jika informasi tersebut berasal dari komputer, itu pasti benar. Pada saat yang sama, jika algoritma AI membuat keputusan yang salah, siapa yang salah? Selain itu, jika Anda merasa hasilnya tidak adil atau akurat, apa jalan keluar Anda? Ini adalah masalah yang harus diatasi untuk mencapai manfaat penggunaan AI.

Penggunaan AI oleh JP Morgan merupakan contoh mengesankan tentang betapa efisiennya AI. Raksasa keuangan ini menggunakan perangkat lunak AI untuk melakukan tugas-tugas seperti menafsirkan perjanjian pinjaman komersial dan menjalankan fungsi-fungsi sederhana yang berulang seperti memberikan akses ke sistem perangkat lunak dan menanggapi permintaan TI, dan mereka berencana untuk mengotomatiskan pengajuan hukum yang kompleks. Menurut Bloomberg Markets, perangkat lunak ini "melakukan dalam hitungan detik apa yang dibutuhkan pengacara selama 360.000 jam."

Di sisi lain, perusahaan perdagangan multinasional Cargill mulai memasukkan AI ke dalam strategi bisnisnya. Pada awal 2018, *Financial Times* melaporkan bahwa Cargill sedang merekrut ilmuwan data untuk mencari cara memanfaatkan data yang semakin tersedia dengan lebih baik. Menurut *Times*, "ketersediaan data yang lebih luas – mulai dari pola cuaca hingga pergerakan kapal – telah mengurangi nilai pengetahuan internal tentang pasar komoditas."<sup>23</sup>

Tindakan Cargill menggambarkan dua poin penting. Strategi bisnis Anda mungkin akan mendapat manfaat dari penggunaan AI, meskipun Anda belum mengetahui cara melakukannya. Terlebih lagi, mengingat banyaknya data yang tersedia, kecanggihan algoritma AI yang terus berkembang, dan rekam jejak perusahaan-perusahaan sukses yang telah mengadopsi AI, inilah saat yang tepat untuk menentukan strategi AI Anda sekaligus mulai menerapkannya. Buku ini dirancang untuk membantu Anda melakukan keduanya. Sebagai permulaan, kita akan membahas apa itu AI dan bagaimana algoritma AI bekerja.

## **BAB 2**

### **APA ITU AI DAN BAGAIMANA CARA KERJANYA?**

Konsep AI bukanlah hal baru. Manusia telah membayangkan mesin yang dapat berkomputasi sejak zaman kuno, dan gagasan ini terus berlanjut hingga Abad Pertengahan dan seterusnya. Pada tahun 1804, Joseph-Marie Jacquard menciptakan alat tenun yang "diprogram" untuk membuat kain tenun menggunakan hingga 2.000 kartu berlubang. Mesin tersebut tidak hanya dapat menggantikan penenun, tetapi juga membuat pola yang mungkin membutuhkan waktu berbulan-bulan bagi manusia untuk menyelesaikannya, dan dapat mereplikasinya dengan sempurna.

Namun, baru pada akhir abad ke-20 AI mulai tampak seperti tujuan yang dapat dicapai. Bahkan saat ini, kecerdasan buatan bukanlah istilah yang didefinisikan secara tepat. Dalam sebuah artikel yang diterbitkan pada 14 Februari 2018, Forbes menawarkan enam definisi, yang pertama berasal dari *The English Oxford Living Dictionary*: "Teori dan pengembangan sistem komputer yang mampu melakukan tugas-tugas yang biasanya membutuhkan kecerdasan manusia, seperti persepsi visual, pengenalan suara, pengambilan keputusan, dan penerjemahan antarbahasa." Ini adalah tempat yang masuk akal untuk memulai karena contoh-contoh dalam definisi tersebut adalah jenis AI yang saat ini sedang digunakan: AI yang lemah atau sempit.

#### **2.1 PERKEMBANGAN AI SEMPIT**

Komputer digerakkan oleh algoritma: langkah-langkah logis yang ditulis dalam kode yang dirancang untuk menghasilkan hasil yang diinginkan. "Komputer pintar" paling awal menggunakan sejenis algoritma yang disebut sistem pakar. Sistem pakar didasarkan pada bagaimana seorang pakar di bidang tertentu akan menemukan jawaban atas pertanyaan yang diberikan, yang dihasilkan sebagai serangkaian aturan yang diikuti oleh algoritma tersebut. Antara tahun 1959 dan 1980-an, ilmuwan komputer berfokus pada pengembangan sistem pakar (beberapa di antaranya masih kita gunakan hingga saat ini), membangunnya berdasarkan pengetahuan spesifik domain yang diciptakan oleh para pakar di berbagai bidang. Selama masalah yang harus dipecahkan melibatkan aturan logika formal yang relatif mudah diungkapkan, sistem pakar memadai untuk tugas tersebut.

Komputer catur IBM yang terkenal, Deep Blue, adalah sebuah sistem pakar, dan keberhasilannya disebabkan oleh sifat permainan catur itu sendiri. Pemain catur yang baik mempertimbangkan setiap kemungkinan langkah dalam situasi tertentu dan mencoba memikirkan opsi-opsi ini di kepala mereka sejauh mungkin. "Jika saya menggerakkan pion ini, apa yang akan dilakukan lawan saya selanjutnya? Apa yang akan dia lakukan lima langkah dari sekarang?" Intinya, itulah yang diprogram untuk dilakukan Deep Blue. Pemain ahli "mengajari" para programmer Deep Blue langkah-langkah catur dan strategi yang dapat memenangkan permainan.

Sistem pakar merupakan teknologi mutakhir ketika, pada tahun 1996 dan 1997, Deep Blue menghadapi grandmaster catur Rusia, Gary Kasparov, dan memenangkan pertandingan kedua dari dua pertandingan mereka. Namun, ada banyak hal yang tidak dapat dilakukan sistem pakar, seperti memahami ucapan atau penglihatan, atau penalaran tentang dunia fisik. Tugas-tugas ini jauh kurang terstruktur, dan bisa sangat sulit atau bahkan mustahil untuk dijelaskan atau dikodifikasikan menjadi langkah-langkah logis. Untuk melakukannya – untuk benar-benar "bersaing" dengan manusia – komputer harus melakukan dua hal yang jauh lebih rumit, hal-hal yang mudah dilakukan manusia tetapi sulit dijelaskan: komputer harus mampu belajar dari pengalaman dan membangun intuisi.

Intuisi lebih dari sekadar perasaan. Intuisi adalah proses yang dikuasai otak manusia untuk memahami segudang detail luar biasa dan seringkali membingungkan yang disajikan dunia setiap hari. Ketika kita melihat seekor burung, entah itu merah atau biru, besar atau kecil, kepalanya tegak atau kakinya tegak, atau bahkan setengah membusuk, bagaimana kita menentukan bahwa itu adalah burung? Apakah ia berdiri dengan dua kaki, bukan empat? Manusia juga melakukannya. Apakah ia bertelur? Begitu juga kura-kura. Bagaimana dengan fakta bahwa ia terbang? Pesawat terbang.

Bagaimana manusia mengambil langkah dari semua informasi yang kita miliki di otak kita tentang burung, seperti perbedaan ukuran, warna, bentuk paruh, kebiasaan, dan cara terbangnya, ke sebuah keputusan yang memungkinkan kita mengetahui bahwa makhluk tertentu termasuk dalam kategori yang disebut burung? Seorang anak berusia dua tahun dapat melakukan ini. Namun, bagi komputer, ini adalah tugas yang menantang: tidak mudah untuk membuat seperangkat aturan yang dapat menentukan kapan komputer tersebut melihat gambar seekor burung.

Orang cenderung meremehkan jumlah pengetahuan yang dibutuhkan untuk mencapai perilaku seperti manusia dalam komputer sekaligus meremehkan hasil sistem yang dapat mendekatinya. Ada tugas yang dapat dilakukan manusia secara efisien, dan ada tugas yang dapat dilakukan mesin dengan sangat baik, dan tugas-tugas ini cenderung sangat berbeda. Tugas ilmiah atau teknis sering kali memiliki seperangkat aturan yang ketat, meskipun mungkin canggih, yang memandu perilakunya. Jenis pengetahuan ini relatif mudah dikodifikasi; itulah sebabnya beberapa sistem pakar masih layak bahkan 40 tahun yang lalu.

Komputer juga mampu menangani data dalam jumlah yang luar biasa besar. Kemampuan ini membuat para ilmuwan bertanya-tanya: dengan data yang cukup, akankah ada cara bagi komputer untuk mencapai kesimpulan berdasarkan informasi ini tanpa intuisi manusia yang setara? Bisakah mereka mengumpulkan dan mengkategorikan sebanyak mungkin pengetahuan dari "pengalaman" tanpa memerlukan para ahli untuk memberi tahu mereka cara melakukan sesuatu? Bisakah komputer belajar sendiri dari contoh?

World Wide Web adalah salah satu hal yang memungkinkan ide ini terwujud. Para ilmuwan tiba-tiba memiliki aliran data daring yang sangat besar yang dapat mereka manfaatkan. Jika komputer dapat berpikir seperti manusia, belajar dari semua data ini untuk mencapai kesimpulan, sesuatu yang mendekati AI mungkin saja terjadi. Namun, pengembangan di bidang ini masih terbatas. Saat ini terdapat banyak pendekatan dalam AI

yang dapat diterapkan pada berbagai masalah, tetapi selama 10 tahun terakhir, pembelajaran mesin telah menjadi yang paling umum. Namun dalam lima tahun terakhir, jaringan saraf dalam, versi pembelajaran mesin yang jauh lebih canggih, telah mulai menggantikan kemampuan manusia di banyak bidang seperti pengenalan gambar. Jaringan saraf ini pada akhirnya dapat menjadi komponen AI yang kuat.

## 2.2 JARINGAN SARAF PERTAMA

Awalnya, AI terinspirasi oleh logika: jika ini, maka itu. Namun, logika adalah sesuatu yang dipelajari orang di kemudian hari. Misalnya, anak usia tiga tahun tidak berpikir secara eksplisit dan logis; mereka belajar melalui pengamatan pola. Kita belum mengetahui secara pasti semua aspek tentang cara kerja otak, terutama selama tahun-tahun awal tersebut. Namun, terdapat beberapa model teoretis yang relevan, dan para ilmuwan komputer memutuskan bahwa salah satu model ini mungkin merupakan paradigma yang jauh lebih baik untuk AI daripada logika. Begitulah lahirnya jaringan saraf tiruan.

Teori ilmu saraf saat ini menyatakan bahwa neuron (sel otak pada manusia dan simpul dalam program komputer) tersusun berlapis-lapis, dengan informasi pertama-tama melewati lapisan "terbawah", lalu lapisan berikutnya, dan seterusnya. Informasi tersebut terus disempurnakan saat melewati rantai tersebut. Jaringan saraf tiruan (JST) sederhana telah digagas sejak tahun 1940-an. Pada tahun 1957, Frank Rosenblatt dari Universitas Cornell membangun sebuah prototipe yang ia sebut Perceptron. Hanya terdiri dari dua lapisan neuron – lapisan masukan dan lapisan keluaran – prototipe tersebut berhasil membedakan antara kartu yang ditandai di sebelah kiri dan kartu yang ditandai di sebelah kanan. Perceptron adalah algoritma pembelajaran pertama.

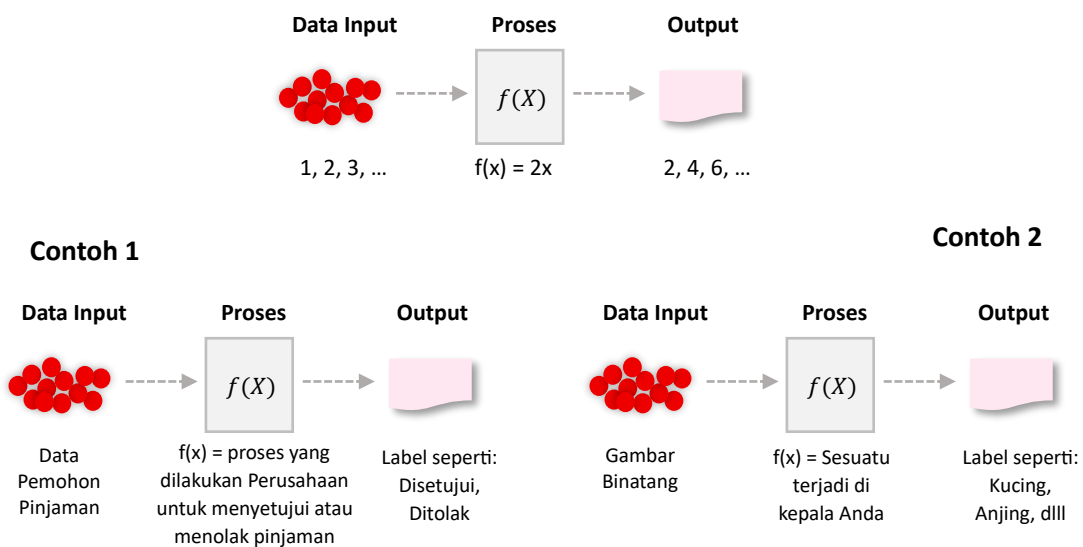
Pada tahun 1958, Arthur Samuel, seorang pelopor di bidang ilmu komputer dan AI, memperkenalkan istilah *pembelajaran mesin* untuk mencakup berbagai cara komputer dapat belajar, termasuk menggunakan Perceptron. Sejak itu, istilah tersebut telah digunakan untuk merujuk pada berbagai teknik AI di mana mesin belajar dari pengalaman.

## 2.3 PEMBELAJARAN MESIN

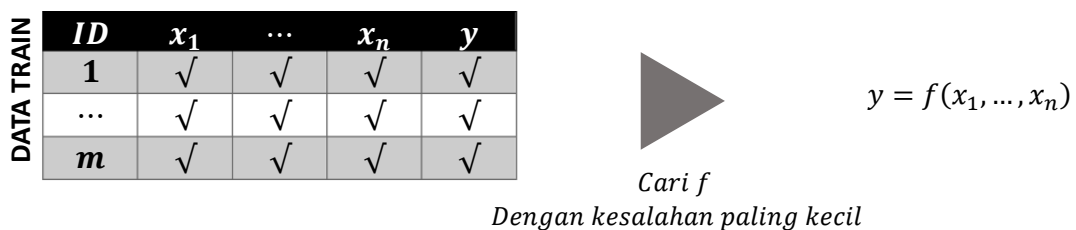
Pembelajaran mesin didasarkan pada algoritma yang dapat belajar dari data tanpa bergantung pada pemrograman eksplisit berbasis domain atau berbasis aturan – yaitu, pemrograman yang dirancang khusus untuk memecahkan masalah tertentu. Tidak ada yang secara eksplisit memprogram komputer atau membuat kode logika secara manual untuk memungkinkannya melakukan tugas tertentu. Algoritma dirancang untuk menentukan atau memperkirakan fungsi yang memprediksi keluaran berdasarkan serangkaian masukan. Pembelajaran mesin merupakan pendekatan yang berguna ketika terdapat data representatif yang memadai dan akurat, dan ketika mungkin sulit atau mahal untuk memodelkan pengetahuan domain secara manual. Alih-alih secara eksplisit "mengajarkan" sistem dengan memodelkan pengetahuan seperti manusia, sistem dirancang untuk belajar dari data.

Tujuan pembelajaran mesin adalah untuk mempelajari fungsi tersebut dari sejumlah besar observasi historis nilai masukan dan nilai keluarannya yang sesuai, serta untuk secara

akurat memprediksi nilai keluaran di masa mendatang berdasarkan nilai masukan di masa mendatang. Proses ini mungkin terdengar relatif sederhana, tetapi fungsi-fungsi ini bisa sangat rumit, seringkali terlalu rumit untuk diturunkan oleh manusia. Fungsi yang diperkirakan dapat berasal dari proses apa pun (lihat Gambar 2.1). Misalnya, proses tersebut mungkin mengambil angka masukan dan mengalikannya dengan 2, lalu menghasilkan angka keluaran. Sebagai alternatif, proses tersebut mungkin mengambil aplikasi pinjaman (data) sebagai masukan dan membuat keputusan pinjaman, lalu memberikan label "disetujui" atau "ditolak" sebagai keluaran. Sebagai alternatif, proses tersebut mungkin berupa apa pun yang terpikirkan oleh Anda untuk mengambil gambar masukan dan memberinya label "kucing" atau "bukan kucing" sebagai keluaran. Berdasarkan masukan dan keluaran ini,



**Gambar 2.1** Contoh fungsi  $f(x)$  yang dapat diestimasi dengan menggunakan pembelajaran mesin pada kumpulan data masukan dan keluaran.



**Gambar 2.2** Menggunakan data pelatihan untuk pelanggan 1 hingga  $m$  untuk memperkirakan  $f$  yang akan memprediksi  $y$  jika diberikan  $x_1, \dots, x_n$ .

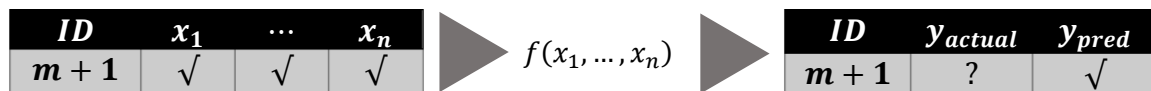
algoritma pembelajaran mesin akan memperkirakan fungsi yang meniru proses ini sekaligus mengoptimalkannya untuk menghasilkan kesalahan terendah.

Mari kita pertimbangkan contoh lain. Misalkan Anda memiliki data historis tentang perilaku pelanggan. Data ini mungkin mencakup berapa kali mereka menghubungi dukungan pelanggan, berapa banyak yang mereka belanjakan untuk berlangganan produk atau layanan

Anda, dan sebagainya. Sebut perilaku ini  $x_1, \dots, x_n$ . Misalkan Anda juga tahu pelanggan mana yang menutup akun dan keluar (churn) atau tetap berlangganan. Sebut ini  $y$ .

Pembelajaran mesin mengekstrak pola dari data untuk mencari fungsi  $f$  yang paling akurat memprediksi  $y$  dari  $x_1, \dots, x_n$ . Fungsi ini akan menjadi fungsi estimasi dengan kesalahan terendah, misalnya,  $f$  dengan persentase positif palsu dan negatif palsu terendah.  $f$  ini adalah model pembelajaran mesin Anda (lihat Gambar 2.2). Saat Anda mendapatkan informasi  $x_1, \dots, x_n$  tentang pelanggan baru, Anda dapat meneruskan informasi tersebut ke  $f$ , dan model tersebut akan memprediksi apakah pelanggan ini kemungkinan akan keluar atau tidak dalam interval kepercayaan tertentu (lihat Gambar 2.3). Kita akan membahas model AI langkah demi langkah yang terperinci untuk memecahkan masalah churn pelanggan ini di Bab 13.

Apa yang saya jelaskan di sini adalah penyederhanaan yang berlebihan berdasarkan regresi statistik dari kelas enam, yang ditemukan pada tahun 1889 oleh Francis Galton dan telah digunakan sejak saat itu. Sering kali terdapat kebingungan tentang



**Gambar 2.3 Menggunakan model pembelajaran mesin ( $f$ ) untuk memprediksi apakah pelanggan nomor  $m + 1$  akan berhenti berlangganan.**

bagaimana pembelajaran mesin berbeda dari statistik atau pemodelan statistik. Pembelajaran mesin adalah salah satu bentuk pembelajaran statistik, tetapi terdapat perbedaan krusial di antara keduanya: model pembelajaran mesin digunakan untuk membuat prediksi tentang data masa depan (yang belum terlihat), sedangkan model statistik menjelaskan hubungan antara data masukan historis dan variabel hasil (model tersebut dimaksudkan untuk bersifat deskriptif dan tidak membuat prediksi tentang kumpulan data masa depan). Model statistik melihat ke belakang, sedangkan model pembelajaran mesin melihat ke masa depan. Kesamaan ini muncul karena keduanya menggunakan konsep dasar probabilitas. Kita akan mengeksplorasi konsep "data yang belum terlihat untuk dapat memprediksi hasil di masa depan" di Bab 8. Pembelajaran mesin juga sering kali memiliki nonlinieritas ekstrem yang tertanam di dalamnya dalam jaringan saraf. Dalam contoh sebelumnya, jika Anda mengganti  $f$  dengan jaringan saraf dalam dengan 100.000 parameter untuk dioptimalkan, Anda mendapatkan pembelajaran mendalam. Selain itu,  $x_1, \dots, x_n$  dapat berupa data apa pun, seperti intensitas setiap piksel pada kamera mobil tanpa pengemudi, dan  $y$  dapat berupa label "rambu berhenti".

### Jenis-jenis Penggunaan Pembelajaran Mesin

Pembelajaran mesin memungkinkan pengguna untuk mengelola data mereka secara efektif dan menghasilkan prediksi darinya. Sebagian besar masalah yang dapat diatasi dengan pembelajaran mesin secara efektif terdiri dari tiga kategori berikut.

**Klasifikasi** adalah proses memprediksi kategori terbaik untuk masukan data baru tertentu, dari serangkaian kategori yang telah ditentukan sebelumnya. Klasifikasi digunakan jika keluarannya terdiri dari serangkaian kelas tetap – misalnya, mengkategorikan aplikasi

pinjaman menjadi yang harus disetujui dan yang tidak. Model dilatih untuk mengklasifikasikan masukan baru ke dalam salah satu dari dua kelas ini.

**Pengelompokan** adalah proses menemukan pengelompokan atau struktur dalam data, mempartisi data ke dalam klaster yang memiliki properti spesifik yang tidak berasal dari serangkaian kategori yang telah ditentukan sebelumnya tetapi muncul dari data. Metode ini sering digunakan dalam segmentasi pelanggan untuk memahami preferensi mereka berdasarkan informasi profil dan perilaku daring mereka.

**Regresi** adalah proses memprediksi nilai keluaran kontinu yang sesuai dengan masukan data baru yang diberikan. Contohnya adalah memprediksi suhu besok, yang bisa jadi 98,1 derajat, 98,2 derajat, atau bahkan 98,123 derajat. Prediksi suhu melibatkan kemungkinan hasil yang tak terbatas jumlahnya. Tidak seperti klasifikasi, intinya bukanlah memprediksi di kelas mana data baru tersebut berada.

### **Jenis-jenis Algoritma Pembelajaran Mesin**

Untuk melakukan klasifikasi, pengelompokan, dan regresi, pembelajaran mesin menggunakan berbagai teknik atau algoritma. Contoh-contoh berikut hanyalah beberapa metodologi yang tersedia saat ini. Bagi para manajer, memahami dasar-dasar beberapa algoritma lebih bermanfaat daripada menjadi ahli dalam masing-masing algoritma untuk membangun intuisi tentang apa yang terjadi dalam proses pemodelan. Hal ini membantu dalam pengambilan keputusan berdasarkan keluaran model karena Anda memiliki pemahaman yang lebih baik tentang bagaimana model tersebut sampai pada kesimpulannya. Algoritma baru terus dikembangkan, dan ini adalah sesuatu yang perlu terus diperbarui oleh para ilmuwan AI di organisasi.

**Pohon keputusan** adalah cara untuk menganalisis aliran data yang melibatkan pembuatan sistem percabangan, yang simpul-simpulnya umumnya membagi data menjadi dua kategori: misalnya, orang yang menyukai film dan orang yang tidak. Simpul-simpul berikutnya membagi data dengan cara yang sama, mengembangkan pohon, cabang demi cabang. Bahkan ada yang disebut tunggul data, yang hanya menanyakan satu pertanyaan dari data Anda. Variasi pohon keputusan adalah hutan acak, dan ini terdiri dari banyak pohon keputusan di mana rata-rata tertimbang digunakan di seluruh pohon keputusan yang berbeda.

**Regresi logistik** digunakan untuk klasifikasi (bukan regresi). Misalkan Anda memiliki kasus dengan dua fitur, seperti berapa kali pelanggan mengeluh dan jumlah tagihan bulanan rata-rata mereka. Salah satu cara untuk memahami regresi logistik adalah dengan memplot titik-titik data pada grafik dengan keluhan pada satu sumbu dan jumlah tagihan pada sumbu lainnya, lalu menemukan garis lurus yang memisahkan titik-titik data tersebut menjadi dua kategori – dengan semua titik data di satu sisi garis termasuk dalam satu kategori (misalnya, "kemungkinan besar akan berhenti berlangganan") dan semua titik data di sisi lain garis termasuk dalam kategori lain (misalnya, "kemungkinan besar tidak akan berhenti berlangganan"). Semakin baik posisi garis yang membagi kedua kelompok, semakin akurat klasifikasi Anda. Dengan lebih dari dua fitur, konsepnya tetap sama, dan lebih banyak dimensi yang digunakan.

**Mesin vektor pendukung** juga mencari garis yang memisahkan data menjadi dua kategori dalam dua dimensi, tetapi garis-garis ini tidak harus lurus; dalam tiga dimensi atau lebih, visualisasi ini menjadi lebih rumit, tetapi sekali lagi, konsepnya tetap sama.

**Model ansambel** adalah model di mana beberapa model dilatih dan digunakan bersama-sama. Pendekatan sederhana untuk ini adalah melatih model yang sedikit berbeda dan menggunakan rata-rata keluaran dari semuanya untuk masukan tertentu. Ansambel seringkali memberikan akurasi rata-rata yang lebih tinggi untuk data baru. Ada pendekatan lain untuk model ansambel, termasuk *bagging* dan *boosting*. Bagging melatih algoritma yang identik, seperti hutan acak, pada berbagai subset data dan menerapkan ensemble ini ke seluruh dataset. Boosting melatih satu model demi satu; setiap model berikutnya berfokus pada pembelajaran dari kekurangan model sebelumnya. Misalnya, model kedua akan berfokus pada prediksi untuk masukan data di mana keluarannya salah pada model pertama. Dalam Bab 13, kita akan melihat jenis akurasi yang diberikan ketika kita beralih dari model regresi logistik ke model boosting gradien ekstrem.

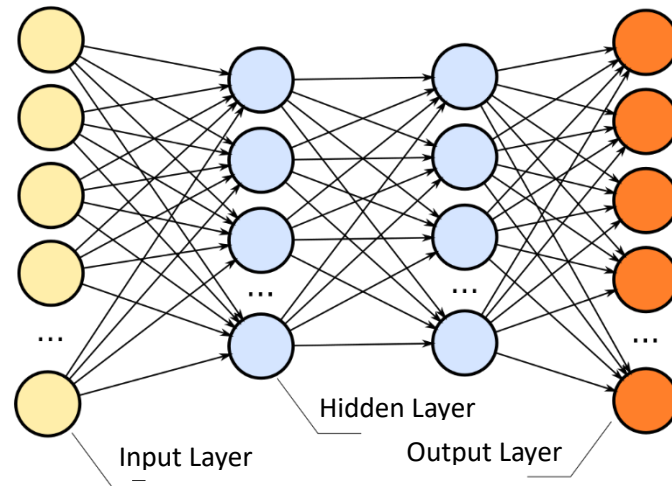
**Pembelajaran mendalam** merupakan evolusi dari Perceptron karya Frank Rosenblatt. Perceptron hanya memiliki satu lapisan neuron. Pelopor AI, Marvin Minsky, menunjukkan pada tahun 1969 bahwa ada beberapa fungsi bermanfaat yang tidak dapat dipelajari oleh Perceptron. Minsky percaya bahwa jaringan saraf berlapis-lapis, atau dalam, mungkin bekerja lebih baik, menghasilkan perceptron berlapis-lapis (MLP). Dua peneliti di Universitas Toronto, Geoffrey Hinton dan Yann LeCun, sependapat dengan teori Minsky. Pada tahun 1980-an, mereka berteori bahwa lebih banyak lapisan simpul saraf merupakan kunci menuju tingkat persepsi yang lebih dalam.

"Dalam" dalam *pembelajaran mendalam* tidak mengacu pada kedalaman pemahaman komputer yang sebenarnya tentang data (atau pengetahuan domain yang tersirat oleh data), tetapi pada struktur jaringan saraf tiruan itu sendiri. Dalam jaringan saraf dalam, terdapat lebih banyak lapisan simpul di antara simpul masukan dan simpul keluaran (lihat Gambar 2.4). Dalam beberapa kasus, mungkin terdapat ratusan. Keunggulan pembelajaran mendalam adalah daya ekspresifnya yang luar biasa tinggi, artinya pembelajaran mendalam dapat dilatih untuk mempelajari fungsi-fungsi yang sangat kompleks.

Untuk memahami cara kerja jaringan saraf dalam dan pembelajaran mendalam, mari kita lihat contoh penentuan apakah suatu gambar merupakan gambar burung atau bukan. Bayangkan informasi yang masuk ke lapisan pertama jaringan saraf, lapisan input, bersifat visual, berupa kumpulan piksel dengan warna dan tingkat kecerahan yang berbeda. Lapisan jaringan saraf selanjutnya mungkin mendeteksi tepi pada setiap gambar, menemukan bagian-bagian yang satu sisinya lebih gelap dan sisi lainnya lebih terang. Proses ini akan memberikan gambaran kasar bentuk-bentuk pada gambar.

Lapisan neuron berikutnya mungkin menerima keluaran lapisan pertama dan belajar mendeteksi faktor-faktor yang lebih kompleks seperti titik pertemuan tepi pada sudut tertentu, misalnya, di sudut-sudut. Sekelompok neuron ini mungkin merespons dengan kuat, misalnya, sudut paruh burung, meskipun jaringan saraf tidak akan menyadari bahwa ia sedang melihat sebuah "paruh"; neuron-neuron tersebut hanya akan mengidentifikasi pola tertentu.

Tingkat berikutnya mungkin menemukan konfigurasi yang lebih rumit, seperti pengelompokan tepi yang tersusun melingkar. Ketika neuron-neuron ini aktif, mereka mungkin merespons kepala burung yang melengkung,



**Gambar 2.4** Contoh jaringan saraf dalam.

meskipun jaringan saraf tersebut tidak akan mengetahui konsep kepala. Pada tingkat yang lebih tinggi, neuron-neuron mungkin mendeteksi penjajaran berulang sudut-sudut seperti paruh di dekat lingkaran-lingkaran seperti kepala, yang dapat menjadi sinyal kuat bahwa jaringan tersebut "melihat" kepala burung tersebut.

Demikian pula, neuron pada setiap lapisan berikutnya merespons konsep dengan kompleksitas dan abstraksi yang lebih tinggi, berevolusi dari pengenalan bentuk yang lebih lugas menuju, misalnya, kompleksitas mengidentifikasi garis luar bulu. Terakhir, neuron di tingkat teratas, atau lapisan keluaran, berkorespondensi dengan konsep "burung" atau "bukan burung". Hanya satu dari neuron ini yang akan aktif di lapisan keluaran berdasarkan hal-hal yang dipicu di lapisan sebelumnya.

Namun, jika ingin belajar, jaringan saraf harus melakukan lebih dari sekadar mengirim pesan ke lapisan-lapisan tersebut. Jaringan saraf harus memastikan bahwa ia mendapatkan hasil yang tepat dari lapisan teratas berdasarkan data berlabel. Jika tidak, jaringan saraf harus mengirim pesan kembali agar semua lapisan sebelumnya dapat menyempurnakan keluaran mereka untuk meningkatkan hasilnya. Anda dapat menganggap ini sebagai penalti karena tidak mencocokkan label "burung" untuk gambar masukan seekor burung. Penalti ini menyesuaikan kembali simpul mana di lapisan sebelumnya yang memicu atau tidak, berdasarkan gambar masukan. Jika lapisan keluaran sekarang cocok dengan label, maka pelatihannya selesai. Jika tidak, jaringan akan melakukan penyesuaian lagi, dan seterusnya hingga label keluaran sesuai dengan gambar masukan.

Oleh karena itu, semakin dalam jaringan saraf tiruan, yaitu semakin banyak tingkatan yang dimilikinya, semakin baik hasilnya hingga mencapai tingkat kedalaman tertentu. Seiring waktu, sistem ini mulai disebut jaringan saraf tiruan dalam, dan konsep ini membentuk dasar

pembelajaran mendalam. Lapisan tambahan dalam jaringan saraf tiruan dalam memungkinkan algoritma untuk menemukan fitur hierarkis, fitur yang secara langsung bersesuaian dengan masukan, diikuti oleh fitur yang bersesuaian dengan elemen tingkat pertama, dan seterusnya. Representasi bertingkat ini membuat jaringan pembelajaran mendalam lebih baik untuk memecahkan lebih banyak jenis masalah.

Hinton dan dua rekannya menulis sebuah makalah yang menawarkan solusi algoritmik yang mengurangi kesalahan yang dihasilkan dalam jaringan saraf tiruan. Pendekatan mereka, yang dikenal sebagai backpropagation, menjadi fondasi gelombang kedua jaringan saraf tiruan yang semakin akurat. Backpropagation adalah cara untuk menentukan secara lebih akurat cara mengkalibrasi koneksi antara setiap neuron di satu lapisan dengan neuron di lapisan yang lebih tinggi sehingga sistem dapat menghasilkan keluaran yang semakin akurat.

Jaringan saraf tiruan semakin mendalam, memungkinkan komputer mempelajari tugas-tugas yang lebih rumit, seperti pengenalan suara, pengenalan objek tiga dimensi, dan pemrosesan bahasa alami. Arsitektur pembelajaran mendalam baru—bagaimana simpul-simpul dalam jaringan saraf tiruan terhubung satu sama lain di semua lapisan—sedang bermunculan, masing-masing sesuai untuk berbagai jenis masalah. Jaringan saraf tiruan konvolusional (CNN), misalnya, paling baik dalam pengenalan gambar, dan jaringan saraf tiruan berulang (RNN) baik dalam pemrosesan bahasa alami.

Model pembelajaran mendalam jauh lebih baik dalam banyak masalah dibandingkan algoritma pembelajaran mesin klasik (yaitu, non-dalam). Namun, model ini lebih sulit dikembangkan karena membutuhkan pengetahuan tentang arsitektur jaringan saraf tiruan dan teknik optimasi. Model ini juga membutuhkan daya komputasi yang besar dan data pelatihan yang sangat banyak. Menggunakan metode pembelajaran mesin yang lebih lugas selain jaringan saraf tiruan dalam mungkin lebih efisien atau lebih murah untuk beberapa tugas, terutama jika kinerjanya cukup baik, dan akurasi model inkremental tidak diperlukan. Namun, seiring dengan penurunan biaya komputasi, aplikasi komersial untuk pembelajaran mendalam telah meningkat. Saat ini, misalnya, sistem pembelajaran mendalam digunakan untuk mencari transaksi perbankan yang mencurigakan secara real time, memicu pesan teks tentang pembelian daring yang mungkin tidak Anda lakukan sendiri.

## **2.4 PEMBELAJARAN TERAWASI, TANPA PENGAWASAN, DAN SEMI-TERAWASI**

Dari tahun 1950-an hingga 1980-an, seiring dengan semakin canggihnya algoritma, mereka menjadi mampu mempelajari hal-hal yang semakin rumit. Masalahnya, tidak ada cukup data yang tersedia untuk membuat prediksi yang akurat. Dengan munculnya World Wide Web pada tahun 1990-an dan digitalisasi proses bisnis, sejumlah besar data mulai tersedia. Para peneliti mulai berpikir tentang bagaimana mereka dapat menggunakan informasi baru ini untuk memungkinkan komputer mempelajarinya. Salah satu langkah yang mereka ambil adalah memberi label pada data tersebut.

*Data berlabel* adalah data yang sudah dikategorikan, seringkali oleh manusia. "Burung" adalah sebuah kategori. "Bukan-burung" juga merupakan sebuah kategori. Untuk melatih jaringan saraf agar mengenali burung, Anda dapat memberi label pada seribu gambar yang

memuatnya sebagai "burung", serta seribu gambar yang tidak memuat burung sebagai "bukan-burung". Bersama-sama, kedua set data berlabel ini dapat digunakan untuk melatih jaringan saraf agar mengenali gambar burung.

Penggunaan data berlabel dengan cara ini dikenal sebagai pembelajaran terawasi. Dalam pembelajaran terawasi, data pelatihan yang diberikan ke sistem mencakup keluaran yang diinginkan, atau kesimpulan, untuk setiap masukan. Data tersebut juga dapat mencakup contoh kesimpulan positif dan negatif. Pembelajaran terawasi saat ini digunakan untuk segala hal, mulai dari mengidentifikasi sel kanker hingga mengenali spam. Filter spam mempertimbangkan banyak faktor saat menyortir email, seperti asal pesan dan perangkat lunak apa yang digunakan. Spammer cenderung menggunakan perangkat lunak, atau mesin spam, yang mengirimkan sejumlah besar pesan yang tidak dapat dilacak dengan cepat. Filter spam juga mencari melalui rangkaian karakter dalam email yang mengkategorikan spam yang sudah diketahui – seperti pesan tentang Viagra atau permintaan dari seorang pangeran Nigeria.

Namun, seiring semakin cerdasnya spammer, filter spam berbasis aturan sederhana mulai gagal. Untuk mengatasi masalah ini, ilmuwan komputer mulai menggunakan pembelajaran mesin terawasi untuk membedakan spam dari bukan spam. Para ilmuwan telah memiliki banyak data tentang seperti apa rupa spam, sebagian berkat data yang diberi label secara manual. Jadi, mereka mengembangkan algoritma yang memungkinkan program mempelajari mana yang spam dan mana yang bukan, menggunakan contoh email yang sudah ada yang dikategorikan sebagai spam atau bukan spam. Bahkan saat ini, ketika Anda menandai email sebagai spam, data tersebut akan diberi label, yang kemudian digunakan untuk melatih ulang algoritma.

Mengidentifikasi kanker agak berbeda dengan mengidentifikasi spam, tetapi pembelajaran terawasi juga telah digunakan untuk mencapai hal ini. Tentu saja, tidak ada filter sel kanker yang dapat menjadi dasar pengembangan algoritma. Lebih penting lagi, untuk mendeteksi sel kanker, komputer perlu menganalisis gambar; dan setiap gambar dari setiap sel terdiri dari sejumlah besar titik data. Mesin harus mencari tahu apa dalam kombinasi semua titik data ini yang membuat suatu sel abnormal. Untungnya, manusia dapat dilatih untuk mengidentifikasi sel kanker, dan kita sudah memiliki sejumlah besar data tentang sel-sel abnormal ini, sehingga memungkinkan untuk membangun algoritma yang memungkinkan komputer belajar membedakannya.

Prosesnya dimulai dengan memasukkan sejumlah besar gambar sel ke dalam sistem. Seperti dalam model spam, dua set data digunakan: satu set yang gambarnya sebelumnya telah ditentukan, kemungkinan besar oleh dokter, sebagai kanker, dan satu set sel sehat lainnya. Dengan belajar dari data ini, komputer dapat melihat gambar yang belum pernah dilihat sebelumnya dan mengklasifikasikannya sebagai kanker atau non-kanker. Baik algoritma spam maupun sel kanker termasuk dalam kategori pembelajaran terawasi.

Dalam keduanya, terdapat (atau dibuat) data berlabel yang telah dikumpulkan sebelumnya yang dapat diorganisasikan dan digunakan agar komputer dapat mengenali pola dalam data. Umumnya, masalah seperti deteksi kanker terlalu rumit untuk dianalisis oleh

komputer menggunakan pembelajaran tanpa pengawasan. Sebagian besar aplikasi pembelajaran mesin saat ini mengandalkan pembelajaran terawasi. Namun, pembelajaran tanpa pengawasan dapat bermanfaat dalam banyak kasus.

Pada tahun 1997, dua ilmuwan komputer mengembangkan contoh konkret pembelajaran tanpa pengawasan menggunakan pengelompokan.<sup>2</sup> Idenya adalah untuk membantu bisnis pemberi kredit memprediksi cara membuat keputusan kelayakan kredit yang andal tentang pelanggan baru tanpa data klasifikasi sebelumnya tentang pelanggan dan informasi tentang riwayat kredit sebelumnya. Sebaliknya, mereka memanfaatkan data tentang pelanggan yang sudah ada. Ilmuwan AI menciptakan algoritma yang mengelompokkan pelanggan yang sudah ada ke dalam kelompok-kelompok berdasarkan faktor-faktor seperti penggunaan kartu kredit dan apakah mereka melunasi kartu tepat waktu. Dengan informasi ini, mereka menciptakan model yang membantu perusahaan-perusahaan ini mengetahui di mana seorang pelanggan akan berada dalam berbagai kelompok kelayakan kredit pada saat pendaftaran, tanpa akses ke peringkat kredit spesifik dari setiap pelanggan tertentu.

Pengelompokan data tanpa label, di mana komputer diminta untuk menentukan kategori tanpa manusia melihat data sebelumnya, disebut *pembelajaran tanpa pengawasan*. Dalam pembelajaran tanpa pengawasan, keluaran yang diinginkan tidak disediakan. Tujuan pembelajaran tanpa pengawasan adalah agar komputer mengungkap struktur dan pola yang melekat dalam data, atau hubungan di antara masukan. Biasanya, pembelajaran tanpa pengawasan digunakan untuk mengidentifikasi klaster dalam data yang ada, yang memungkinkannya mengkategorikan data masukan baru ke dalam salah satu klaster tersebut. Pembelajaran tanpa pengawasan juga digunakan untuk deteksi anomali, di mana masukan yang merupakan outlier – yaitu yang tidak termasuk dalam kelompok tempat sebagian besar data berada – dapat diidentifikasi. Deteksi anomali dapat digunakan dalam manufaktur untuk menentukan apakah suatu komponen memiliki bentuk yang diinginkan atau tidak, misalnya, untuk memeriksa apakah gigi pada roda gigi telah terkelupas.

Kategori tambahan yang dikenal sebagai *pembelajaran semi-supervised* memanfaatkan teknik yang digunakan dalam pembelajaran terawasi dan tanpa pengawasan. Pembelajaran semi-supervised sering digunakan ketika label untuk suatu dataset tidak lengkap atau rawan kesalahan. Dalam pembelajaran semi-supervised, algoritma menggunakan klaster, seperti yang dilakukan dalam pembelajaran tanpa pengawasan, tetapi juga menggunakan dataset kecil berlabel. Dengan demikian, ketika set berlabel ini muncul dalam klaster tertentu, algoritma memiliki informasi tambahan tentang sifat klaster tersebut. Algoritme ini tidak akan beroperasi dengan kepastian sempurna, tetapi dapat "mengenali" kemungkinan bahwa sampel lain dalam kelompok yang sama harus diberi label serupa. Tentu saja, pembelajaran semi-supervised memerlukan peninjauan dan analisis manual untuk memvalidasi atau membatalkan model yang dihasilkan. Saat ini, pembelajaran semi-supervised digunakan dalam aplikasi anti-pencucian uang dan deteksi penipuan.

### **Bilah Samping: ImageNet**

Pembelajaran terawasi dan tak terawasi banyak digunakan saat ini. Masing-masing dapat dan akan terus berlaku dalam keadaan tertentu. Namun, pembelajaran terawasilah yang

akan menghasilkan salah satu perkembangan besar dalam AI. Pada tahun 2009, basis data gambar yang ekstensif disediakan daring oleh para peneliti dari departemen ilmu komputer di Princeton. Setahun kemudian, sebuah kontes diluncurkan: ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Tujuannya adalah untuk melihat berapa banyak gambar yang dapat dikategorikan dengan tepat oleh suatu algoritma: yaitu, berapa banyak foto yang dapat diklasifikasikan dengan tepat. Gambar kucing harus masuk dalam kategori kucing, gambar anjing masuk dalam kategori anjing, dan seterusnya.

Empat ilmuwan komputer di Stanford, dua di antaranya adalah mahasiswa Hinton, mengikuti kontes pada tahun 2012, menerapkan arsitektur jaringan saraf dalam yang telah mereka kembangkan ke basis data ImageNet. Perangkat lunak mereka mengidentifikasi objek dengan akurasi dua kali lipat dari pesaing terdekat mereka. Peningkatan ini merupakan perkembangan yang mencengangkan, memberikan bukti konklusif bahwa jaringan saraf dalam dapat bekerja jauh lebih baik daripada metode AI sebelumnya. Tiga tahun kemudian, dalam kontes yang sama, algoritma AI baru tim Stanford melampaui kinerja manusia dalam tugas identifikasi untuk pertama kalinya. Di masa depan, ImageNet dapat dianggap sebagai versi Rosetta Stone masa kini untuk visi komputer.

Di Google, Ng terlibat dalam eksperimen inovatif lainnya. Timnya mengekspos jaringan saraf besar ke 10 juta gambar mini video YouTube yang tidak berlabel dan kemudian membiarkan algoritma mereka mempelajari cara mengidentifikasi kucing tanpa pengawasan. Ketika mereka menguji algoritma tersebut dengan data baru, algoritma tersebut mengidentifikasi kucing dengan benar sebanyak 74,8%. Pembelajaran tanpa pengawasan belum pernah digunakan dalam skala sebesar ini sebelumnya dan dengan tingkat keberhasilan seperti ini.

## 2.5 MEMBUAT DATA LEBIH BERMANFAAT

Kemampuan komputer untuk belajar tidak hanya bergantung pada seberapa banyak data yang tersedia. Kemampuan ini juga sangat bergantung pada bagaimana data tersebut direpresentasikan. Artinya, agar komputer dapat mengidentifikasi seekor burung secara akurat sebagai burung, entah itu gagak, burung hantu, pelikan, kuntul, atau ayam, komputer harus "mengetahui" apakah gambar yang dilihatnya tegak lurus atau terbalik, atau dalam kabut, salju, matahari, atau bayangan. Komputer harus mengetahui bahwa gambar seekor burung adalah burung meskipun hanya sebagian dari burung itu yang terlihat, mungkin hanya paruhnya. Komputer harus mengecualikan foto apa pun yang mungkin terlihat seperti burung, seperti hiasan kepala berbulu atau kemoceng. Analisis ini tidak mudah, itulah sebabnya (saat tulisan ini dibuat) tes daring untuk menentukan apakah Anda manusia mencakup berbagai gambar yang tidak lengkap dari berbagai hal seperti mobil atau rambu jalan.

Untuk mengimbangi beberapa tantangan yang dihadapi komputer dalam representasi data, ilmuwan komputer mengembangkan *rekayasa fitur*. Fitur adalah atribut atau properti data yang menjadi dasar analisis komputer. Rekayasa fitur dirancang untuk memilih fitur terbaik dari data, menciptakan representasi data terbaik sehingga mesin dapat belajar lebih efisien. Awalnya, tugas ini sulit dan mahal, terutama karena harus dilakukan "dengan tangan"

oleh ilmuwan AI. Saat ini, ada berbagai metode rekayasa fitur otomatis yang memfasilitasi penyesuaian data, sebuah perkembangan penting karena, terlepas dari waktu atau biaya, rekayasa fitur sering kali diperlukan untuk bekerja dengan data secara efektif.

Sebagai ilustrasi dari kebutuhan untuk ini, pertimbangkan perbedaan antara angka Romawi dan Arab - misalnya 10 versus X. Jika suatu masalah melibatkan representasi numerik, katakanlah, menambahkan kolom 10 angka besar, representasi angka mana yang akan menjadi pilihan ideal: Romawi atau Arab? Mana yang akan membuatnya lebih mudah, pada kenyataannya, mungkin, untuk perhitungan yang dilakukan oleh manusia? Dan agar komputer dapat menjumlahkan angka-angka tersebut, akan lebih mudah untuk mengubahnya ke representasi biner – jadi 10 akan menjadi 1010. Rekayasa fitur membantu menciptakan representasi data yang tepat untuk algoritma pembelajaran mesin dengan cara yang serupa, sehingga algoritma tersebut dapat melakukan pembelajaran komputasi yang menjadi tugasnya.

Di luar rekayasa fitur, *reduksi dimensionalitas* adalah cara untuk memilih fitur yang paling penting dari semua fitur yang tersedia, mengabaikan fitur yang mungkin bukan prediktor ideal untuk suatu hasil. Misalnya, jika seseorang ingin memprediksi seberapa besar kemungkinan calon pelanggan membeli produk baru, mengetahui nama pelanggan tersebut tidak akan memberikan hasil yang bermanfaat, tetapi mengetahui berapa banyak yang telah mereka belanjakan untuk produk tersebut di masa lalu kemungkinan besar akan bermanfaat.

Rekayasa fitur kurang diperlukan untuk model pembelajaran mendalam. Pembelajaran mendalam menghasilkan representasi data, sehingga manusia tidak perlu melakukan rekayasa fitur. Model-model ini menggunakan data pelatihan yang memadai untuk mencari tahu apa dalam data yang mengarah pada jawaban yang benar. Perlu diingat, seperti yang telah disebutkan sebelumnya, bahwa fitur yang digunakan dalam pembelajaran mendalam belum tentu memiliki konsep yang jelas yang sesuai dengan konsep yang mudah dipahami manusia. Misalnya, jaringan saraf pembelajaran mendalam yang mengenali gambar burung mungkin tidak memiliki fitur seperti kepala atau paruh. Jika Anda bertanya kepada salah satu sistem ini, "Apa yang membuat Anda memutuskan bahwa ini adalah gambar burung?", sistem tersebut mungkin akan menjawab dengan jawaban yang tidak dapat ditafsirkan manusia dalam hal ciri-ciri burung.

Itu bukanlah jawaban yang mencerahkan, sehingga mempelajari cara menggunakan sistem seperti ini dalam bisnis bisa sedikit rumit. Lebih sulit lagi untuk membenarkan mengapa keputusan individual dibuat oleh model pembelajaran mendalam, terutama dalam situasi yang diatur atau berhadapan langsung dengan pelanggan seperti ketika seorang pelanggan ingin tahu mengapa ia ditolak pinjaman hipotek. Sistem ini sering disebut sebagai kotak hitam, dan dapat menyebabkan berbagai masalah. (Kami akan membahas cara memitigasi masalah kotak hitam di Bab 10.) Membangun sistem yang dapat mengatakan "Saya menemukan 'kucing' karena ia memiliki telinga, mata, ekor, bulu, dan cakar" akan memiliki keuntungan nyata. Salah satu jenis sistem AI yang dapat melakukan hal ini saat ini menggunakan jenis AI yang berbeda dari pembelajaran mesin: pemodelan semantik dan penalaran kausal, atau singkatnya penalaran semantik.

## 2.6 PENALARAN SEMANTIK

Model pembelajaran mesin sampai pada kesimpulannya dengan mengenali pola atau korelasi antara masukan dan keluaran. Model ini menjawab pertanyaan seperti "Dari semua sel, mana yang bersifat kanker?" atau "Dari semua email, mana yang merupakan spam?" dengan mempelajari pola dari data berlabel dan memprediksi hasil untuk data yang tidak berlabel. Namun, dalam situasi tertentu, sistem ini kurang ideal. Sistem pembelajaran mesin membutuhkan sejumlah besar data untuk melatih algoritmanya – data yang mungkin tidak tersedia. Misalnya, model pembelajaran mesin yang digunakan untuk mencocokkan hasil dari berbagai protokol medis dengan anomali genetik pada masing-masing pasien akan membutuhkan banyak pasien dengan anomali tersebut untuk mencapai hasil yang bermakna, dan mungkin tidak banyak pasien yang tersedia. Sistem penalaran semantik, sebagian, dapat mengatasi kesulitan ini.

Model penalaran semantik tidak memerlukan jumlah data berlabel yang sama dengan pendekatan pembelajaran mesin, meskipun model ini membutuhkan seorang ontologis yang memiliki pemahaman mendalam tentang informasi dan aturan yang perlu ditangkap. Mereka adalah sistem berbasis aturan yang dapat memiliki informasi spesifik tentang entitas dan hubungannya, dan mereka memanfaatkan informasi ini untuk membuat inferensi. Dalam penalaran semantik, dunia dijelaskan melalui serangkaian konsep, deskripsinya, dan hubungan di antara mereka. Sistem penalaran semantik membuat inferensi dari sebab hingga akibat, memungkinkan mereka untuk menyimpulkan konsekuensi logis dari serangkaian fakta dan konsep yang diberikan. Model pengetahuan dan mesin inferensi memungkinkan sistem ini untuk memodelkan pengetahuan dengan fidelitas yang lebih tinggi dan mendukung penalaran yang lebih canggih saat menjawab pertanyaan.

Untuk mengilustrasikan cara kerjanya, kita dapat mengambil kalimat "John menonton *Mission Impossible* kemarin" dan mengubahnya menjadi model semantik (lihat Gambar 2.5) entitas atau konsep, atribut, dan hubungan. Entitas mewakili sesuatu di dunia nyata, misalnya seseorang. Demikian pula, konsep mewakili ide abstrak yang bukan entitas, misalnya waktu, suatu aktivitas. Atribut mewakili sesuatu tentang suatu objek atau konsep, seperti nama atau usia. Hubungan menghubungkan berbagai entitas dan konsep.

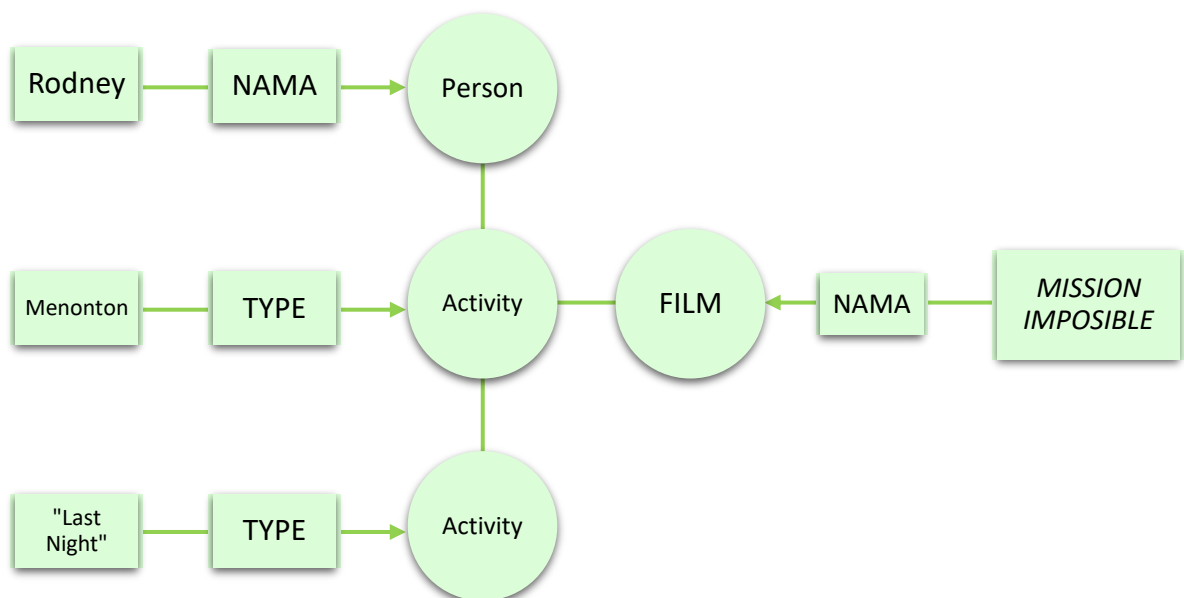
Penalaran semantik dapat menggeneralisasi di luar data yang dikumpulkan. Misalnya, penalaran semantik dapat mencakup konsep seperti kucing dan mamalia, dan fakta seperti "Sylvester adalah kucing hitam putih." Kemudian, penalaran semantik dapat menyimpulkan berbagai hal, seperti fakta bahwa Sylvester kemungkinan memiliki ekor dan dua telinga. Mengajari model pembelajaran mesin untuk mengenali, misalnya, dalam kondisi apa gen manusia mungkin mati sehingga seseorang dapat menciptakan obat yang dapat mencegah hal ini terjadi, secara pragmatis tidak layak menggunakan pembelajaran mesin saja, karena pembelajaran mesin didasarkan pada pengenalan pola atau korelasi, bukan kausalitas, dan korelasi tersebut terkadang tidak mudah diinterpretasikan oleh manusia.

Jalur biologis adalah serangkaian interaksi kausal antar molekul yang menyebabkan perubahan dalam sel, seperti mengaktifkan dan menonaktifkan gen. Pembelajaran mesin

memungkinkan Anda untuk melihat sejumlah besar data untuk menemukan interaksi molekuler mana yang paling umum ketika gen tertentu dimatikan, tetapi tidak akan memberi tahu Anda interaksi mana yang menonaktifkan gen tersebut. Pembelajaran mesin dapat menemukan korelasi tetapi bukan kausalitas. Karena ratusan ribu interaksi molekuler yang berkorelasi terjadi dan tanpa mengetahui jalur kausal yang tepat, akan sangat mahal untuk menciptakan obat yang dapat mencegah gen dinonaktifkan hanya dengan menggunakan pembelajaran mesin. Sebaliknya, dibutuhkan sistem AI yang dapat menangani penalaran semantik.

Jelas mengapa mengajarkan komputer untuk menggunakan penalaran semantik bisa sangat bermanfaat. Hal ini akan menghilangkan kebutuhan akan data yang sangat banyak atau waktu komputasi yang sangat besar, sehingga memungkinkan berbagai aplikasi, mulai dari terapi gen hingga memahami apa yang membuat sebuah kampanye pemasaran menjadi viral. Model pembelajaran mesin yang ada saat ini tidak dapat melakukan hal tersebut. Namun, model ini dapat memberi tahu Anda apakah terdapat pola di antara kampanye yang viral versus yang tidak viral.

Selain itu, terdapat area lain di mana penalaran semantik dapat meningkatkan hasil, misalnya, di bidang penegakan hukum, dengan menggunakan data lokasi dari TKP dan area sekitarnya, konteks di mana kejahatan dilakukan, pengetahuan kosakata yang canggih, dan konteks (perbedaan, misalnya, antara penangkapan dalam jargon polisi dan penangkapan di bidang medis). Kehalusan di bidang ini dan area lainnya akan sangat diuntungkan oleh pemodelan pengetahuan semantik.



**Gambar 2.5** Contoh jenis grafik pengetahuan.

Sejauh ini, hanya ada keberhasilan yang belum sempurna di bidang ini. Penalaran semantik mengalami beberapa kemajuan pada tahun 1980-an dan 1990-an, tetapi keberhasilannya masih minim. Sementara itu, kemajuan AI yang paling menonjol baru-baru ini

terus berasal dari pembelajaran mendalam. Itulah sebabnya sebagian besar buku ini akan berfokus pada pembelajaran mesin dan pembelajaran mendalam. Meskipun penalaran semantik umumnya tidak skalabel maupun fleksibel, penelitian menunjukkan bahwa penalaran semantik dapat digunakan bersama pembelajaran mesin dan pembelajaran mendalam untuk memungkinkan penalaran AI. Saat ini, kombinasi pembelajaran mendalam dan kondisi penalaran semantik saat ini mungkin terbukti menjadi pendekatan yang paling bermanfaat.<sup>5</sup>

Sistem saat ini sangat baik dalam memprediksi data yang ada pada domain tertentu. Sistem ini baru-baru ini telah digunakan dengan cukup sukses, misalnya, dalam keamanan siber, di mana algoritma mempelajari kode di komputer dan mampu mengetahui di mana kerentanan mungkin berada dan bagaimana cara memperbaikinya. Namun, sistem pembelajaran mesin kurang baik dalam abstraksi: menerapkan pengetahuan yang telah dipelajari dan menggunakannya di tingkat lain. Untuk melakukan ini, peneliti mungkin harus mengambil pendekatan hibrida. Pendekatan ini membutuhkan dua komponen kunci: model yang luas, mendalam, dan berfidelitas tinggi yang dapat digunakan komputer dari domain dan pengetahuan latar belakang yang relevan (data), serta mesin penalaran (inferensi) yang mampu menggabungkan pengetahuan ini secara efisien untuk menjawab pertanyaan atau mencapai kesimpulan.

Saat ini, meskipun beberapa aspek dari model pengetahuan tertentu dapat disimpulkan, sebagian besar konten pengetahuan harus dikurasi secara manual, yang membutuhkan perbaikan berkelanjutan, yang bisa sangat mahal. Model-model ini juga dapat bergantung pada pendapat masing-masing pakar, yang mungkin tidak lengkap atau bahkan salah. Paket perangkat lunak semantik yang tersedia bervariasi dalam hal kemampuan representasi dan fidelitasnya, serta dalam hal luas basis pengetahuan atau jumlah data yang dibutuhkan sebagai titik awal untuk pemodelan pengetahuan selanjutnya. Paket perangkat lunak ini menyediakan dasar untuk melakukan beberapa penalaran, tetapi terbatas dalam hal daya representasi, serta dalam hal pengetahuan domain yang harus dibangunnya.

Jalur lain menuju penalaran semantik mungkin berasal dari pembelajaran mendalam. Meskipun masih terlalu dini untuk mengetahui apakah ini akan berhasil, pembelajaran mendalam tetap menjadi bidang penelitian akademis yang aktif. Hanya dalam beberapa tahun terakhir, kemajuan besar telah dicapai di bidang pemrosesan bahasa alami. Kini terdapat model bahasa yang tidak hanya melakukan penambangan data yang jauh lebih baik daripada sebelumnya, tetapi juga tampaknya mensimulasikan jenis penalaran yang sangat mendasar. The New York Times baru-baru ini melaporkan bahwa untuk pertama kalinya, sebuah sistem AI mampu lulus ujian sains kelas 8 dan kelas 12.6 Aristo, sebutan untuk sistem ini, menunjukkan kemajuan di bidang ini.

#### **Bilah Samping: Cyc dan ConceptNet**

Pada awal 1990-an, para ilmuwan mulai mengembangkan cara-cara baru untuk merepresentasikan data dan hubungan antar elemen data. Cara-cara ini disebut kerangka kerja representasi. Di salah satu ujung spektrum terdapat bahasa seperti Unified Modeling Language (UML) dan Web Ontology Language (RDF/OWL), yang didasarkan pada standar

World Wide Web yang dikenal sebagai Resource Description Framework, yang merupakan salah satu kerangka kerja representasional ini. Masing-masing kerangka kerja pemodelan pengembangan yang bersifat umum ini dimaksudkan untuk menyediakan cara standar dalam merepresentasikan desain suatu sistem. Kerangka kerja ini dapat mencakup hal-hal seperti komponen individual suatu sistem (Anda dapat menganggapnya sebagai tampilan statis, atau kata benda, dari sistem) dan bagaimana komponen-komponen tersebut berinteraksi satu sama lain (tampilan dinamis, atau kata kerja, dalam sistem).

Di ujung spektrum yang lain adalah platform Cyc yang bergantung pada data (dibuat oleh Cycorp®). Cyc adalah proyek AI tertua di dunia, sebuah upaya untuk menciptakan pengumpulan data terbesar yang pernah ada yang menjelaskan cara kerja dunia tetapi berfokus pada hal-hal yang jarang ditulis atau dikatakan, yang berarti hal-hal tersebut tidak akan menjadi aliran data "standar". Jumlah data ini kemungkinan akan memungkinkan program untuk lebih fleksibel dalam situasi baru dan tak terduga.

Cyc menyediakan basis pengetahuan yang sudah ada yang berisi ratusan ribu konsep yang saling terkait melalui puluhan ribu jenis relasi. Cyc juga menyediakan mesin penalaran yang kuat dan efisien yang mendukung inferensi maju dan mundur serta penalaran deduktif, induktif, dan abduktif. Basis pengetahuan dan mesin penalaran Cyc memungkinkannya untuk menyediakan bahasa pemodelan logika tingkat tinggi yang mendukung fitur-fitur seperti relasi paritas tinggi (yaitu tidak terbatas pada pernyataan subjek-kata kerja-objek), dan kemampuan untuk membuat pernyataan (pernyataan) tentang pernyataan lain: sebuah mekanisme yang ampuh untuk menempatkan pengetahuan dalam konteks.

Pendekatan lain untuk menciptakan basis pengetahuan manusia yang luas agar komputer dapat bekerja adalah proyek Open Mind dari MIT, yang mengumpulkan jutaan fakta tentang dunia secara crowdsourced dalam bahasa Inggris, alih-alih dalam representasi formal yang dapat digunakan komputer. Idennya adalah untuk mengkonsolidasikan potongan-potongan pengetahuan umum ini, yang menurut MIT terdiri dari apa yang kita sebut "akal sehat", dan mengubahnya menjadi bentuk yang dapat digunakan komputer. Pengetahuan ini (sebagian) ditangkap dalam ConceptNet, sebuah grafik ekstensif berisi konsep-konsep yang saling terkait yang dapat mendukung beberapa tugas, seperti pencarian semantik. Pencarian semantik mengacu pada peningkatan akurasi pencarian dengan memahami maksud pencari menggunakan makna kontekstual istilah-istilah dalam ruang data.

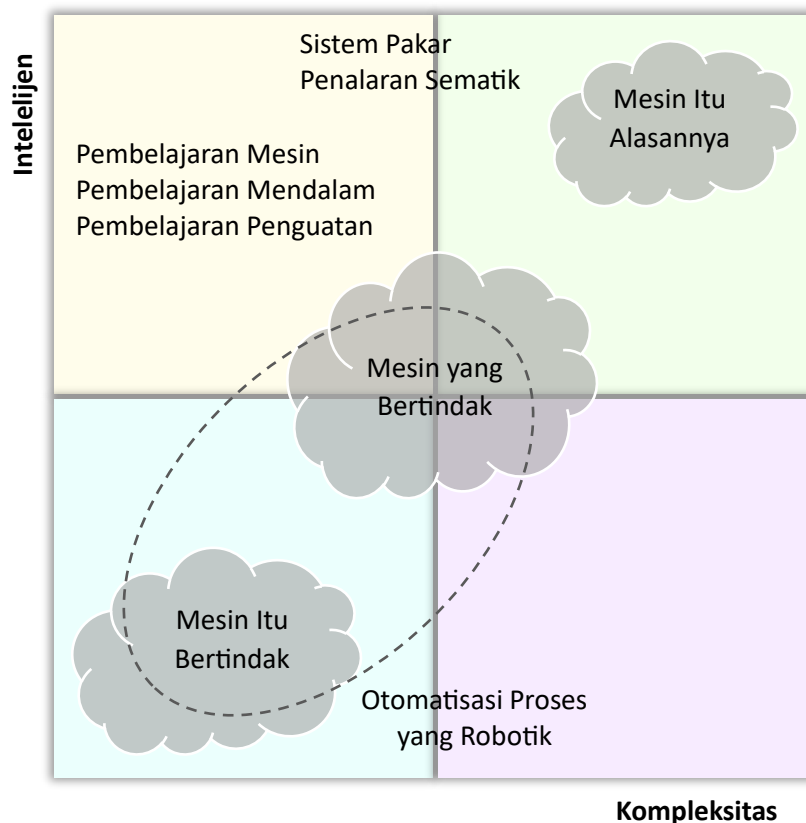
## 2.7 PENERAPAN AI

Secara umum, penerapan AI oleh perusahaan dapat dikelompokkan menjadi tiga area. Beberapa perusahaan *menghilangkan tugas-tugas berulang* bagi karyawan – membebaskan pekerja dari tugas-tugas pengetahuan yang repetitif dan membosankan yang hanya membutuhkan sedikit upaya kognitif. Otomatisasi ini meningkatkan penghematan biaya dan akurasi data, dan seringkali dilakukan menggunakan otomatisasi proses robotik (RPA), yang menggunakan aturan untuk meniru apa yang akan dilakukan pengguna pada keyboard di berbagai aplikasi bisnis.

Perusahaan lain *menghasilkan wawasan* – mengekstrak pengetahuan dan wawasan yang dapat ditindaklanjuti, non-sepele, dan sebelumnya tidak diketahui dari data terstruktur dan tidak terstruktur. Hal ini seringkali dilakukan menggunakan pembelajaran mesin dan pembelajaran mendalam. Wawasan ini kemudian ditindaklanjuti untuk mendapatkan nilai darinya. Perusahaan lain juga *meningkatkan kecerdasan manusia* dengan menyediakan pengetahuan dan dukungan kontekstual, untuk membantu pelanggan dan karyawan melakukan tugas dengan cara yang semakin mudah dan efektif. Hal ini seringkali dilakukan menggunakan asisten virtual atau proses yang terintegrasi dalam konteks yang sesuai dalam aplikasi dan alur kerja yang ada.

Anda dapat menganggap ketiga aplikasi AI ini termasuk dalam kategori mesin yang bertindak, mesin yang belajar, dan mesin yang bernalar (lihat Gambar 2.6). Saat ini, hanya ada sedikit sistem yang dapat mendekati penalaran kecuali dalam arti yang sederhana. Selama lima tahun terakhir, banyak inovasi dalam AI, pembelajaran mesin, dan pembelajaran mendalam telah merambah bisnis. Selain itu, mayoritas adopsi perusahaan terjadi pada mesin yang bertindak (misalnya RPA) dan mesin yang belajar.

Di buku ini, kita akan melihat beberapa kasus penggunaan di industri tertentu untuk memberikan gambaran tentang beragam aplikasi AI yang digunakan saat ini. Penggerak industri di balik kasus penggunaan AI yang akan kita lihat meliputi pengurangan biaya dengan mengotomatiskan tugas, pengurangan risiko melalui prediksi yang lebih akurat, peningkatan layanan pelanggan, pencapaian kepatuhan terhadap berbagai peraturan nasional dan internasional, serta peningkatan pendapatan.



**Gambar 2.5 Jenis sistem AI.**

## **BAB 3**

### **AI DALAM E-COMMERCE DAN RITEL**

Bisnis e-commerce dan ritel konvensional saat ini memanfaatkan wawasan yang diperoleh dari algoritma AI yang canggih untuk meningkatkan dan mengembangkan perusahaan mereka. Hal ini menghasilkan kolaborasi yang produktif antara divisi penjualan, layanan pelanggan, serta periklanan dan pemasaran, yang semakin banyak berbagi data, platform AI, dan tim analitik. Data dari periklanan, pemasaran, transaksi pelanggan, dan layanan pelanggan juga diintegrasikan ke dalam fungsi rantai pasokan untuk meningkatkan peramalan dan pemenuhan permintaan, serta optimalisasi pengembalian. Seiring waktu, kemungkinan akan ada integrasi yang lebih signifikan dari fungsi-fungsi ini. Berikut ini adalah beberapa area di mana model AI berhasil digunakan dalam bisnis e-commerce dan ritel.

#### **3.1 PERIKLANAN DIGITAL**

Mungkin keberhasilan terbesar penggunaan AI dalam skala besar adalah dalam periklanan dan pemasaran digital. Pada tahun 2019, untuk pertama kalinya, bisnis akan menghabiskan lebih banyak uang untuk periklanan dan pemasaran digital daripada media tradisional, seperti TV, radio, dan koran.<sup>1</sup> Banyak inovasi AI berasal dari pengiklan terbesar – Google, Facebook, Alibaba, dan Amazon. Algoritma AI telah membantu perusahaan yang memanfaatkan periklanan digital dalam berbagai bentuknya: iklan pencarian bersponsor, iklan kontekstual, iklan bergambar, dan lelang penawaran waktu nyata. Model AI secara akurat dan cepat memprediksi rasio klik-tayang iklan berdasarkan segmen pelanggan, yang menciptakan dasar untuk nilai iklan. Meskipun sebagian besar fokusnya adalah pada periklanan digital, periklanan tradisional juga telah diuntungkan oleh AI – misalnya, beberapa perusahaan menggunakan model segmentasi pelanggan berbasis AI dari iklan digital untuk membantu menginformasikan periklanan tradisional seperti iklan TV.

Ada dua cara utama untuk beriklan online: selama pencarian, ketika pengguna memilih untuk melihat iklan, dan di situs web dan aplikasi seluler. Ketika seseorang menggunakan mesin pencari, mesin pencari tersebut mencocokkan serangkaian iklan potensial dengan pencarian tersebut berdasarkan kata kunci spesifik yang dipilih pengiklan. Mekanisme lelang kemudian memilih iklan mana yang akan ditampilkan kepada pengguna tersebut dan menentukan tarif yang akan dibayarkan pengiklan kepada mesin pencari untuk menampilkan iklan tersebut. Kemampuan memprediksi apakah iklan yang ditampilkan akan diklik menjadikan AI komponen penting dari lelang ini karena model AI memiliki kemampuan prediktif yang lebih baik daripada metode apa pun yang digunakan sebelumnya. Memilih iklan mana yang akan ditampilkan di situs web bekerja dengan cara yang serupa, tetapi alih-alih mengaitkan iklan dengan kueri pencarian, iklan ditargetkan kepada pengguna yang menjelajahi situs, berdasarkan demografi dan informasi lainnya.

Periklanan digital real-time melibatkan banyak aktor: penerbit, seperti majalah mobil online yang ingin menjual ruang iklan di situs web mereka; pengiklan, seperti bank yang ingin

menampilkan iklan untuk pinjaman mobil; dan serangkaian penyedia layanan yang membantu melakukan pencocokan. Penerbit menugaskan platform sisi penawaran (SSP) untuk menawarkan ruang iklan ke bursa iklan. Bursa iklan bertindak sebagai pasar tempat terjadinya pencocokan penawaran-permintaan (atau ruang iklan). Pengiklan menggunakan platform sisi permintaan (DSP) untuk bersaing dan menawar ruang iklan yang tersedia, memanfaatkan AI untuk mencapai rasio klik-tayang dan rasio konversi yang tinggi. Platform manajemen data (DMP) menawarkan profil pemirsa individual dan menjual data minat untuk mendukung keputusan dalam proses lelang.

Dalam periklanan digital waktu nyata (real-time), ruang atau slot iklan individual terjual dalam beberapa milidetik setelah pemirsa mengeklik situs web: misalnya, ketika pemirsa mengeklik situs majalah mobil. Saat halaman web dimuat dari server, permintaan diajukan ke SSP untuk slot tempat iklan perlu ditampilkan. SSP menggunakan spesifikasi ruang iklan, seperti ukuran dan posisi slot, harga minimum, jenis iklan yang diizinkan untuk ditampilkan, dan informasi pengguna yang relevan, untuk memicu lelang di bursa iklan. Bursa iklan menentukan DSP mana yang mungkin sesuai untuk slot ini dan meneruskan permintaan lelang kepada mereka. DSP kemudian memutuskan kampanye pengiklan mana yang relevan. Mereka biasanya mempertimbangkan semua informasi yang tersedia tentang pemirsa saat ini. Jika terdapat cookie, perangkat, atau pengenalan pengguna lainnya, DSP akan mencari informasi tambahan tentang pengguna dari DMP, seperti minat atau demografi pengguna. Informasi ini kemudian digunakan untuk membuat keputusan dalam proses penawaran.

Ketika bursa iklan menerima semua tawaran, ia akan menetapkan pemenang. DSP yang menang mengirimkan lokasi server iklan untuk media tersebut, seperti gambar atau video iklan, ke bursa. Informasi ini kemudian dikirim kembali ke halaman web, yang masih dimuat, karena semua aktivitas ini hanya berlangsung dalam hitungan milidetik. Server iklan menayangkan media dan melacak semua aktivitas pengguna selanjutnya yang terkait dengan materi iklan. DSP juga menggunakan informasi pelacakan pengguna ini untuk iklan mendatang yang akan ditawarkan.

Penggunaan model AI memungkinkan pengiklan untuk membuat keputusan tentang apakah akan menawar ruang iklan berdasarkan respons pengguna yang diprediksi. Inilah sebabnya sebagian besar pengiklan online menggunakan sistem prediksi klik. Prediksi sering kali didasarkan pada segmentasi pelanggan pengguna, yang dilakukan menggunakan pengelompokan berbasis pembelajaran mesin. AI sangat berguna dalam ruang iklan digital karena perilaku pelanggan sangat beragam dan kuat, dan algoritma AI adaptif secara unik mampu memprediksi perilaku berdasarkan model yang dipelajari dari data historis.

### **3.2 PEMASARAN DAN AKUISISI PELANGGAN**

E-commerce ritel tumbuh pesat; nilainya melampaui Rp. 500 Triliun pada tahun 2018. Sebagian besar pertumbuhan ini disebabkan oleh AI dan big data. Beberapa penggunaan AI dalam pemasaran untuk mendapatkan pelanggan baru meliputi kampanye pemasaran berbasis AI untuk menargetkan calon pelanggan, serta pemodelan peningkatan dan atribusi untuk mengoptimalkan pengeluaran pemasaran.

Di masa lalu, menentukan calon pelanggan mana yang menjelajahi situs e-commerce yang akan berkonversi dan menghasilkan pendapatan terkadang sulit. Jika calon pelanggan meninggalkan situs tanpa membeli, sulit untuk membawanya kembali ke corong pembelian. Umumnya, penargetan ulang generik telah digunakan, di mana iklan untuk produk pengecer ditampilkan kepada semua individu yang telah berinteraksi dengan situs pengecer. Namun, saat ini, teknologi "penginderaan calon pelanggan" berbasis AI dan pemasaran berbasis AI membuat penargetan dan penargetan ulang lebih tepat – tidak hanya di ritel, tetapi juga untuk e-commerce di seluruh industri.

Kunci sukses dalam penginderaan prospek adalah kemampuan untuk mengumpulkan data prospek dan pelanggan dan menindaklanjuti data tersebut dengan tepat. Kita menyebut seseorang sebagai *prospek* jika kita tidak tahu siapa mereka tetapi tahu bahwa mereka mungkin *pelanggan* potensial; pelanggan adalah seseorang yang masuk atau yang telah membeli dari pengecer di masa lalu. Untuk prospek, sebagian besar interaksi online, seperti penelusuran atau iklan digital, dilacak melalui pengenalan perangkat (ID) pada aplikasi seluler atau sumber daya pelacakan lain seperti pengenalan yang masuk, cookie, skrip, piksel, atau gambar pada browser. Dalam sebuah opini *New York Times*, Farhad Manjoo membahas eksplorasinya tentang seberapa banyak pelacakan digital terjadi melalui situs web dan menemukan "semua yang Anda lakukan online dicatat dengan sangat detail" dan dia "terpukau oleh skala dan detail pelacakan tersebut." Data pelacakan diilustrasikan dengan indah oleh Nadieh Bremer.

Jika pengguna adalah pelanggan tetap, sebagian besar perusahaan e-commerce akan menyimpan ID pelacakan di basis data pelanggan, bersama dengan email dan informasi lainnya. Bahkan jika pelanggan menjelajah tanpa masuk, ID pelacakan mereka saat ini dapat dicocokkan dengan yang ada di sistem manajemen hubungan pelanggan (CRM) untuk mengetahui siapa mereka, dan situs atau aplikasi dapat dipersonalisasi berdasarkan informasi ini. Jika pengguna adalah calon pelanggan yang menjelajahi aplikasi atau situs web e-commerce bisnis secara anonim, hanya informasi penjelajahan sebelumnya yang terkait dengan ID pelacakannya yang dapat digunakan. ID pelacakan ini penting dalam mengumpulkan data yang cukup untuk mempelajari dan memprediksi perilaku calon pelanggan atau pelanggan. ID pelacakan juga digunakan dalam menayangkan iklan digital ke perangkat atau peramban calon pelanggan tersebut, tanpa memerlukan informasi identitas pribadi (PII).

Jenis data yang dikumpulkan saat calon pelanggan menggunakan aplikasi atau mengunjungi situs mencakup data tentang asal pengguna (misalnya, dari iklan atau dengan menggunakan mesin pencari) dan apa yang dilakukannya saat berada di situs tersebut. Jika sebuah iklan ditayangkan di ponsel pintar, informasi lokasinya dapat dikumpulkan, jika ia telah memberikan akses ke aplikasi atau peramban untuk mengaktifkan pelacakan ini. Beberapa prospek ini akhirnya menjadi pelanggan, meninggalkan jejak digital jalur mereka menuju konversi.

Peristiwa konversi itu sendiri – yaitu, apakah pembelian telah dilakukan – dapat dianggap sebagai label untuk pembelajaran mesin yang diawasi. Perusahaan menggunakan data yang relevan untuk menghasilkan linimasa interaksi yang komprehensif dengan setiap

pelanggan, yang dikenal sebagai *perjalanan pelanggan*. Peta perjalanan ini memungkinkan kemampuan prediktif yang jauh lebih baik dan cara-cara baru dan produktif untuk menangani prospek dan pelanggan – misalnya, menentukan tidak hanya siapa yang menjadi target tetapi juga kapan dan melalui saluran mana untuk berkomunikasi. Sangat penting untuk membantu menjangkau pelanggan dengan pemikiran perjalanan adalah model AI di balik konten dinamis dan pengalaman terbaik berikutnya.

Banyak perusahaan memanfaatkan data ini dalam kampanye pemasaran berbasis AI mereka untuk memprediksi perilaku pengguna secara lebih akurat, menggunakan model AI untuk membuat pesan dan penawaran yang dipersonalisasi, yang meningkatkan pendapatan dan laba atas investasi (ROI) pemasaran. Akibatnya, pemasaran beralih dari model pasar massal, satu ukuran untuk semua, menuju ideal satu-ke-satu atau satu-ke-Anda yang kini dimungkinkan oleh kemampuan AI dan ketersediaan data. Misalnya, perusahaan e-commerce menggunakan AI untuk belajar dari data yang disebutkan sebelumnya dan, berdasarkan perilaku prospek baru, dapat memprediksi apakah prospek tersebut kemungkinan akan membeli atau tidak. Jika calon pelanggan kemungkinan besar tidak akan berkonversi, menghindari penargetan ulang dapat menghemat pengeluaran pemasaran. Jika calon pelanggan diperkirakan akan membeli tetapi tidak menyelesaikan transaksi, ia dapat ditargetkan ulang secara lebih tepat dengan iklan. Dengan cara ini, AI dapat mendorong prospek yang relevan kembali ke dalam perjalanan pembelian.

Penargetan calon pelanggan yang belum menjadi pelanggan terutama dimungkinkan melalui data tayangan iklan milik pihak ketiga dan individu, seperti riwayat tayangan iklan yang dijelaskan sebelumnya. Bayangkan seorang wanita yang sedang mencari TV layar lebar. Pertama, ia mengunjungi Google dan mengetikkan kueri penelusuran "TV layar lebar terbaik". Google mencantumkan situs-situs yang menyediakan informasi tentang TV, seperti TechRadar atau CNET. Peritel yang menjual TV memasang iklan di situs-situs ini. Dengan memasang iklan, calon pelanggan dilacak dengan ID cookie atau, jika mereka menggunakan aplikasi seluler, ID perangkat. Meskipun calon pelanggan tidak mengklik iklan untuk masuk ke situs yang diiklankan, peritel yang memenangkan tawaran untuk memasang iklan tersebut sudah mengetahui sesuatu tentangnya karena iklan tersebut telah dimuat atau dilihat. Peritel kini dapat memasang kembali iklan yang lebih personal di CNET atau TechRadar agar ia dapat melihatnya saat ia online lagi. Jika ia mengunjungi situs e-commerce peritel tersebut, peritel dapat mempersonalisasi situs tersebut untuk menyorot TV, khususnya TV layar lebar.

Dengan tidak adanya data individual, atau terkadang bersamaan dengannya, banyak perusahaan menggabungkan data pihak ketiga lainnya seperti demografi atau cuaca pada tingkat segmen pelanggan, termasuk data seperti tingkat pendapatan rata-rata berdasarkan kode pos. Misalnya, pertimbangkan dealer mobil yang ingin menjual mobil kepada calon pelanggan. Dealer tersebut memiliki beberapa data pihak pertama tentang perilaku pelanggan mereka sebelumnya. Namun, tidak ada informasi yang tersedia tentang pendapatan atau kebiasaan belanja calon pelanggan. Agar dapat mengirimkan selebaran tertarget yang menawarkan penawaran paling tepat kepada calon pelanggan di berbagai bagian kota, dealer tersebut memperoleh data sensus yang tersedia untuk umum. Dealer tersebut bahkan dapat

memutuskan untuk membeli data tentang jenis mobil yang diparkir di lingkungan tempat mereka memperoleh data sensus ini. Hal ini memungkinkan dealer untuk menggunakan model AI guna menentukan lingkungan terdekat mana yang sangat sensitif terhadap harga dan merek serta model mobil apa yang kemungkinan akan dikendarai oleh penduduk. Dealer kemudian menargetkan calon pelanggan di area tersebut dengan penawaran khusus untuk merek dan model mobil tertentu.

Perusahaan seperti Publicis Epsilon, Acxiom, Liveramp, dan Experian menyediakan layanan yang mengambil data pengguna dari suatu bisnis, menghubungkannya dengan data pengguna lain dari situs lain, dan mengembalikan set data lengkap ini ke perusahaan, menghapus semua PII. Set data terkonsolidasi ini memungkinkan peritel untuk mengetahui lebih banyak tentang apa yang dijelajahi dan dilihat oleh calon pelanggan. Tentu saja, "dia" tidak diidentifikasi - terkadang disebut sebagai nama samaran. Namun, jika ID pelacaknya muncul di masa mendatang, peritel akan menampilkan iklan spesifik yang disesuaikan dengan profilnya.

Dalam pemasaran, AI juga digunakan untuk *pemodelan peningkatan* untuk mengoptimalkan pengeluaran pemasaran. Pemodelan peningkatan menentukan pelanggan mana yang harus ditargetkan dengan suatu penawaran dan pelanggan mana yang tidak dengan memprediksi perkiraan perubahan tingkat konversi karena kampanye tertentu. Misalkan seorang peritel merencanakan kampanye pemasaran untuk meningkatkan konversi dengan menawarkan diskon untuk suatu produk. Setiap pelanggan mendapatkan skor peningkatan yang mewakili peningkatan probabilitasnya melakukan pembelian jika ia menerima diskon. Jika pelanggan sudah memiliki kemungkinan pembelian yang sangat tinggi, diskon akan menurunkan pendapatan karena kemungkinan besar ia akan tetap membeli. Jika ia memiliki skor peningkatan yang tinggi tetapi kemungkinan pembeliannya rendah, diskon akan meningkatkan pendapatan. Jika ia memiliki skor peningkatan yang rendah, menargetkannya dengan penawaran dan iklan kemungkinan besar akan membuang-buang anggaran pemasaran. Jika pelanggan memiliki skor peningkatan negatif, yaitu jika ia cenderung tidak berkonversi karena kampanye, maka menargetkan pelanggan khusus ini akan kehilangan pendapatan, baik karena hal itu akan membuatnya menjauh maupun karena anggaran pemasaran akan dibelanjakan secara tidak perlu. Dalam beberapa kasus, peningkatan digunakan untuk memodelkan optimasi harga. Misalnya, banyak perusahaan memberikan diskon untuk produk mereka. Anggaran untuk insentif ini seringkali 10 kali lebih besar daripada anggaran pemasaran, dan mengoptimalkannya memiliki dampak yang signifikan.

Aspek lain dari pengelolaan anggaran pemasaran adalah memutuskan berapa banyak yang akan dialokasikan untuk berbagai kampanye dan saluran. Untuk melakukan ini, pemasar harus mengetahui berapa banyak pendapatan yang dihasilkan oleh kampanye dan saluran sebelumnya dan mengaitkan pendapatan yang dihasilkan dengan pengeluaran pemasaran digital mereka. Menentukan atribusi pemasaran ini bisa jadi sulit bagi banyak tim pemasaran digital. Masalahnya sering kali disebabkan oleh independensi berbagai komponen platform untuk kampanye lintas kanal, yang mengakibatkan data berada di beberapa sistem yang tidak terkait, sehingga menyulitkan tim untuk melacak ROI dengan yakin. Menentukan model

atribusi dan seberapa besar bobot yang akan diberikan pada kanal atau titik kontak mana dalam perjalanan pelanggan merupakan tantangan besar tanpa data yang memadai. Meremehkan satu sumber dalam model atribusi dan melebih-lebihkan sumber lain dapat menyebabkan keputusan pengeluaran yang buruk dan hasil pemasaran yang buruk. Dengan menggunakan data perjalanan pelanggan omni-channel, banyak perusahaan menggunakan AI untuk membuat model yang lebih akurat guna mengatribusikan hasil ke kanal tertentu.

### 3.3 CROSS-SELLING, UP-SELLING, DAN LOYALITAS

Bagi calon pelanggan yang telah berkonversi, tersedia lebih banyak data bermanfaat yang dapat digabungkan dengan data pihak ketiga, dengan mempertimbangkan privasi dan peraturan. Wawasan yang diperoleh sejak pengguna tersebut menjadi calon pelanggan akan disalin ke profil pelanggan setelah ia menjadi pelanggan. Data ini kemudian dikumpulkan dalam danau data yang disebut *platform data pelanggan*, yang kemudian digunakan untuk mengembangkan model AI.

Di masa lalu, penargetan pelanggan tidak cukup terperinci untuk prediksi dengan akurasi tinggi. Paling banter, hal ini didasarkan pada segmentasi pelanggan yang sangat rinci. Secara tradisional, segmentasi pelanggan berfokus pada pembagian basis pelanggan ke dalam kelompok-kelompok yang memiliki kesamaan spesifik yang relevan untuk pemasaran yang ditargetkan, seperti usia, jenis kelamin, dan lokasi. Kini, dengan lonjakan data yang tersedia seperti riwayat penelusuran, tayangan iklan, dan media sosial, perusahaan menggunakan algoritma AI untuk menciptakan *segmentasi pelanggan* yang lebih rinci dan lebih spesifik terhadap pelanggan. Segmentasi ini kemudian digunakan sebagai input dalam model lain untuk akurasi personalisasi yang lebih tinggi.

*Pemodelan nilai seumur hidup pelanggan* adalah cara lain untuk mengelompokkan pelanggan, berdasarkan jumlah pendapatan atau laba yang akan dihasilkan pelanggan bagi perusahaan selama hubungan tersebut. Model-model ini memanfaatkan pembelajaran mesin untuk memprediksi pembelian di masa mendatang berdasarkan riwayat pembelian, membantu memprioritaskan rekomendasi dan penawaran. Hal ini memungkinkan bisnis untuk mengonversi pelanggan dengan nilai tertinggi: mereka yang akan loyal dan membeli lebih banyak melalui cross-selling dan up-selling di masa mendatang.

Bisnis e-commerce dan ritel menggunakan model AI untuk memprediksi perilaku pelanggan seperti kecenderungan pembelian: menilai orang berdasarkan faktor-faktor yang paling prediktif terhadap pembelian dalam periode tertentu. Model AI untuk memprediksi perilaku pelanggan ini disesuaikan dengan industri tertentu dan kemudian disegmentasikan berdasarkan hal-hal seperti perilaku dalam industri tersebut, daya beli pelanggan, pangsa dompet, dan pendapatan. Informasi yang dihasilkan digunakan untuk personalisasi, penargetan cross-sell dan up-sell, dan ajakan bertindak, sehingga meningkatkan konversi penjualan. Model seperti *kecenderungan untuk membeli* dan data yang mendasarinya juga diperluas untuk meningkatkan area bisnis lainnya, yang memungkinkan tidak hanya pengoptimalan pengeluaran iklan pada pelanggan yang paling mungkin melakukan pembelian,

yang meningkatkan ROI, tetapi juga membantu meningkatkan perkiraan permintaan dalam perencanaan rantai pasokan.

Salah satu pendekatan untuk meningkatkan keterlibatan pelanggan adalah menggunakan *personalisasi situs* untuk menyediakan versi unik situs web atau aplikasi bagi setiap pelanggan. Banyak elemen situs yang dapat dikustomisasi, seperti gambar atau video banner, warna, konten yang ditampilkan, dan rekomendasi produk atau layanan; algoritma yang disebut *mesin rekomendasi* menyediakan konten yang dipersonalisasi yang ditampilkan atau produk yang ditawarkan kepada setiap pengguna. Sebagian besar mesin rekomendasi menggunakan serangkaian teknik algoritmik yang disebut *pemfilteran kolaboratif*. Teknik ini menganalisis data berdasarkan kategori yang berfokus pada kesamaan di antara calon pelanggan dan antar produk. Seseorang dianggap serupa jika, misalnya, mereka membeli banyak produk yang sama, atau memiliki beberapa karakteristik yang sama, seperti demografi, minat, dan riwayat belanja. Produk dikategorikan serupa jika pembeli cenderung membelinya bersamaan, misalnya, pel dan pembersih lantai.

Model AI mencari aspek apa dari profil lengkap pengguna (termasuk transaksi, perilaku penelusuran, dan wawasan apa pun dari data pihak ketiga) yang cocok dengan jenis konten yang mereka habiskan waktunya. Dalam skenario di mana mungkin tidak ada data historis, seperti desain individual baru untuk situs web, *pengujian A/B* digunakan. Pengujian A/B, atau lebih umum pengujian multivariat, adalah cara untuk membandingkan variasi alternatif situs (atau versi alternatif model AI) dengan menunjukkan setiap variasi kepada pengunjung yang berbeda secara acak untuk menentukan mana yang berkinerja lebih baik dengan tujuan tertentu, seperti konversi.

Meskipun penyaringan kolaboratif menggunakan pembelajaran mendalam telah meningkatkan akurasi prediksi, algoritma AI yang lebih modern telah muncul, seperti model lebar dan dalam, yang jauh lebih menjanjikan dalam hal akurasi. Model jaringan saraf ini menggabungkan pemahaman tentang interaksi pengguna dan produk (jaringan luas) dan pemahaman yang mendalam tentang karakteristik produk (jaringan dalam). Dengan menggunakan mesin rekomendasi berbasis AI, perusahaan memprediksi produk apa yang paling mudah dibeli oleh pelanggan – dan merekomendasikan produk-produk ini melalui saluran daring atau pesan langsung. Dengan algoritma yang tepat, tidak sulit bagi, misalnya, peritel pakaian untuk menentukan produk mana yang mungkin diminati pelanggan dengan melihat kemiripan antara pelanggan ini dan pelanggan sebelumnya yang telah membeli barang serupa.

Meningkatkan peluang cross-sell dan up-sell memiliki potensi keuntungan yang sangat besar. Dengan menggunakan berbagai teknik AI yang memanfaatkan prediksi niat pelanggan dan mesin rekomendasi, peritel dapat menciptakan langkah optimal berikutnya yang paling mungkin dalam proses penjualan untuk meningkatkan penjualan produk tambahan.

Banyak perusahaan menggunakan mesin rekomendasi untuk menciptakan siklus positif yang terus meningkat melalui eksperimen dan eksplorasi berulang. Data berkualitas menghasilkan wawasan tentang bagaimana pengguna memperhatikan atau mengabaikan produk atau mengikuti rekomendasi; perusahaan memanfaatkan wawasan ini untuk terus

meningkatkan siklus positif ini. Keputusan yang dihasilkan dapat mencakup apakah perusahaan ingin menetapkan harga atau menggabungkan produk secara berbeda, atau meningkatkan atau menghilangkan fitur tertentu dari suatu produk sepenuhnya. Hasil dari keputusan ini memasuki siklus sebagai data dan pada gilirannya menghasilkan wawasan baru, dan seterusnya.

Algoritma AI juga membantu perusahaan memprediksi dan mengurangi *churn pelanggan*. Algoritma ini mengidentifikasi pelanggan yang tidak puas sejak dini, memberi perusahaan kesempatan untuk menawarkan insentif agar mereka tetap berlangganan. Insentif tersebut dapat mencakup peningkatan, fitur gratis, atau diskon untuk layanan di bulan mendatang. Kami akan menggunakan churn pelanggan sebagai contoh detail di Bab 13, dengan melihat catatan historis pelanggan dan membangun serta melatih model pembelajaran mesin untuk memprediksi perilaku pelanggan baru. Kasus penggunaan yang dekat dengan churn pelanggan adalah *model pengisian ulang*. Model-model ini memprediksi produk apa yang telah dibeli pelanggan yang mungkin akan habis dan kapan. Sebelum diprediksi kehabisan, mereka akan dikirimkan pengingat untuk segera mengisi ulang produk tersebut agar pembelian tidak dilakukan dengan produk atau peritel alternatif.

### 3.4 KECERDASAN PELANGGAN BISNIS-KE-BISNIS

Perusahaan bisnis-ke-konsumen (B2C) bukan satu-satunya yang mengadopsi AI untuk meningkatkan kinerja; perusahaan bisnis-ke-bisnis (B2B), mulai dari yang menjual produk ke perusahaan hingga perusahaan jasa keuangan dan energi, juga ikut serta. Secara historis, tenaga penjualan B2B mengandalkan pendekatan berbasis intuisi dengan menggunakan data terbatas (sering kali dalam bentuk spreadsheet), melacak wawasan mereka secara manual dalam proses yang memakan waktu. Namun, banyak B2B telah mendigitalkan penjualan mereka menggunakan solusi cloud dari Salesforce dan Microsoft.

Seperti kunjungan ritel fisik, setiap kali calon pembeli mengunjungi properti digital penjual, penjual tersebut memperoleh informasi tentangnya. Informasi ini mencakup asal pengunjung dan detail perusahaannya, seperti nama dan ukuran. Perusahaan yang mempekerjakan pengunjung diidentifikasi menggunakan pencarian terbalik alamat IP-nya. Penjual juga dapat melacak aktivitas pengunjung di situs dengan mengumpulkan informasi tentang waktu yang dihabiskan pengunjung di halaman situs dan aktivitas yang dilakukannya di sana, sehingga memberikan wawasan yang lebih mendalam tentang minat dan niatnya. Misalnya, jika pengunjung menghabiskan sebagian besar waktunya untuk membandingkan produk atau layanan, kemungkinan besar ia sudah lebih maju dalam riset dan kemungkinan besar akan membeli. Bahkan seberapa jauh ia menggulir halaman ke bawah menunjukkan posisi pengunjung dalam perjalanan riset dan pembeliannya.

Semua informasi ini dikumpulkan untuk setiap orang dari perusahaan tertentu yang mengunjungi situs web penjual. Jika perusahaan tersebut merupakan pelanggan B2B, informasi tambahan tentang perusahaan akan digunakan: informasi seperti riwayat pembelian, kondisi penggunaannya, layanan dan produk yang digunakan, bagaimana pengguna berinteraksi dengan aplikasi seluler jika tersedia, dan hasil kunjungan sebelumnya.

Perusahaan B2B menggabungkan kecerdasan pelanggan semacam ini berdasarkan perusahaan, memungkinkan penjual untuk memiliki pandangan holistik tentang perilaku semua pengguna dari bisnis tertentu. Data yang dihasilkan memberikan wawasan tentang lingkungan bisnis dan niat pengguna, memungkinkan personalisasi situs sehingga pengalaman pengguna menjadi lebih relevan.

Perusahaan B2B juga menggunakan metode lain untuk mengumpulkan data tentang perusahaan pengguna mereka, menggunakan pengikisan web dari situs berita dan lowongan pekerjaan, serta dari pembelian data untuk memahami lebih lanjut tentang mereka. Wawasan yang diperoleh dari data terstruktur atau tidak terstruktur ini dapat mencakup lonjakan perekrutan, investasi yang dilakukan, perubahan dalam dewan direksi, atau arah strategis baru. Jika digabungkan secara efektif, data yang dihasilkan menciptakan pandangan tentang "perilaku" perusahaan. Berdasarkan informasi ini, penjual menggunakan model AI untuk menilai kemungkinan pengguna di perusahaan tersebut akan membeli produk dan layanan penjual, dan jika ya, kapan. Wawasan ini kemudian memungkinkan manajer hubungan untuk berfokus pada segmen dan pembeli yang paling mungkin dari akun target.

Algoritma AI tidak hanya dapat menilai prospek terbaik perusahaan, tetapi juga dapat terus menyempurnakan skor tersebut seiring pembeli menghubungi divisi penjualan dan pemasaran. Algoritma terkadang bahkan digunakan untuk mencocokkan perwakilan penjualan tertentu dari penjual dengan peluang spesifik untuk mendapatkan hasil terbaik. Dengan menggunakan model AI, manajer akun mendapatkan wawasan baru tentang aktivitas pelanggan, yang memungkinkan mereka dengan cepat mengidentifikasi cara terbaik untuk menyesuaikan produk atau layanan mereka atau pesan tentang layanan tersebut guna meningkatkan kesuksesan pelanggan. Perusahaan jasa profesional, perusahaan penjualan produk industri, perusahaan energi, dan lainnya menggunakan pendekatan jenis ini.

### **3.5 PENETAPAN HARGA DINAMIS DAN OPTIMALISASI RANTAI PASOKAN**

Pemogokan di pelabuhan atau kekurangan tenaga kerja di lokasi pembeli adalah beberapa data pihak ketiga yang mungkin dimasukkan ke dalam sistem AI yang digunakan untuk memprediksi permintaan. Perusahaan menggunakan AI untuk memanfaatkan data kontekstual dan membangun model AI yang meningkatkan prediksi. Informasi seperti hari dalam seminggu atau minggu dalam setahun, perilaku penelusuran pelanggan, dan kampanye pemasaran terkini semuanya dapat memengaruhi permintaan, dan berpotensi, harga.

Cuaca, misalnya, memiliki efek dramatis pada cara orang membeli atau apa yang mereka beli. Cuaca buruk baik untuk e-commerce. Orang-orang tetap di dalam ruangan, menjelajah, dan membeli secara online daripada pergi ke toko. Menghitung informasi ini sebelumnya dan menggabungkannya dengan titik data lain memberi tahu peritel apakah mereka akan menjual lebih banyak atau lebih sedikit pada hari itu. Kondisi cuaca juga dapat memengaruhi rantai pasokan, yang dapat berdampak pada hal-hal seperti pergudangan dan menipisnya inventaris. Peningkatan *peramalan permintaan* dengan kecerdasan buatan untuk memperkuat rantai pasok ritel telah menunjukkan manfaat yang signifikan, memungkinkan perusahaan membuat peramalan akurat tentang produk apa yang mereka butuhkan, pada

musim apa, kapan, atau di mana. Hal ini meningkatkan kemampuan pengecer untuk menyimpan produk dalam stok, menghemat biaya penyimpanan inventaris, dan bahkan biaya pembusukan.

Penggunaan model AI untuk penetapan *harga dinamis* juga menarik pelanggan yang sensitif terhadap harga tanpa mengorbankan pendapatan dari pelanggan yang kurang sensitif terhadap harga. Memiliki titik data ini berarti bisnis dapat menetapkan harga yang berbeda ketika diperlukan. Munculnya penjualan kilat merupakan hasil langsung dari pengetahuan semacam ini, meskipun pada tingkat yang lebih agregat. Algoritma AI dapat menetapkan harga optimal secara mendekati waktu nyata (*real-time*) untuk meningkatkan pendapatan dan keuntungan bagi bisnis serta menghindari kebutuhan akan penjualan kilat. Hal ini lebih mudah bagi bisnis e-commerce, karena semua bagian transaksi telah terdigitalisasi; tetapi layanan juga muncul di perusahaan ritel fisik, di mana label harga dinamis di rak dalam bentuk layar mini memungkinkan proses serupa. Label harga ini tidak ditujukan untuk individu tetapi lebih berdasarkan waktu dan tempat, dan sekali lagi, bahkan cuaca.

Model AI juga membantu *pengembalian produk*, yang menjadi momok bagi banyak bisnis. Di Amerika Serikat saja, Statista memperkirakan biaya pengiriman pengembalian akan mencapai Rp. 550 Triliun per tahun pada tahun 2020. Tingkat pengembalian lebih buruk untuk belanja daring dibandingkan dengan belanja di toko, tetapi menurut berbagai survei, pengiriman gratis dan pengembalian gratis adalah beberapa alasan penting mengapa pembeli cenderung berbelanja daring. Untuk mengatasi dikotomi ini, banyak peritel menggunakan model pembelajaran mesin untuk menemukan akar penyebab pengembalian, seperti ukuran yang tidak pas, hasil akhir yang tidak diharapkan, atau karena efek "ruang ganti". Efek "ruang ganti" mengacu pada situasi di mana pelanggan membeli barang seolah-olah mereka sedang mencobanya di ruang ganti toko fisik, dengan niat untuk tidak melakukan pembelian sampai mereka melihat bagaimana pakaian tersebut terlihat pada mereka. Dalam e-commerce, ini berarti membeli banyak barang dan kemudian, setelah mencobanya, mengembalikan sebagian besar atau semuanya.

Perusahaan juga membangun model AI yang mengevaluasi keranjang belanja seseorang berdasarkan perilaku sebelumnya – misalnya, secara konsisten memesan ukuran yang salah – untuk menentukan seberapa besar risiko mereka mengembalikan produk. Mereka kemudian menerapkan hukuman atau insentif jika pengembalian kemungkinan terjadi. Ini termasuk tindakan pencegahan, seperti menaikkan biaya pengiriman, atau insentif, seperti menawarkan kupon sebagai imbalan atas pembelian yang tidak dapat dikembalikan.

Kombinasi ritel daring dan fisik tradisional, yang dikenal sebagai ritel omni-channel, telah menghasilkan perkembangan lain: penggunaan lokasi toko tradisional sebagai tempat untuk memenuhi atau mengembalikan pesanan. Pengecer sering menggunakan AI untuk *optimasi pemenuhan*, menentukan cara paling hemat biaya untuk menangani pesanan: pengiriman dari pabrik, pusat pemenuhan, atau toko lokal. Dengan menggabungkan model pengiriman terbaik dengan model inventaris global, AI memungkinkan para pengecer ini untuk mengganti aturan statis seperti "kirim dari gudang terdekat" menjadi, sebagai gantinya, mengoptimalkan profitabilitas secara dinamis.

Beberapa pengecer besar sedang bereksperimen dengan, atau menggunakan, robot dan drone di gudang mereka. Mereka sering bermitra dengan perusahaan robotika dan perusahaan rintisan untuk pekerjaan ini. Meskipun sebagian besar robotika menggunakan AI dan pembelajaran mesin untuk mengoperasikan robot, topik ini sendiri sangat luas dan berada di luar cakupan buku ini.

### 3.6 ASISTEN DIGITAL DAN KETERLIBATAN PELANGGAN

Interaksi layanan pelanggan telah berkembang seiring waktu. Awalnya, semua interaksi dilakukan melalui telepon dengan perwakilan layanan pelanggan (CSR) secara langsung. Kemudian, berbagai bisnis menerapkan sistem respons suara interaktif (IVR): perintah suara interaktif yang terkadang mengganggu yang mengarahkan Anda untuk menjawab berbagai pertanyaan sebelum menindaklanjuti permintaan Anda atau meneruskannya kepada seseorang yang dapat membantu Anda ("Untuk berbicara dengan operator, tekan nol"). Sistem ini bekerja melalui sinyal multifrekuensi nada ganda (DTMF), teknologi yang memungkinkan Anda merespons melalui keypad nada sentuh ponsel, dan mengurangi beban CSR. Kemudian, pesan ditambahkan: sebuah cara untuk menggantikan panggilan telepon dengan mengobrol dengan CSR. Hal ini memungkinkan perusahaan untuk menyusun percakapan dan menangani masalah secara lebih efisien daripada hanya menggunakan perintah suara telepon. Sifat asinkron dari saluran ini menghemat waktu perusahaan, karena kecepatan percakapan dapat diatur tanpa membuat pelanggan merasa sedang menunggu, meskipun responden dapat menangani beberapa pelanggan secara bersamaan.

Kini, pengiriman pesan atau obrolan semakin sering terjadi antara pelanggan dan asisten digital berbasis AI, yang terkadang disebut *asisten virtual* atau chatbot. AI memungkinkan *pemrosesan bahasa alami* (NLP) dan ucapan ke teks serta penerjemahan teks ke ucapan dalam asisten digital ini. Hal ini memungkinkan asisten untuk memahami pertanyaan pelanggan dan menemukan respons atau jawaban yang tepat. Juniper Research menemukan bahwa chatbot berbasis AI saja dapat menghemat Rp. 80 Triliun per tahun bagi bisnis pada tahun 2022, dengan sektor kesehatan dan perbankan yang paling diuntungkan.

Pemrosesan bahasa alami adalah kemampuan sistem untuk memilah makna dari bahasa. Kemampuan untuk memahami bahasa alami tertulis (atau lisan dan transkripsi) ini memungkinkan model AI ini untuk menentukan informasi spesifik dalam dokumen teks berukuran besar, serta memfilter dan mengelompokkan informasi ini, yang memungkinkan perusahaan untuk mengotomatiskan respons mereka terhadap permintaan pelanggan. Kondisi NLP saat ini "memahami" pertanyaan "bertanya" dan menjawab pelanggan ketika penalaran tidak diperlukan – di mana jawabannya ditulis dalam beberapa data teks tidak terstruktur yang perlu ditemukan dan divalidasi sebagai kemungkinan jawaban. Sebagian besar asisten digital berbasis AI yang berhadapan langsung dengan pelanggan saat ini terintegrasi dengan eskalasi agen, seperti menekan angka nol untuk berbicara dengan operator. Hal ini dikarenakan teknologi NLP berbasis AI belum matang hingga mencapai titik di mana semua interaksi pelanggan dikelola secara otonom.

Dengan memanfaatkan NLP, AI digunakan untuk memeriksa transkrip panggilan historis di pusat panggilan. Hasilnya adalah peningkatan kemampuan untuk memahami tidak hanya frekuensi dan waktu panggilan, tetapi juga sentimen penelepon serta jenis dan topik permintaan: misalnya, jika musim memengaruhi jumlah panggilan tentang subjek tertentu atau durasi permintaan tertentu. Pengetahuan mendalam ini memungkinkan algoritma AI untuk memperkirakan jumlah panggilan yang diharapkan per jam per topik secara lebih tepat. Peningkatan yang dihasilkan dalam pencocokan dan penjadwalan agen pusat panggilan berarti pusat kontak cenderung tidak kekurangan staf, yang berdampak negatif pada layanan pelanggan, atau kelebihan staf, yang memerlukan biaya yang tidak perlu.

Penggunaan NLP juga memungkinkan pusat kontak untuk mengarahkan panggilan dengan cara yang lebih efisien dan bahkan penuh kasih sayang dengan membebaskan karyawan untuk menangani situasi yang lebih sensitif. Manusia tidak selalu harus menangani permintaan; asisten digital mungkin sesuai jika penelepon tidak akan merasa terasing dalam prosesnya. Namun, jika seseorang menelepon untuk melakukan pembelian atau memperbarui akun yang terkait dengan, misalnya, duka cita, pusat panggilan yang disempurnakan dengan AI kemungkinan besar memiliki agen pusat panggilan terlatih yang tersedia untuk menangani situasi tersebut. Pendekatan ini memindahkan volume tinggi interaksi yang lebih sederhana dari agen manusia ke asisten digital, memberi agen lebih banyak waktu untuk mengelola kebutuhan pelanggan yang kompleks. Beberapa perusahaan menggunakan interaksi dan transkrip pusat panggilan ini untuk terus memberi makan grafik pengetahuan organisasi, meningkatkan percakapan masa depan yang dapat didukung oleh NLP. (Lihat Bab 12 untuk informasi lebih lanjut tentang pola ini.)

Dengan bantuan sistem AI, peritel saat ini mengoptimalkan saluran mana yang terbaik untuk jenis pesan apa (telepon, teks, email, web, atau aplikasi). Sistem AI memungkinkan peritel untuk mengukur respons setiap pelanggan terhadap saluran serta pesan itu sendiri, termasuk panjang pesan dan susunan katanya. Bahkan perubahan yang relatif kecil dalam cara peritel menjangkau pelanggan berarti pelanggan mungkin lebih mungkin melihat dan menanggapi pesan-pesan ini. Perubahan tersebut juga dapat menyebabkan peningkatan substansial dalam tingkat keberhasilan kampanye pemasaran peritel.

Asisten digital memiliki beragam kegunaan baik di arena ritel maupun di banyak industri lainnya, kegunaan yang mungkin relatif sederhana tetapi semakin berharga bagi bisnis ini maupun pelanggan yang mereka layani. Tersedia 24/7, asisten digital ini disiapkan untuk menangani berbagai jenis permintaan pelanggan. Asisten digital yang lebih canggih memandu pelanggan melalui proses tertentu, seperti pembelian peralatan rumah tangga atau mobil yang rumit. Dalam layanan keuangan, misalnya, asisten digital digunakan untuk memandu pelanggan melalui langkah-langkah yang diperlukan untuk mengajukan pinjaman hipotek, mengubah apa yang sebelumnya merupakan proses melelahkan dengan mengisi aplikasi yang panjang di situs web menjadi interaksi percakapan yang lebih singkat.

Interaksi ini membantu meningkatkan konversi pelanggan, karena pelanggan yang ragu atau memiliki pertanyaan yang belum terjawab lebih mungkin untuk keluar dari proses pembelian. Pembelian yang didukung asisten digital lebih mudah menjawab dan menjawab

banyak pertanyaan pelanggan. Pelanggan yang sudah familiar dengan, misalnya, Facebook Messenger, merasa pendekatan ini nyaman dan menghemat waktu. Pengalaman berbasis suara atau obrolan ini dimungkinkan oleh kemajuan dalam pembelajaran mesin dan pemrosesan bahasa alami, dan ketersediaannya berarti bahwa bisnis dapat melengkapinya interaksi karyawan manusia yang mahal dengan solusi otomatis.

Peritel juga memanfaatkan meningkatnya adopsi perangkat rumah pintar seperti Microsoft Cortana, Apple Siri, Amazon Echo, atau Google Home. Mereka mengintegrasikan perangkat cerdas ini ke dalam sistem AI mereka untuk mendukung percakapan pelanggan otomatis yang canggih, dengan "keahlian" khusus perusahaan yang ditambahkan ke kapabilitas perangkat dasar. Perangkat rumah pintar bekerja serupa dengan membuka aplikasi di ponsel pintar, tetapi alih-alih mengeklik ikon untuk membuka aplikasi, Anda mengatakan "Alexa, buka aplikasi MASUKKAN\_MEREK\_ANDA\_DI\_SINI saya." Setelah aplikasi aktif di Alexa (misalnya sebagai keahlian Alexa), pelanggan berinteraksi dengan aplikasi peritel untuk mengautentikasi, mengajukan pertanyaan, dan bertransaksi. Meskipun peritel harus mempertimbangkan produk apa yang bekerja dengan baik dalam lingkungan suara saja, banyak yang berfokus pada pembelian ulang.

Saat ini, penelitian sedang dilakukan pada skenario yang semakin kompleks, seperti memungkinkan AI untuk mendeteksi, langsung dari nada suara atau penggunaan bahasa pelanggan, ketika ia tidak senang dan kemungkinan akan meninggalkan bisnis, meskipun bisnis belum menerapkan kemampuan ini dalam skala besar. Bidang yang relatif baru ini, yang dikenal sebagai *komputasi afektif*, melibatkan penggunaan AI dalam bentuk analisis ucapan, dan pengenalan ekspresi wajah untuk memahami dan mengukur respons emosional terhadap berbagai rangsangan dengan menetapkan nilai pada teks positif, negatif, atau netral. Komputasi afektif menambah analisis sentimen, jenis penambangan teks yang menggunakan NLP untuk menentukan pendapat orang, memungkinkan sistem AI untuk menganalisis interaksi pelanggan, misalnya, untuk mendeteksi situasi di mana klien mengekspresikan ketidakbahagiaan.

Banyak perusahaan sekarang juga menggunakan teknologi AI seperti pembelajaran mesin dan NLP untuk mendukung karyawan agar menanggapi pertanyaan layanan pelanggan baik teks maupun suara dengan tepat. Hal ini sering dilakukan dengan menerapkan asisten digital ke CSR, alih-alih langsung ke pelanggan. Ketika seorang pelanggan menelepon, CSR manusia mengakses asisten digital yang mengikuti percakapan tersebut. Dengan cara ini, asisten digital ada untuk mendukung agen untuk membantu pelanggan, tetapi pelanggan tidak berinteraksi langsung dengan asisten digital. Asisten digital menyarankan kepada agen manusia tindakan terbaik berikutnya untuk pelanggan ini, dan agen menyampaikan informasi ini kepada pelanggan dengan cara yang paling tepat. Beberapa perusahaan juga telah mengembangkan algoritma rekomendasi produk dan layanan dalam asisten digital ini untuk membantu penjualan silang dan penjualan atas. Transaksi pengguna dan pola navigasi diproses untuk menghasilkan rekomendasi dan memungkinkan agen untuk menawarkan pelanggan produk atau penawaran terbaik yang tersedia. Rekomendasi tersebut kemungkinan

besar akan sama dengan rekomendasi yang akan diterima pelanggan jika ia menggunakan aplikasi seluler yang dipersonalisasi.

Teknologi AI yang tidak berhadapan langsung dengan pelanggan ini, yang memungkinkan peningkatan layanan pelanggan, cenderung lebih banyak digunakan karena penggunaannya secara internal berarti dapat diuji dan disempurnakan tanpa risiko mengasingkan pelanggan. Hal ini mencegah rekomendasi yang tidak pantas langsung sampai ke pelanggan tanpa penilaian manusia dan mempertahankan interaksi pelanggan yang empatik melalui agen manusia.

Pusat dukungan sering kali menerima email dari pelanggan dengan tajuk yang hanya bertuliskan "Bantuan" di baris subjeknya. NLP saat ini merupakan cara otomatis terbaik untuk mengubah email ini menjadi sesuatu yang dapat ditindaklanjuti dengan menggali lebih dalam teks, mengetahui siapa di perusahaan yang menangani pelanggan atau masalah tertentu, dan secara otomatis *mengarahkan tiket masalah* ke orang tersebut. Aplikasi NLP lainnya saat ini adalah untuk memahami umpan balik atau komentar pelanggan guna mendapatkan wawasan yang memungkinkan perusahaan membuat perubahan dan meningkatkan proses.

Salah satu penggunaan AI yang lebih menarik yang sedang diujicobakan oleh beberapa peritel adalah "pengalaman ritel generasi mendatang". Ini mencakup input suara dan gestur yang terintegrasi dengan layar sentuh atau augmented reality (AR) atau virtual reality (VR) di kios ritel dan ruang pameran eksperimental lainnya.

Dari pemasaran yang tepat sasaran hingga rekomendasi yang dipersonalisasi secara individual, solusi AI yang dijelaskan sebelumnya sudah digunakan dalam e-commerce omni-channel di industri ritel dan industri lainnya. Lebih banyak lagi aplikasi AI yang diantisipasi di masa mendatang.

#### **Bilah Sampung: Percakapan Alami**

Salah satu tujuan utama AI adalah menciptakan algoritma yang memungkinkan pelanggan melakukan percakapan alami dengan komputer melalui telepon, alih-alih mengandalkan interaksi IVR yang kaku atau agen manusia yang mahal, yang merupakan norma saat ini. Percakapan alami berarti pemanfaatan waktu yang lebih baik dan keuntungan yang lebih tinggi tanpa mengurangi kepuasan pelanggan. Dalam upaya mencapai cita-cita ini, pada 8 Mei 2018, Google mengumumkan Google Duplex, sebuah teknologi yang memungkinkan orang melakukan percakapan alami melalui telepon dengan komputer. Saat ini, teknologinya sangat terbatas, hanya ditujukan untuk tugas-tugas tertentu seperti membuat janji temu di salon. Bahkan membatasi Duplex pada domain tertutup semacam ini pun tetap membutuhkan pelatihan yang mendalam. Namun, ini merupakan satu langkah menuju kemampuan berinteraksi alami dengan komputer.

## BAB 4

### AI DALAM LAYANAN KEUANGAN

Meskipun bank, manajer aset, dan perusahaan jasa keuangan lainnya cenderung sepakat bahwa AI penting, mereka sangat bervariasi dalam hal mengapa, bagaimana, dan kapan mengadopsi teknologi baru ini. Menurut sebuah artikel di *McKinsey Quarterly*, lebih dari selusin bank Eropa sekarang menggunakan teknik pembelajaran mesin alih-alih pendekatan pemodelan statistik yang lama, dan beberapa sekarang melihat peningkatan hingga 10% dalam penjualan produk baru; penghematan 20% dalam belanja modal; dan penurunan churn pelanggan sebesar 20%. Pada tahun 2018, Forum Ekonomi Dunia melaporkan bahwa "semakin banyak lembaga keuangan yang menerapkan AI pada saran dan interaksi nasabah, yang meletakkan dasar bagi keuangan mandiri." Menurut survei McKinsey, salah satu bank Eropa dikutip di *Financial Times* memiliki 500–800 karyawan yang bekerja di bidang AI. Meskipun metrik ini tampak mengesankan, kemungkinan besar banyak contoh yang merupakan bukti konsep yang lebih kecil daripada aplikasi pembelajaran mesin di seluruh perusahaan.

Lembaga keuangan memiliki data terstruktur dalam jumlah besar, yang seringkali sudah berkualitas tinggi, sehingga cocok untuk penggunaan AI. Banyak bank menggunakan AI untuk menilai dan mengelola risiko, seperti dalam deteksi penipuan dan mengelola risiko kredit untuk menyetujui pinjaman; untuk layanan nasabah, seperti asisten digital dan saran keuangan yang disesuaikan dan prediktif; untuk menghasilkan alpha yang lebih tinggi dengan perdagangan algoritmik; dan untuk meningkatkan efisiensi operasional di dalam bank.

Seiring AI terus membuktikan kemanjurannya di perbankan, serta di semakin banyak industri lainnya, perusahaan investasi telah menggunakannya untuk membuat keputusan perdagangan, dan mereka bersedia mempertaruhkan sejumlah besar uang untuknya – baik uang klien maupun uang mereka sendiri. Beberapa kasus penggunaan AI lebih relevan bagi perusahaan yang mencari pertumbuhan jangka panjang, seperti perusahaan ekuitas swasta; beberapa kasus penggunaan mungkin relevan untuk imbal hasil jangka menengah, seperti perusahaan manajemen aset dan kekayaan. Kasus lain mungkin paling tepat bagi dana lindung nilai kuantitatif yang mencari perputaran harian saat memasuki dan keluar dari posisi ekuitas (membeli dan menjual pada hari yang sama atau dalam hitungan milidetik).

Dua kasus penggunaan AI yang signifikan dalam industri perbankan investasi adalah di bidang riset investasi dan perdagangan algoritmik. Penggunaan utama ketiga adalah dalam organisasi penjualan dan distribusi perusahaan manajemen aset sehingga mereka dapat memahami perilaku klien dan menargetkan layanan dengan lebih baik untuk memenuhi kebutuhan klien mereka, seperti menciptakan wawasan dan pengalaman personal yang berdampak bagi klien dan menyediakan informasi relevan kepada manajer hubungan manusia atau penasihat untuk membantu investor. Kasus-kasus penggunaan ini hanyalah puncak gunung es dalam hal penerapan AI di bank atau perusahaan manajemen aset, dan di masa

mendatang, lembaga jasa keuangan akan semakin memperluas cakupan kasus penggunaannya. Berikut ini adalah beberapa area spesifik di mana model AI mencapai hasil yang sukses.

#### **4.1 ANTI PENCUCIAN UANG**

AI menjadi instrumen penting dalam menangani salah satu masalah paling umum yang dihadapi industri perbankan: pencucian uang. Menurut Kantor Perserikatan Bangsa-Bangsa untuk Narkoba dan Kejahatan (UNODC), "Perkiraan jumlah uang yang dicuci secara global dalam satu tahun adalah 2%–5% dari PDB global, atau \$800 miliar–\$2 triliun dalam dolar AS saat ini."<sup>4</sup> Selain itu, terdapat berbagai biaya tambahan dari pencucian uang, termasuk kegiatan kriminal seperti penggelapan pajak, perdagangan narkoba, perdagangan manusia, dan pendanaan kegiatan teroris.

Pencucian uang adalah upaya untuk menyembunyikan dana yang diperoleh secara ilegal dengan melewatkannya melalui sistem keuangan, terutama untuk menyembunyikan atau menyamarkan tiga hal: dari mana dana tersebut berasal, siapa yang memiliki atau mengendalikannya, dan di mana dana tersebut akhirnya disimpan atau dibelanjakan. Pelaku pencucian uang menyetorkan dana yang tampak sah tetapi berasal dari kegiatan kriminal ke dalam rekening bank, pialang, atau lembaga keuangan lainnya. Selanjutnya, mereka melakukan serangkaian transaksi untuk menyembunyikan transaksi pertama tersebut. Misalnya, pelaku pencucian uang mungkin menarik dana mereka dari lembaga pertama dan menyetorkannya ke satu atau lebih bisnis beberapa kali. Setelah itu, mereka mungkin menarik uang tersebut dan membelanjakannya atau menginvestasikannya.

Semua bank memiliki tim anti-pencucian uang (AML) untuk memerangi kejahatan keuangan. Mereka memantau aktivitas nasabah yang mungkin merupakan indikasi pencucian uang. Tim-tim ini bertanggung jawab untuk menyediakan laporan aktivitas mencurigakan (SAR) kepada lembaga regulator (termasuk pemerintah atau lembaga internasional). Di Amerika Serikat, unit Jaringan Penegakan Kejahatan Keuangan (FinCEN) dari Departemen Keuangan AS mengatur industri perbankan. Laporan aktivitas mencurigakan diperlukan setiap kali transaksi yang mencurigakan tercatat.

Transaksi mencurigakan yang mungkin mengindikasikan potensi penipuan paling mudah dideteksi dengan memisahkan yang asli dari yang palsu: yaitu, membedakan transaksi biasa yang diduga dari transaksi kriminal. Untuk melakukannya, tim AML menggunakan apa yang dikenal sebagai deteksi anomali. Pertama, mereka menghubungkan titik-titik di antara beragam kumpulan data dari berbagai sumber, yang kemungkinan tersimpan dalam sejumlah sistem yang berbeda. Data ini mungkin mencakup informasi tentang SAR sebelumnya atau data AML yang ada. Kemudian, mereka mencari anomali dalam transaksi.

Hingga saat ini, bank telah mampu menandai berbagai ketidaksesuaian sederhana, misalnya, transaksi nasabah dengan jumlah setoran lebih dari dua kali lipat dari jumlah normal, yang dapat konsisten dengan pencucian uang. Ketika pola yang cocok diamati, baik yang menyerupai pencucian uang sebelumnya atau pengulangan berkelanjutan dari transaksi yang tidak biasa yang sama, peringatan akan dikeluarkan dan kasus tersebut dirujuk ke tim AML

bank untuk ditinjau secara manual. Setelah beberapa putaran evaluasi, jika penyelidik menyimpulkan perilaku tersebut merupakan indikasi pencucian uang, bank akan mengajukan SAR kepada badan pengaturnya.

Kesulitannya adalah meskipun bank telah menggunakan komputer untuk memantau transaksi selama beberapa dekade, komputer-komputer ini menggunakan sistem berbasis aturan. Setiap sistem berbasis aturan pada dasarnya kaku dan oleh karena itu belum tentu mampu mengurai interaksi kompleks di antara berbagai cara yang digunakan untuk mencuci uang. Itulah sebabnya tingkat positif palsu yang dihasilkan oleh sistem berbasis aturan tersebut mencapai 90% atau lebih. Hal ini membuat kemampuan mereka untuk berhasil melacak semua penipuan nyata menjadi sangat mahal, karena transaksi yang ditandai membutuhkan banyak jam kerja manusia yang mahal untuk diselidiki.

AI, di sisi lain, menggunakan data yang tersedia untuk membuat kumpulan model AI. Mesin AI kemudian memilih model yang paling baik memprediksi kasus mana dalam data yang menyerupai kasus yang pada akhirnya menghasilkan pengajuan SAR sebelumnya, sehingga mendefinisikan pola yang lebih akurat sebelum meneruskan informasi ke tangan manusia. Ini mengurangi positif palsu secara signifikan, sambil mengambil lebih banyak positif benar. Umumnya, sistem AI jenis ini bekerja dengan awalnya menerapkan pembelajaran tanpa pengawasan pada data, yang memungkinkan bank untuk menemukan pengelompokan berdasarkan pola serta potensi anomali atau outlier. Setelah itu, jika ada kasus pencucian uang yang diketahui, titik-titik data tersebut disorot di setiap kluster, yang mengarah pada peninjauan kelompok data di dekat kasus penipuan tersebut. Ketika contoh-contoh tambahan ditemukan, baik melalui proses ini maupun dengan menemukannya melalui investigasi lain, data yang cukup dikumpulkan untuk memanfaatkan pembelajaran terawasi, yang memprediksi apakah transaksi baru kemungkinan merupakan penipuan.

Banyak bank sudah menggunakan AI untuk menandai aktivitas yang tidak biasa, memungkinkan mereka untuk menentukan kapan seorang analis perlu menyelidiki lebih lanjut suatu transaksi. Penggunaan AI ini memberi tim AML cara yang sangat baik untuk menilai kepatuhan terhadap peraturan AML, juga mengurangi tingkat positif palsu untuk kasus-kasus yang dipilih untuk peninjauan manual. Dengan demikian, penggunaan AI menghasilkan tim kepatuhan yang lebih efisien. Keuntungan lain dari sistem AI adalah kemampuannya untuk berkembang seiring dengan perubahan perilaku pencuci uang dari waktu ke waktu, yang berarti bahwa tim AML tidak perlu merevisi dan menerapkan kembali aturan secara eksplisit. Namun, penting untuk menyadari bahwa jenis-jenis penipuan atau anomali baru terus bermunculan, sehingga pembelajaran tanpa pengawasan akan terus digunakan.

## **4.2 PINJAMAN DAN RISIKO KREDIT**

Dengan meningkatnya ekspektasi dari nasabah komersial dan ritel serta potensi untuk meningkatkan pendapatan sekaligus menurunkan risiko, bank mulai memanfaatkan AI untuk keputusan persetujuan pinjaman mereka. Persetujuan kredit selama ini didasarkan pada logika berbasis aturan dengan menggunakan faktor-faktor statis seperti sektor bisnis dan omzet industri (untuk pinjaman bisnis), peringkat kredit pribadi dan bisnis, status perkawinan, dan

faktor-faktor lainnya. Pendekatan ini telah menjadi fondasi manajemen siklus hidup kredit untuk menyetujui kartu kredit, pinjaman pribadi dan bisnis, hipotek, dan jalur kredit lainnya selama beberapa waktu.

Faktor-faktor ini cenderung berkembang perlahan meskipun kondisi pasar berubah, meskipun kesehatan kredit konsumen atau bisnis dapat memburuk dengan cepat. Tujuan utama penggunaan AI adalah untuk meningkatkan tingkat persetujuan otomatis demi biaya transaksi yang lebih rendah, meningkatkan waktu pengambilan keputusan, dan meningkatkan kualitas keputusan dengan risiko dan bias yang berkurang. Untuk mencapai hal ini, AI memanfaatkan kumpulan data besar untuk proses pengambilan keputusan, seperti perilaku transaksional pelanggan melalui pembayaran gaji, sewa, dan utilitas di samping informasi profil statis dan faktor kredit tradisional seperti rasio utang terhadap pendapatan, untuk meningkatkan kualitas keputusan pinjaman. Data ini kemudian digunakan untuk membangun model AI untuk memprediksi kelayakan kredit.

Pendekatan yang diambil beberapa bank pertama-tama mengkategorikan transaksi ke dalam pengelompokan yang relevan. Kategorisasi ini memberi bank cara untuk memahami jenis transaksi apa yang dilakukan setiap pelanggan. Selanjutnya, mereka melihat analisis deret waktu, misalnya, mencari prediktor kesulitan keuangan atau risiko perilaku lainnya seperti melakukan pembayaran pinjaman bulanan dengan kartu kredit mereka. Setelah itu, mereka mengelompokkan pelanggan berdasarkan risiko gagal bayar kredit – sekarang berdasarkan masukan ini daripada hanya pada profil atau data statis seperti peringkat kredit eksternal, sehingga meningkatkan kualitas keputusan kredit.

#### **4.3 LAYANAN DAN SARAN PREDIKTIF**

Kedalaman keterlibatan nasabah selalu menjadi komponen utama dalam meningkatkan nilai jangka panjang bagi nasabah perbankan, dan sistem AI digunakan untuk meningkatkan keterlibatan dan kepuasan nasabah. Sebelumnya, bank akan menghubungi nasabah utama mereka untuk menyampaikan saran atau penawaran, dan tim penjualan serta tim sukses nasabah mencatat wawasan mereka untuk diskusi di masa mendatang. Hal ini hanya mungkin dilakukan dengan persentase kecil nasabah yang berharga. Kini, sistem AI membantu menciptakan keterlibatan yang disesuaikan untuk semua nasabah, berdasarkan pelacakan komunikasi dan interaksi setiap nasabah dengan bank serta wawasan yang diekstraksi dari informasi transaksional dan data tentang interaksi mereka. Kombinasi kemampuan AI untuk menangani data dalam jumlah besar dan kapasitas untuk mengeksekusi model canggih memungkinkan bank menganalisis pola pengeluaran dan transaksi nasabah tertentu relatif terhadap situasinya serta tren dan insiden keuangan global dan lokal yang relevan. Analisis canggih ini, yang digerakkan oleh AI, membantu bank menentukan strategi jangka pendek atau jangka panjang terbaik untuk setiap nasabah.

Sistem AI menggunakan data perilaku untuk menentukan kondisi nasabah tersebut. Data ini berasal dari berbagai sumber, seperti transaksi bank dan kartu kredit, riwayat penelusuran nasabah di situs web atau aplikasi bank, kontak dengan pusat panggilan atau cabang, atau interaksi lainnya dengan bank. Informasi ini memungkinkan lembaga keuangan

memperoleh wawasan substantif secara real-time tentang kebutuhan, minat, tujuan, dan kekhawatiran nasabah. Wawasan ini memungkinkan mereka memberikan pengalaman yang disesuaikan dan menyeluruh kepada nasabah dan klien, meningkatkan segalanya mulai dari memberikan nasihat keuangan hingga menentukan metode komunikasi terbaik untuk setiap nasabah, baik melalui telepon, email, SMS, atau pesan lain seperti WhatsApp atau Facebook.<sup>5</sup>

Bank-bank mengambil isyarat dari mesin rekomendasi e-commerce di industri ritel untuk menawarkan nasihat keuangan prediktif kepada nasabah mereka yang sesuai dengan kebutuhan keuangan dan kenyamanan emosional mereka. Analisis ini memungkinkan bank untuk memberikan rekomendasi yang dirancang khusus untuk setiap klien, menciptakan hubungan yang lebih personal dan produktif. Layanan berbasis aplikasi AI seperti ini siap menjadi semakin penting bagi profitabilitas bank. Transformasi ini berkembang seiring dengan transformasi personalisasi di bidang ritel, di mana ada perbedaan besar antara merekomendasikan barang paling populer dan membuat rekomendasi yang dipersonalisasi.

Salah satu contohnya adalah penggunaan AI untuk memprediksi transaksi nasabah di masa mendatang, berdasarkan beragam data nasabah, termasuk informasi pengeluaran historis dari tagihan kartu kredit dan informasi transaksi bank. Transaksi mendatang yang diprediksi digunakan untuk memperkirakan pengeluaran dan saldo rekening, dan berdasarkan informasi ini, bank membuat rekomendasi tentang tindakan yang dapat diambil nasabah untuk menghindari cerukan. Menghindari cerukan dapat menjadi manfaat yang signifikan bagi nasabah – pada tahun 2018, nasabah bank membayar biaya cerukan sebesar Rp. 350 miliar. Sebaliknya, jika nasabah diprediksi memiliki lebih banyak uang dari biasanya, sistem dapat merekomendasikan transfer uang ke rekening tabungan atau investasi atau melakukan pembayaran tambahan atas pinjaman, jika ini merupakan pilihan yang lebih baik.

Contoh lain dari aplikasi AI adalah untuk memprediksi perilaku nasabah di masa mendatang, seperti pembelian peralatan rumah tangga atau mobil. Prediksi ini dapat didasarkan pada transaksi terkini seperti pembayaran perbaikan mobil dengan kartu debit bank atau mengklik iklan pinjaman mobil digital bank di situs web mobil. Dalam kasus ini, kredit disetujui terlebih dahulu, dengan asumsi pelanggan memenuhi syarat, dan pelanggan menerima pesan untuk mengantisipasi pembelian. Contoh lain termasuk memprediksi pajak penghasilan berdasarkan pendapatan dan pengeluaran tahunan, atau merekomendasikan pembiayaan kembali hipotek berdasarkan perubahan suku bunga. Karena data transaksi bank merupakan kumpulan data yang sangat kaya, berbagai jenis prediksi dimungkinkan, dan berbagai bank mengambil pendekatan yang berbeda.

Jenis AI yang sama yang digunakan oleh penasihat keuangan juga diterapkan pada berbagai kasus penggunaan dalam manajemen kekayaan. Pada tahun 2017, divisi manajemen kekayaan Morgan Stanley menghasilkan hampir setengah dari pendapatan tahunan perusahaan. Menambah grup manajemen kekayaan dengan AI dapat meningkatkan profitabilitas perusahaan secara dramatis. Karena manajer hubungan memediasi sebagian besar manajemen kekayaan, bank mengambil pendekatan hibrida, menawarkan kepada nasabah berbagai alat yang berhadapan langsung dengan klien yang didukung AI, seperti onboarding mandiri, portal, dan dasbor interaktif. Manajer hubungan menggunakan alat-alat

yang didukung AI ini untuk mendapatkan rekomendasi tindakan terbaik berikutnya yang harus diambil dengan klien-klien terbaik mereka. Mengintegrasikan AI ke dalam proses pengelolaan kekayaan membuka tingkat baru fokus pada klien, efisiensi organisasi, dan penghematan biaya.

Seiring perusahaan jasa keuangan meluncurkan algoritma pengelolaan kekayaan kognitif, fungsi front-office berubah dalam dua hal utama. Manajer hubungan dan tim pengelolaan kekayaan kini menggunakan waktu mereka secara lebih efisien dan efektif, dan klien memiliki pilihan untuk mendapatkan dukungan manusia atau virtual. Dengan menggunakan otomatisasi berbasis AI, alih-alih interaksi antarmanusia yang mahal, bank juga meningkatkan keuntungan dengan menjadikannya hemat biaya untuk melayani populasi klien yang lebih besar dari kalangan kurang mampu, sehingga membawa serangkaian nasabah baru ke dalam organisasi.

#### **Masalah Kotak Hitam dalam Kasus Penggunaan Keuangan**

Perusahaan keuangan yang mencoba memahami secara tepat bagaimana algoritma AI membuat keputusan harus menghadapi tantangan yang melekat pada teknologi tersebut. Tantangan utama di antaranya adalah fakta bahwa keputusan AI terjadi di dalam kotak hitam. Dalam situasi ini, Anda tahu data apa yang Anda masukkan ke dalam sistem dan prediksi apa yang dihasilkannya. Namun, mungkin sulit atau hampir mustahil untuk mengetahui bagaimana sistem membuat prediksi tersebut, tergantung pada kompleksitas modelnya. Kurangnya transparansi ini dapat menjadi masalah serius ketika bisnis harus menjelaskan mengapa algoritma mereka bermasalah dan merugikan pelanggan mereka. Hal ini perlu diingat ketika mencari kasus penggunaan AI di industri apa pun. Namun, terlepas dari keyakinan para skeptis bahwa hanya AI tingkat manusia yang akan mampu menangani sejumlah besar informasi yang memengaruhi pasar keuangan, para ilmuwan AI terus mengubah cara lembaga keuangan menjalankan bisnis dengan mengatasi tantangan ini menggunakan AI yang dapat diinterpretasikan.

#### **4.4 PERDAGANGAN ALGORITMIK DAN OTONOM**

Perdagangan algoritmik telah ada dalam berbagai bentuk selama beberapa dekade, terkadang dengan nama perdagangan otomatis dan perdagangan frekuensi tinggi. Pasar keuangan Amerika mulai menggunakannya pada tahun 1970-an. Perdagangan algoritmik tidak harus menggunakan AI; misalnya, seorang pedagang dapat memutuskan untuk menjual ketika rata-rata tiga tick terakhir (harga saham) lebih tinggi daripada angka tertentu, seperti rata-rata pergerakan historis. Ini adalah contoh algoritma atau aturan perdagangan yang memanfaatkan keahlian domain (meskipun contoh spesifik ini terlalu sederhana). Perdagangan algoritmik dengan cepat mendapatkan pangsa pasar karena memungkinkan organisasi perdagangan menghasilkan keuntungan yang sangat besar. Seiring dengan semakin matangnya algoritma AI, penerapannya pada perdagangan algoritmik menjadi hal yang tak terelakkan.

Perdagangan frekuensi tinggi, suatu bentuk perdagangan algoritmik, memiliki waktu eksekusi beberapa detik ketika diperkenalkan pada tahun 1983. Pada tahun 2010, waktu eksekusi ini telah berkurang menjadi milidetik atau kurang. Pada tahun 2016, firma riset pasar

modal Tabb Group memperkirakan bahwa perdagangan frekuensi tinggi menyumbang hampir 50% dari rata-rata volume perdagangan harian. Namun, ketika laba menurun seiring meningkatnya persaingan, para pedagang mulai mencari sesuatu yang dapat memberi mereka keunggulan kompetitif, dan mereka secara aktif mulai menggunakan AI. Insinyur komputer mengembangkan algoritma pembelajaran mesin (ML) yang mencakup informasi seperti batas eksposur, kelas aset, dan biaya perdagangan. Ketika algoritma ini diberikan data yang cukup, mereka dapat membandingkannya dengan pola historis, menghasilkan wawasan yang tidak dapat dilihat oleh analis manusia.

AI mampu mempercepat pencarian strategi perdagangan yang sukses, menyederhanakan proses yang sebelumnya membosankan dan memakan waktu. AI juga meningkatkan jumlah pasar yang dapat diamati dan ditindaklanjuti oleh pedagang. Menemukan asosiasi secara manual dari data kini menjadi masa lalu; AI memungkinkan pedagang untuk mengidentifikasi dan beradaptasi dengan tren yang berkembang. Hal ini membuat perdagangan algoritmik menjadi lebih menguntungkan. Salah satu kasus penggunaan yang dilaporkan adalah program AI LOXM milik JP Morgan, yang mengeksekusi perdagangan di seluruh bisnis ekuitas globalnya. Tidak seperti perdagangan berbasis aturan, LOXM belajar dari perdagangan sebelumnya untuk menetapkan parameternya, alih-alih hanya mengandalkan masukan pedagang. Kemudian, program ini mengeksekusi pesanan dengan kecepatan tercepat dan harga terbaik. Kecerdasannya yang terus berkembang memungkinkan LOXM untuk mengatasi masalah seperti cara terbaik melepas saham ekuitas besar tanpa mengubah harga pasar. Di masa mendatang, LOXM dapat mengenal setiap nasabah lebih dekat, sehingga memungkinkan LOXM untuk mempertimbangkan perilaku dan tujuan nasabah saat memutuskan cara terbaik untuk berdagang portofolio mereka.

Menggunakan AI dalam perdagangan algoritmik berarti AI mempelajari aturan yang seharusnya, menentukan struktur data, dan membuat prediksi berdasarkan hal tersebut. Pembelajaran mesin dan pembelajaran mendalam semakin banyak digunakan untuk membuat keputusan transaksi yang sangat kompleks dan berkecepatan tinggi ini. Pendekatan utama untuk hal ini adalah memprediksi *deret waktu*. Deret waktu adalah urutan titik data, seperti harga saham, dalam urutan kronologis dan berjarak waktu yang sama. Dalam perdagangan algoritmik, AI digunakan untuk memahami struktur data ini guna memperkirakan deret waktu dalam waktu dekat.

Kesulitan dari tugas ini adalah data yang mungkin tampak acak – dan mustahil untuk memperkirakan variabel acak secara akurat. Namun, ilmuwan komputer berhasil menggunakan berbagai teknik untuk mengatasi tantangan ini. Misalnya, mereka menggunakan rekayasa fitur pada data sumber dengan mengagregasi berdasarkan jumlah peristiwa, alih-alih murni kronologis, seperti setiap kali volume saham tertentu diperdagangkan, alih-alih berdasarkan interval waktu. Selain jenis rekayasa fitur ini, pembelajaran mendalam semakin banyak digunakan dalam komunitas kuantitatif – yaitu, bisnis yang menggunakan strategi sistematis atau algoritmik – yang mengarah pada apa yang disebut sebagian orang sebagai fase "Kuantitatif 2.0" Wall Street.

Perdagangan algoritmik berbasis AI menjadi fokus dana lindung nilai kuantitatif, yang berada dalam domain perdagangan frekuensi tinggi. Bank investasi dan manajer aset telah lebih fokus pada penggunaan AI dalam domain perdagangan frekuensi menengah hingga rendah untuk memanfaatkan informasi yang tersedia untuk umum dengan lebih cepat dan dalam skala yang lebih besar. Artinya, bank dan manajer aset menggunakan AI untuk mengekstrak informasi yang berguna dari data yang tersedia untuk umum sebelum pihak lain bertindak berdasarkan sinyal ini, dan melakukan hal ini di seluruh ekuitas, industri, sektor, dan wilayah (lihat bagian selanjutnya). Penggunaan AI untuk membuat keputusan investasi melibatkan analisis riset untuk transaksi bulanan atau triwulanan. Tidak seperti perdagangan algoritmik, ini adalah transaksi otonom berfrekuensi rendah.

#### 4.5 RISET INVESTASI DAN WAWASAN PASAR

Manajemen aset aktif adalah tentang arbitrase informasi dan seberapa cepat dan murah Anda dapat menyampaikan informasi ini kepada para pengambil keputusan atau manajer portofolio sehingga mereka dapat menindaklanjutinya. Manajemen informasi selalu mendorong kesuksesan dalam investasi, mulai dari pengumpulan data hingga analisis dan keputusan. Hal terpenting adalah mendapatkan wawasan yang paling relevan dari informasi dengan kecepatan yang tepat dan dengan biaya terendah. Analisis riset investasi menghabiskan banyak waktu untuk pengumpulan dan pengorganisasian data, tetapi produktivitas mereka menjadi semakin penting seiring dengan semakin kompetitifnya industri keuangan. Saat ini, sebagian besar pekerjaan manajemen data yang memakan waktu tersebut dapat dengan mudah diotomatisasi dan dianalisis menggunakan AI.

Data keuangan yang tersedia bagi analis dulunya terdiri dari dua jenis. Jenis pertama berupa informasi yang diterbitkan sendiri oleh perusahaan tertentu dalam bentuk laporan tahunan, pengajuan SEC, siaran pers, dan sejenisnya. Jenis kedua adalah informasi yang berasal dari pialang atau bursa saham, seperti harga saham, riwayat harga, dan rasio harga terhadap pendapatan. Data ini bukan hanya yang paling mudah tersedia; Informasi lain yang berpotensi bermanfaat berupa data tidak terstruktur sehingga sulit digunakan selain secara manual. *Data tidak terstruktur* biasanya padat teks, seperti data dari email atau tweet, meskipun mungkin juga berisi elemen seperti tanggal, angka, dan fakta, serta gambar dan video. Di sisi lain, *data terstruktur* adalah data yang memiliki bidang-bidang yang terdefinisi dengan baik, seperti nama, alamat, dan nomor telepon.

Sebelum tersedianya algoritma AI, data tidak terstruktur terlalu memakan waktu untuk dimanfaatkan secara intensif. Namun, hal itu berubah. Algoritma AI, dalam bentuk pemrosesan bahasa alami, menangani sejumlah besar data teks tidak terstruktur yang dikumpulkan dari berbagai sumber. Ini termasuk data yang terkubur dalam pengajuan, berita, penelitian, broker, dan konten internal serta data dari publikasi yang rumit dengan berbagai elemen data, yang semuanya menjadi penting untuk memahami bagaimana perusahaan beroperasi. Ini sering disebut *data alternatif* – kumpulan data nonfinansial tentang suatu perusahaan yang telah diterbitkan oleh sumber di luar perusahaan.

Kumpulan data ini dapat berasal dari berbagai sumber, termasuk umpan berita, unggahan lowongan kerja, diskusi media sosial, citra satelit, transaksi kartu kredit, dan perangkat seluler. Semua ini merupakan data unik, terperinci, dan berstempel waktu yang memungkinkan pandangan luas tentang tren aktual yang terperinci. Data ini sering kali melengkapi data keuangan yang tersedia, membantu analis membuat keputusan investasi yang lebih baik dengan mendapatkan basis masukan yang lebih luas dalam keputusan mereka.

Model AI juga digunakan untuk menambang data dari organisasi seperti bank sentral dan Biro Statistik Tenaga Kerja Nasional, menyusunnya menjadi model yang memprediksi kinerja pendapatan tetap, berdasarkan perubahan pekerjaan serupa di masa lalu. Inferensi dari model ini dibuat dalam hitungan detik, dan wawasan yang dihasilkan diterapkan dalam hitungan menit setelah data dikumpulkan. Sistem AI memungkinkan perusahaan untuk membaca mesin pengajuan SEC, panggilan pendapatan, transkrip dari hari investor, dan berbagai sumber lainnya. Mereka juga memahami pertanyaan bahasa alami lisan atau tertulis dan memberikan jawaban, meskipun hanya dalam domain tempat mereka dilatih, sehingga meningkatkan produktivitas analis riset dan kualitas rekomendasi mereka.

Setelah semua data terstruktur dan tidak terstruktur ini tersedia, algoritma pemrosesan bahasa alami digunakan untuk "memahami" konten secara semantik serta untuk memodelkan ontologi yang berkembang berdasarkan pemahaman ini. Ketika algoritma AI memetakan dokumen atau bahkan kalimat dalam konten dengan cara ini (misalnya, sebagai grafik pengetahuan), lebih mudah untuk menentukan informasi yang relevan. Bagi analis yang mungkin menghabiskan banyak waktu untuk mencari wawasan, ini memberikan manfaat yang substansial. Satu perusahaan melakukan studi untuk melihat berapa banyak waktu yang dihabiskan pengguna bisnis mereka untuk jenis aktivitas apa dan menemukan bahwa lebih dari 36% waktu mereka dihabiskan untuk mencari informasi. Dengan memanfaatkan algoritma AI dan pemrosesan bahasa alami, data diatur sehingga analis dapat mencari informasi lebih cepat dan tepat, memungkinkan mereka untuk mengidentifikasi pola dan memperkirakan kinerja dengan lebih akurat dan cepat. Dengan melapisi semua data ini, analis investasi tidak hanya menghemat waktu; mereka juga memperoleh wawasan pasar yang lebih signifikan. Sistem AI menelusuri jutaan dokumen relevan dan menemukan tidak hanya dokumen yang paling baik memberikan informasi tersebut, tetapi juga jawaban spesifik atas pertanyaan yang diajukan analis.

Sebelum pengembangan algoritma AI, analis riset di divisi manajemen aset suatu perusahaan akan membaca dokumen untuk menjawab serangkaian pertanyaan yang memungkinkan mereka memutuskan apakah akan merekomendasikan pembelian, penahanan, atau penjualan ekuitas. Mereka mengamati pertanyaan-pertanyaan seperti "Apakah pendapatan perusahaan bertumbuh?" (dari data terstruktur), atau "Apakah perusahaan berada dalam posisi yang cukup kuat untuk mengalahkan pesaingnya di masa depan?" atau "Apakah tim manajemen memiliki keyakinan terhadap perusahaan?" (dari data tidak terstruktur). Dengan perkembangan AI, ilmuwan AI melatih algoritma pemrosesan bahasa alami (NLP) untuk mencari hal-hal yang perlu diperhatikan – misalnya, kata-kata lindung nilai (misalnya, "hampir," "kemungkinan," dll.) yang digunakan oleh manajemen

selama panggilan pendapatan atau diskusi analis dapat menyiratkan bahwa mereka memiliki keyakinan yang rendah terhadap apa yang mereka katakan. Model NLP ini digunakan untuk menemukan jawaban atas pertanyaan-pertanyaan analis. Kemudian, analis riset meninjau laporan yang menyediakan pertanyaan-pertanyaan tersebut, dan jawaban-jawaban berdasarkan pemahaman mesin, dengan tautan ke sumber-sumber dari mana jawaban-jawaban tersebut dikumpulkan. Hal ini memungkinkan analis untuk mendalami lapisan analisis berikutnya jika diperlukan, dan untuk mencakup lebih banyak ekuitas per hari daripada sebelumnya.

Saat mengevaluasi industri yang akan difokuskan, analis mencari jawaban atas pertanyaan seperti "Apakah permintaan di sektor ini diperkirakan akan tumbuh dalam waktu dekat?" atau "Apakah disrupsi industri sedang dibahas?" Keduanya dapat diperoleh menggunakan teks tak terstruktur dari berita atau analisis. Misalnya, NLP digunakan untuk membaca beberapa sumber berita dan mengklasifikasikan apakah laporan tersebut tentang penawaran atau permintaan. Jika tentang permintaan, NLP kemudian mengklasifikasikan permintaan sebagai meningkat atau menurun per item berita. Keluaran dari ini adalah deret waktu sentimen penawaran dan permintaan, yang kemudian digunakan dengan indeks pertumbuhan industri historis sebagai fitur input untuk memprediksi pertumbuhan di masa mendatang.

Secara tradisional, mengetahui lalu lintas pejalan kaki untuk suatu bisnis di area tertentu dilakukan dengan dua cara: dengan penghitung tally, menggunakan clicker untuk menghitung secara manual setiap orang yang lewat, dan dengan memasang kamera keamanan di luar sehingga karyawan dapat menonton rekaman dengan cepat dan memperkirakan jumlah orang yang lewat per jam. Keduanya sangat memakan waktu, dan datanya biasanya tidak tersedia bagi perusahaan investasi. Namun, ada solusi yang dapat diskalakan: menggunakan data tayangan iklan dari ponsel yang terhubung dengan lokasi, yang mengungkapkan lebih banyak tentang penggunanya (usia dan jenis kelamin, misalnya) daripada sekadar penghitungan. Analis keuangan kini memiliki algoritma AI yang menganalisis citra satelit tempat parkir atau geolokasi ponsel untuk memantau berapa banyak orang secara agregat yang mengunjungi toko-toko perusahaan tertentu dari waktu ke waktu – dengan tren lalu lintas pejalan kaki atau penggunaan tempat parkir. Dengan menggunakan informasi ini dari waktu ke waktu, ilmuwan AI memprediksi pendapatan di masa mendatang, dan prediksi ini sebelum panggilan pendapatan perusahaan memberikan keuntungan bagi perusahaan manajemen aset.

Dengan menggabungkan semua kumpulan data ini dan melapisinya pada model keuangan yang ada, analis keuangan meningkatkan prediksi pendapatan mereka untuk toko ritel dan perusahaan lain, serta memperoleh wawasan yang lebih mendalam tentang penawaran dan permintaan serta faktor-faktor ekonomi lainnya. Analis riset ekuitas, dengan bantuan model AI, kini secara otomatis melacak saham dan mengembangkan perspektif jangka pendek dan jangka panjang yang sensitif terhadap waktu tentang apakah saham yang diikuti akan naik atau turun. Selain itu, seiring dengan terus meningkatnya data mentah yang

memengaruhi harga saham, model AI membantu mengubah data yang sulit dikelola ini menjadi wawasan yang dapat digunakan oleh manajer portofolio.

Semua data ini tidak gratis. Seperti kebanyakan komoditas berharga, kini terdapat pasar yang berkembang pesat untuk data alternatif (lihat Gambar 7.1 di Bab 7). Beberapa hedge fund menghabiskan jutaan dolar untuk mendapatkannya dari vendor yang mencari sumber, membersihkan, dan menjualnya kepada komunitas keuangan. Banyak perusahaan juga menginvestasikan uang untuk membangun infrastruktur mereka sendiri dan menciptakan sumber data alternatif mereka sendiri. Pemodal ventura dan perusahaan ekuitas swasta juga menggunakan AI untuk membuat keputusan investasi. Menurut Financial Times, seorang mitra di EQT Ventures Stockholm, Andreas Thorstensson, membuat sekitar 30% keputusan investasinya berkat data yang dianalisis oleh platform AI mereka, Motherbrain. (EQT Ventures adalah divisi modal ventura dari perusahaan Swedia, EQT Partners.) Motherbrain memantau sekitar dua juta perusahaan setiap hari, dan akibatnya, Thorstensson tidak lagi berinvestasi di perusahaan rintisan tertentu. "Data tidak berbohong," katanya.

#### **Ketika Wajah Berbicara Lebih Keras daripada Kata-kata**

Setelah pertemuan kebijakan internal rutin yang diadakan oleh Bank Sentral Eropa (ECP), presiden ECP biasanya mengadakan konferensi pers. Masalahnya, pada konferensi tersebut, ia hanya mengungkapkan sedikit tentang kebijakan moneter di masa mendatang. Namun, selama masa jabatan presiden ECP, Mario Draghi, para peneliti Jepang berhasil mengidentifikasi korelasi antara pola ekspresi wajahnya dan perubahan kebijakan selanjutnya. Untuk melakukan ini, mereka menggunakan API Emosi Microsoft, yang menggunakan algoritma pengenalan visual untuk menguraikan ekspresi wajah bahagia, sedih, terkejut, marah, takut, jijik, dan netral. Dengan menganalisis ekspresi Draghi, para peneliti berharap dapat menentukan kemungkinan perubahan kebijakan sedikit lebih awal, yang mungkin akan memberi mereka keunggulan investasi. Keberhasilan analisis ini belum dipastikan, tetapi ini merupakan contoh yang sangat baik tentang seberapa besar pengaruh AI terhadap keputusan keuangan dan keputusan lainnya.

#### **4.6 OPERASI BISNIS OTOMATIS**

Robotic process automation (RPA) adalah otomatisasi proses bervolume tinggi yang sebelumnya menjadi tanggung jawab manusia. Ini adalah teknologi yang terus berkembang yang menggabungkan otomatisasi penekanan tombol dengan aturan bisnis dan AI untuk mengotomatiskan proses berulang yang tidak memerlukan pengambilan keputusan yang signifikan. Perusahaan saat ini menggabungkan RPA dengan teknologi AI untuk membantu menerapkan otomatisasi proses back-end, proses yang sebelumnya sebagian besar manual sehingga lebih lambat, kurang akurat, dan kurang efisien. "RPA cerdas" ini membuat proses-proses ini lebih cepat, lebih andal, dan lebih efisien, mengurangi kesalahan manusia dan membebaskan manusia untuk tugas-tugas lain yang lebih kompleks. RPA juga memudahkan penskalaan sumber daya untuk memenuhi permintaan yang terus berubah dengan mengaktifkan bot tambahan. Intelligent process automation (IPA) secara substansial menambah tenaga kerja dengan tim robot perangkat lunak.

Dalam layanan keuangan, terdapat banyak kegunaan untuk IPA. Misalnya, bank investasi mengotomatiskan beberapa proses manual yang memastikan bahwa setiap perubahan pada data instrumen yang masuk ke hulu tercermin dalam sistem hilir. IPA mentransformasi fungsi ini, menggunakan robot untuk mengambil pengecualian data dari sistem sumber dan memasukkan informasi ke hilir. Solusi RPA bertenaga AI saat ini juga membantu bank dengan tugas padat karya untuk merekonsiliasi faktur dan pembayaran. Beberapa bank meningkatkan rekonsiliasi langsung pembayaran masuk dengan menggunakan RPA, pembelajaran mesin, dan pengenalan karakter optik (OCR). Penggunaan RPA bertenaga AI ini berarti bank memposting piutang lebih cepat, yang membuat rekonsiliasi piutang dan pencocokan pembayaran lebih efisien.

Satu bank telah melaporkan bahwa implementasi RPA-nya yang terdiri dari 220 bot menghasilkan peningkatan proses yang signifikan, termasuk akurasi 100% dalam validasi penutupan akun, pengurangan 88% dalam waktu pemrosesan, dan peningkatan 66% dalam penyelesaian entri perdagangan. Perusahaan perbankan lain menerapkan proses seperti ini karena mereka berusaha untuk tetap kompetitif. Menurut laporan, IPA telah membantu bank lain meningkatkan produktivitas dan menambah tenaga kerjanya. Aplikasi dalam pembiayaan perdagangan, operasi kas, operasi pinjaman, perencanaan pajak, pendaftaran pelanggan, dan lainnya telah mencapai otomatisasi antara 30% dan 70%, yang meningkatkan kualitas dan mengurangi risiko. IPA juga telah mengurangi waktu yang dibutuhkan untuk melatih karyawan. Dengan mengodekan pengetahuan melalui robotika, informasi menjadi tersedia untuk meningkatkan keterampilan karyawan dan membimbing mereka dalam menjalankan tanggung jawab sehari-hari.

Layanan keuangan dan lembaga perbankan penuh dengan peluang untuk sistem AI. Selain mengimplementasikan solusi mereka, seperti contoh-contoh sebelumnya, organisasi-organisasi ini juga berinvestasi, mengakuisisi, atau mengintegrasikan layanan dari ratusan perusahaan FinTech yang telah muncul selama dekade terakhir yang menawarkan berbagai jenis layanan berbasis AI. Penerapan AI pada layanan keuangan tetap menjadi area yang akan terus mengalami evolusi, bahkan revolusi, di tahun-tahun mendatang.

## **BAB 5**

### **AI DALAM MANUFAKTUR DAN ENERGI**

Pemanfaatan data dari Internet of Things (IoT) dan AI serta pembelajaran mesin dalam manufaktur telah mendapatkan momentum berkat program Industri 4.0 di Jerman dan Inisiatif Manufaktur Lanjutan AS di Amerika Serikat. Inisiatif Manufaktur Lanjutan AS menyarankan peluncuran "Inisiatif Manufaktur Lanjutan" yang akan "mendukung inovasi dalam manufaktur lanjutan melalui program penelitian terapan untuk teknologi-teknologi baru yang menjanjikan, kemitraan publik-swasta seputar teknologi yang dapat diterapkan secara luas dan prakompetitif, penciptaan dan penyebaran metodologi desain untuk manufaktur, dan infrastruktur teknologi bersama untuk mendukung kemajuan dalam industri manufaktur yang ada."

Baik sektor manufaktur maupun energi sangat cocok untuk penggunaan AI karena keduanya menghasilkan data dalam jumlah besar dalam berbagai format, dan hal ini akan terus berlanjut di masa mendatang. Sebagian besar data ini dikumpulkan menggunakan sensor yang memantau segala hal, mulai dari peralatan yang digunakan hingga laju proses kimia. Menghubungkan sensor-sensor ini ke komputer melalui jaringan nirkabel atau kabel memungkinkan algoritma AI menggunakan data ini untuk meningkatkan berbagai hal dalam proses manufaktur seperti kualitas, produktivitas, distribusi, keselamatan karyawan, dan dampak lingkungan.

Produsen industri menangkap data melalui pembacaan sensor seperti tegangan, tekanan, suhu, atau getaran mesin pabrik. Parameter kunci lainnya mencakup informasi tentang proses produksi itu sendiri, seperti jumlah produksi, ketersediaan mesin, dan kualitas. Informasi pelengkap tambahan dapat berasal dari log kesalahan, yang dikumpulkan saat mesin beroperasi dan selama acara pemeliharaan, baik terjadwal maupun tidak terjadwal; dari sistem perencanaan sumber daya perusahaan (ERP); dan dari operator manusia atau di tempat lain. Dengan menerapkan model AI pada data ini, bisnis mengidentifikasi masalah yang meningkatkan biaya produksi dan memprediksi situasi yang mengakibatkan waktu henti yang tidak direncanakan. Mengetahui, misalnya, bagaimana tingkat beban setiap mesin memengaruhi kinerja keseluruhan memungkinkan pengambilan keputusan yang lebih baik selama setiap proses produksi. AI juga membantu menentukan peralatan mana yang terbaik untuk proses tertentu dengan durasi tertentu, yang mengarah pada efisiensi lebih lanjut tidak hanya dalam produksi tetapi juga dalam pemeliharaan.

Perusahaan energi juga memanfaatkan sensor dan jaringan nirkabel untuk mengumpulkan data tentang segala hal mulai dari operasi pabrik hingga pola konsumsi individu hingga peristiwa cuaca global yang menyebabkan pemadaman listrik. Mereka kemudian menggunakan model AI untuk memanfaatkan data ini guna menemukan cara meningkatkan produktivitas dan memangkas biaya. Algoritma mengurai data dan membuat prediksi tentang cara menangani kejadian tak terduga, membantu karyawan dalam hal-hal

seperti menentukan kapan mesin memerlukan perawatan, memantau keselamatan personel, dan menentukan efisiensi produksi.

Baik industri manufaktur maupun energi menyimpan sebagian data yang mereka kumpulkan sebagai deret waktu dalam penyimpanan data terpusat yang dikenal sebagai *sejarawan proses*. Sejarawan proses mungkin memiliki data dari 10 tahun atau lebih yang lalu, memberikan informasi yang cukup untuk menjadikan penggunaan AI tidak hanya layak tetapi juga efektif. Terkadang, sebagian besar data ini tidak dapat diakses dari jarak jauh. Namun, menggabungkannya dalam danau data dan mengunggahnya ke cloud kini lebih umum, memungkinkan perusahaan di berbagai industri untuk menerapkan apa yang dikenal sebagai *kembaran digital*.

Penasihat dan profesor ternama Michael Grieves pertama kali menggunakan frasa "kembaran digital" pada tahun 2002, ketika ia bekerja di Universitas Michigan. Kembaran digital adalah model digital virtual dari suatu proses, produk, atau layanan yang memantau, mengumpulkan, dan merekam data dari mesin yang terlibat dalam proses fisik di fasilitas manufaktur atau pembangkit energi dan mereproduksi proses ini secara virtual. Dalam arti sempit tentang bagaimana istilah tersebut digunakan saat ini, kembaran digital bukanlah komponen yang mengendalikan proses. Biasanya dianggap sebagai model data pada inti representasi digital dari peralatan atau proses fisik. Kembaran digital menyediakan konteks untuk aliran data konstan yang dihasilkan dari sensor dan alat lain dalam operasi produksi. Algoritma kemudian memanfaatkan konteks ini untuk membuat keputusan dan menyampaikannya kembali ke sistem kontrol yang kemudian meneruskan instruksi ke aktuator fisik dan peralatan kontrol. Seiring waktu, kembaran digital akan berevolusi dari sekadar proses produksi fisik menjadi pandangan menyeluruh tentang rantai nilai perusahaan, yang mencakup rantai pasokan, proses produksi, inventaris barang jadi, distribusi, pemasaran, dan pengiriman ke pelanggan akhir. Berkat pengenalan AI dan IoT, penerapan kembaran digital tidak hanya menjadi hemat biaya, tetapi juga Gartner menobatkannya sebagai salah satu dari 10 Tren Teknologi Strategis Teratas untuk tahun 2017.

Setelah mengumpulkan dan menganalisis data, kembaran digital menemukan masalah dan menemukan solusi untuk meningkatkan prosedur, menurunkan biaya operasional, mengurangi cacat produk, dan meningkatkan efisiensi, semuanya dengan sedikit campur tangan manusia. Terhubung dalam ekosistem peralatan digital, kembaran digital menerima data melalui *sensor*, dan mengendalikan mesin melalui apa yang dikenal sebagai *aktuator*. Mereka kemudian berkomunikasi langsung dengan mesin fisik dan mengendalikan proses berdasarkan kesimpulan yang telah mereka capai. Kembaran digital paling sering digunakan untuk meningkatkan kinerja aset fisik yang kompleks, seperti kilang minyak atau peralatan manufaktur, tetapi ada tren yang semakin meningkat untuk menggunakannya pada peralatan konsumen seperti mesin pencuci piring, lemari es, dan mesin pembuat kopi yang terhubung.

Salah satu contohnya adalah Mesin Cuci dan Pengering Smart All-In-One dari Whirlpool,<sup>5</sup> yang memungkinkan pelanggan untuk mengoperasikan siklus pencucian dan pengeringan dari jarak jauh serta melacak kemajuannya menggunakan aplikasi seluler.

## 5.1 OPERASI PABRIK DAN PEMELIHARAAN ASET YANG DIOPTIMALKAN

Agar kompetitif di pasar yang sensitif terhadap biaya, memaksimalkan mesin manufaktur yang mahal sangatlah penting. Perusahaan yang memanfaatkan kembar digital yang dikendalikan AI mencapai optimalisasi proses dengan berbagai cara. Algoritma kontrol telah ada di bidang manufaktur dan energi selama beberapa dekade, dengan pengontrol logika terprogram (PLC) berbasis mikroprosesor mulai digunakan pada tahun 1980-an. Penggunaan AI membantu meningkatkan efisiensi proses industri ini dengan memanfaatkan titik data input dan output historis dari sistem kontrol, dan dampaknya terhadap proses, untuk mengoptimalkan parameter proses. Salah satu contoh krusialnya adalah mengendalikan prosedur berkecepatan tinggi secara otomatis, seperti dalam manufaktur diskrit, untuk mencapai output produksi maksimum, waktu henti paling sedikit, dan pengurangan kegagalan komponen terbesar. Penggunaan pemantauan waktu nyata bersama AI membuat operasi di area produksi lebih efisien dengan meningkatkan jadwal produksi dan mengoptimalkan efisiensi mekanis. Mengelola proses produksi dengan AI membantu menentukan bagaimana tingkat beban memengaruhi kinerja jadwal produksi secara real-time, yang menghasilkan proses produksi yang lebih efisien. Peningkatan ini secara keseluruhan memaksimalkan pemanfaatan peralatan. Selain itu, optimalisasi inventaris dan suku cadang juga dapat menghasilkan penghematan yang substansial.

Model AI membantu manajer menentukan kombinasi mesin terbaik untuk setiap proses produksi. Sebelum produksi, parameter permintaan, bahan baku, dan faktor-faktor lainnya diberikan kepada algoritma AI sehingga dapat menentukan operasi rantai produksi yang optimal. Selama proses produksi, peralatan mencatat dan mengevaluasi data real-time. Kemudian, dengan memperhitungkan setiap perbedaan antara rencana dan aktual, model dijalankan ulang dan rencana pelaksanaan diperbarui agar tetap sesuai dengan tujuan. Riwayat proses produksi fisik ini digunakan untuk menyesuaikan model sesuai kebutuhan. Siemens, misalnya, memiliki sistem AI yang memanfaatkan data dari lebih dari 500 sensor yang terpasang pada turbin gasnya sehingga dapat belajar untuk terus menyesuaikan katup bahan bakar guna menciptakan pembakaran yang optimal terlepas dari kondisi cuaca atau kondisi peralatan. Manfaat tambahannya mencakup emisi yang lebih sedikit dan pengendalian polusi yang lebih baik. “Bahkan setelah para ahli melakukan yang terbaik untuk mengoptimalkan emisi nitrogen oksida turbin,” kata Dr. Norbert Gaus, Kepala Riset Digitalisasi dan Otomasi di Siemens Corporate Technology, “sistem AI kami mampu mengurangi emisi hingga sepuluh hingga lima belas persen tambahan.”

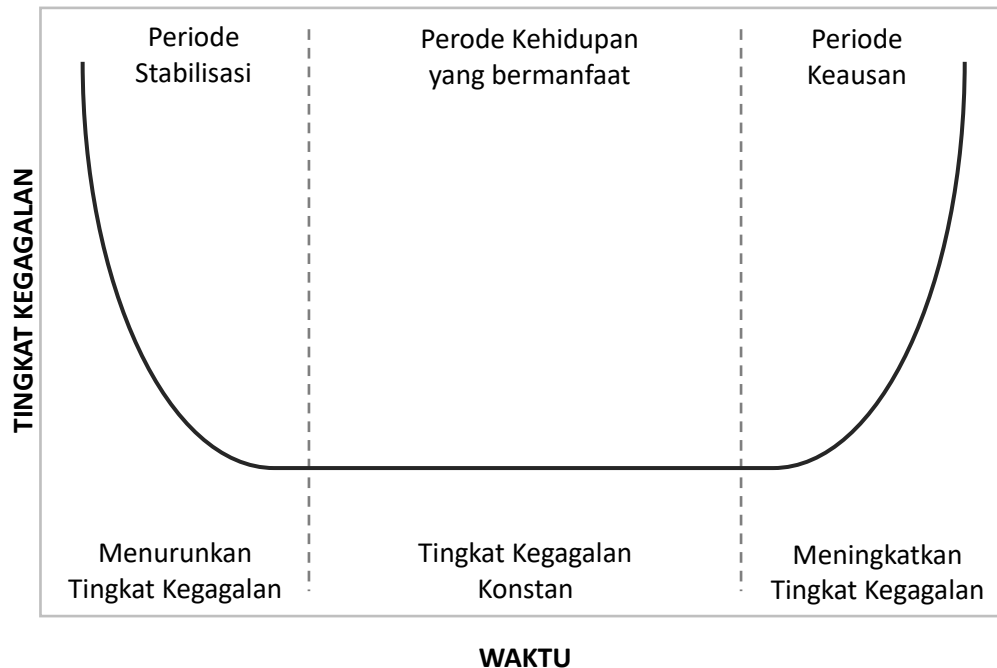
Salah satu masalah terbesar yang terkait dengan manufaktur atau produksi energi adalah waktu henti yang tidak direncanakan yang disebabkan oleh masalah mekanis. Perusahaan sekarang menggunakan model AI untuk mengurangi frekuensi kejadian ini. Aplikasi AI digunakan untuk memprediksi masalah mekanis terlebih dahulu sehingga dapat ditangani sebelum mulai menimbulkan masalah.

Ratusan atau bahkan ribuan sensor pintar dan kemampuan untuk mengumpulkan data pada platform terpusat memungkinkan algoritma AI untuk melakukan *pemeliharaan prediktif* pada mesin, menemukan pola degradasi atau kegagalan sistem berdasarkan data historis.

Pola-pola ini menghasilkan penjadwalan pemeliharaan atau penggantian mesin yang lebih optimal bagi bisnis. Artinya, alih-alih melakukan penghentian darurat, perusahaan menggabungkan tindakan ini ke dalam jadwal waktu henti yang terjadwal secara rutin. Ketika beberapa bagian mesin menjalani pemeliharaan prediktif, waktu henti yang direncanakan dapat dioptimalkan untuk menangani beberapa perbaikan atau penggantian suku cadang. Jika ini tidak direncanakan, setiap komponen dapat menyebabkan waktu henti ketika rusak. Bahkan sedikit peningkatan dalam manajemen aset berpotensi menghemat biaya bisnis dan meningkatkan pendapatan. Departemen Energi AS telah mengindikasikan bahwa dalam industri kelistrikan, misalnya, perusahaan mengalami pengurangan kerusakan hingga 70% dan pengurangan waktu henti sebesar 35% ketika perusahaan memulai program pemeliharaan prediktif yang fungsional.

Pertimbangkan sebuah contoh. Memproduksi polietilena berdensitas rendah (LDPE), termoplastik yang terbuat dari etilena, membutuhkan tekanan yang sangat besar. Tekanan ini dihasilkan oleh sesuatu yang disebut hiperkompresor, yang dapat memberikan tekanan hingga 50.000 pon per inci persegi untuk membentuk etilena menjadi jenis plastik esensial ini. Tekanan sebesar ini memberikan beban yang sangat besar pada hiperkompresor, dan kompresor tersebut akan mengalami penurunan kinerja berkali-kali dalam setahun. Cara untuk memitigasi masalah ini sangat penting bagi bisnis manufaktur LDPE, karena waktu henti dapat merugikan mereka jutaan dolar per tahun.

Dengan menggunakan model AI, perusahaan secara otomatis mengkalibrasi peralatan berdasarkan riwayat kinerja dan pengaturannya, sehingga menghasilkan perbedaan yang sangat besar dalam mengurangi waktu henti yang tidak direncanakan. Gangguan jenis ini menelan biaya sekitar \$50 miliar per tahun<sup>8</sup> di seluruh sektor manufaktur, dan perkiraan menunjukkan bahwa pemeliharaan preventif yang dipelopori oleh model AI akan mengurangi biaya ini sebesar 10% hingga 20%. Beberapa produsen bahkan telah menggunakan pembelajaran mesin untuk tidak hanya memprediksi kapan akan terjadi kegagalan, tetapi juga mengidentifikasi kapan komponen berada di akhir siklus hidupnya dan perlu diganti. Model AI ini dirancang untuk dapat mengetahui apakah komponen tersebut mendekati siklus hidup produktifnya berdasarkan tingkat kegagalan dan informasi lainnya (lihat Gambar 5.1). Data dari pola penggunaan dan harapan hidup peralatan dapat digunakan untuk membuat model AI yang berupaya mengoptimalkan harapan hidup mesin.



**Gambar 5.1** Heuristik yang menunjukkan tingkat kegagalan yang berbeda selama siklus hidup komponen peralatan.

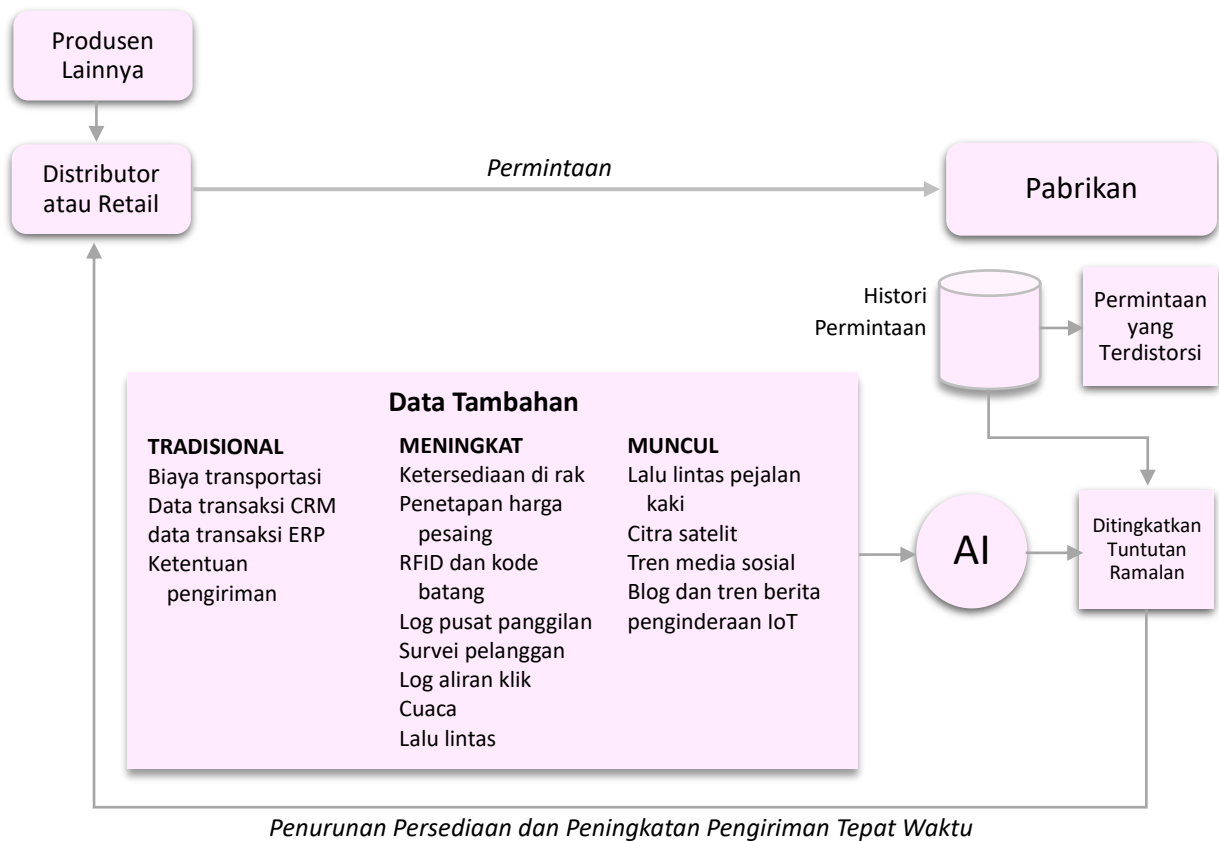
## 5.2 SIKLUS HIDUP PRODUKSI OTOMATIS

Siklus hidup produksi yang sebagian atau sepenuhnya otomatis semakin dikendalikan dan dikoordinasikan oleh model AI. Sistem informasi, terutama di lingkungan manufaktur berkapasitas tinggi, merencanakan, mengoordinasikan, mengendalikan, dan mengevaluasi proses manufaktur produksi massal yang menghasilkan barang-barang seperti mobil, mesin pencuci piring, dan pemanggang roti. Fleksibilitas model AI memungkinkan adaptasi lini produksi ke produksi skala kecil atau produksi yang disesuaikan yang jika tidak akan mengganggu proses produksi normal. Hal ini memungkinkan produsen untuk lebih mudah menciptakan barang-barang yang dipersonalisasi yang semakin diminati konsumen. Sistem informasi juga mendukung pekerja di rantai produksi, menjembatani kesenjangan antara pekerja mesin dan manusia. Ketika pekerja lebih terinformasi tentang bagaimana proses bekerja di stasiun tertentu, semakin baik mereka mampu membuat keputusan yang tepat dan terinformasi.

*Kontrol kualitas* adalah area lain di mana algoritma AI mengambil alih peran manusia. Manusia, bukan mesin, biasanya bertanggung jawab atas pengendalian kualitas karena pengendalian kualitas merupakan fungsi visual atau memerlukan analisis laboratorium. Dulu, manusia lebih mudah melihat warna yang tidak sesuai, label yang tidak sejajar, atau kemasan yang rusak. Namun, model pemrosesan gambar berbasis pembelajaran mendalam kini memungkinkan mesin untuk memeriksa produk dan mengidentifikasi cacat lebih cepat dan akurat daripada manusia.

## 5.3 OPTIMALISASI RANTAI PASOKAN

Rantai pasokan yang efisien memiliki dampak signifikan terhadap kesuksesan perusahaan, dan AI diposisikan dengan baik untuk membuat peningkatan substansial dalam peramalan permintaan, optimalisasi inventaris, dan logistik distribusi. AI digunakan untuk *meramalkan permintaan* agar para peserta dalam rantai pasokan berfungsi secara optimal. McKinsey memprediksi bahwa pembelajaran mesin dapat mengurangi kesalahan peramalan rantai pasokan hingga 50% dan mengurangi kehilangan penjualan hingga 65% dengan membuat produk lebih mudah tersedia saat dibutuhkan. Aplikasi AI juga membantu menurunkan biaya pergudangan dan pengangkutan barang serta biaya manajemen rantai pasokan. Beberapa perusahaan bahkan menggunakan AI untuk meningkatkan skala optimalisasi inventaris di seluruh lokasi distribusi, menggunakan data penjualan yang ada untuk menentukan permintaan serta kinerja pengiriman tepat waktu ke pelanggan.



**Gambar 5.2** Perkiraan permintaan menggunakan penjualan historis dan sumber data baru.

Peramalan permintaan semakin mencakup berbagai jenis data (lihat Gambar 5.2). Produsen mengumpulkan data penjualan dan inventaris di seluruh jaringan pengecer, mengumpulkan informasi tentang penjualan sebelumnya di setiap lokasi berdasarkan produk. Karena sebagian besar produsen tidak sering menjual langsung ke konsumen, mereka memasukkan informasi pemasaran dalam proses perkiraan. Menggunakan informasi pemasaran memerlukan pemrosesan bahasa alami (NLP) dari sumber-sumber seperti berita dan blog, peringkat produk, dan sumber lainnya. Pengecer memiliki keuntungan menggabungkan weblog dan informasi aliran klik pelanggan, data dari program loyalitas,

transkrip pusat panggilan, dan survei pelanggan. NLP kemudian mengekstrak sentimen dan tren permintaan ke dalam deret waktu, yang dapat digunakan untuk mengurangi kesalahan perkiraan permintaan.

Produsen juga mengumpulkan informasi pemasok, baik dengan mewajibkan pemasok untuk memberikan informasi atau mengumpulkannya dari sumber publik. Informasi pemasok sering kali mencakup kapasitas pemasok saat ini dan ukuran pelanggan teratas, serta transaksi pemasok sebelumnya dengan pelanggan lain dengan persyaratan serupa, seperti lokasi pengiriman dan waktu tunggu. Beberapa perusahaan bahkan bereksperimen dengan penyertaan informasi cuaca dan lalu lintas, seperti kemacetan lalu lintas, dalam proses pemodelan AI mereka untuk meningkatkan akurasi.

Untuk situasi di mana tidak ada data penjualan historis, seperti peluncuran produk baru, perusahaan menggunakan proksi dalam proses pemodelan. Peramalan permintaan untuk produk baru sangat penting, karena produk-produk baru ini cenderung mendorong penjualan yang tinggi. Untuk memahami seberapa besar perhatian yang diberikan oleh calon pelanggan terhadap berbagai produk, baik baru maupun lama, model AI diberikan informasi kueri penelusuran: berapa kali seseorang menelusuri istilah tertentu berdasarkan lokasi geografis dan waktu. Informasi ini, bersama dengan data lain yang tersedia seperti penjualan produk terkait, digunakan sebagai input dalam model ML untuk memprediksi permintaan.

#### **5.4 MANAJEMEN INVENTARIS DAN LOGISTIK DISTRIBUSI**

Tujuan setiap produsen atau distributor adalah memaksimalkan penjualan. Kendala bagi mereka adalah ruang rak yang terbatas dan seringkali tidak ada ruang penyimpanan. Stok yang berlebihan pada satu produk dapat menyebabkan penurunan potensi penjualan produk lain yang terjual habis atau tidak pernah tersedia sama sekali. Untuk mengatasi masalah ini, perusahaan mengoptimalkan *manajemen inventaris* mereka dengan memungkinkan algoritma AI untuk menggunakan prakiraan permintaan yang ada dan data lainnya untuk mengoptimalkan tingkat stok dan mengurangi kerugian dengan memprediksi keadaan yang mungkin memengaruhi pengiriman produk.

*Logistik distribusi* telah didefinisikan sebagai mendapatkan produk yang tepat, dalam jumlah yang tepat, dengan kualitas yang tepat, di tempat yang tepat, pada waktu yang tepat, dengan biaya yang tepat untuk pelanggan yang tepat.<sup>10</sup> Produk mungkin perlu didistribusikan dari gudang ke berbagai jenis lokasi ritel, seperti toko elektronik, supermarket, pom bensin, dan toko swalayan. Beberapa pengecer memiliki tantangan tambahan dalam pengiriman dari gudang atau lokasi ritel ke pelanggan akhir. Secara historis, hasil model riset operasi berdasarkan data seperti asal dan tujuan (OND) digunakan untuk menentukan hal-hal seperti rute kendaraan, dan memanfaatkan hubungan logis dan matematis untuk menentukan cara paling efisien untuk mengirimkan barang. Kemampuan model AI untuk belajar dari pengalaman memberikan alternatif yang mendekati waktu nyata bagi model-model ini.

Model AI untuk peramalan permintaan, optimalisasi inventaris, dan logistik distribusi perlu berinteraksi satu sama lain karena saling bergantung. Data yang dihasilkan oleh pemasaran dan penjualan digunakan untuk melatih algoritma AI, seperti model peramalan

permintaan, dalam rantai pasokan. Dalam Bab 3, kita melihat bahwa fungsi periklanan, pemasaran, penjualan, dan layanan pelanggan mulai semakin tumpang tindih karena menggunakan model yang serupa dan kumpulan data yang sama; wawasan yang dihasilkan dibagikan di antara fungsi-fungsi ini. Seiring meningkatnya penggunaan AI dalam industri berat dan perusahaan manufaktur dan energi, akan ada lebih banyak tumpang tindih antara fungsi-fungsi yang berhadapan dengan pelanggan dengan fungsi-fungsi rantai pasokan di sana juga.

## 5.5 PERAMALAN TENAGA LISTRIK DAN RESPONS PERMINTAAN

Jika perusahaan listrik dapat meramalkan permintaan energi dengan lebih baik setiap seperempat jam, volatilitas harga dan pemadaman listrik akan berkurang. Namun, hal ini sulit dilakukan karena variasi permintaan yang halus dan kompleksitas dalam jaringan listrik. Operasi jaringan energi yang sukses membutuhkan kemampuan untuk mengalihkan beban secara terus-menerus guna menyediakan jumlah daya yang dibutuhkan secara tepat setiap jam, baik bagi pelanggan komersial maupun residensial. Prediksi jumlah energi yang dibutuhkan dan kepada siapa energi tersebut harus disalurkan harus sangat akurat, yang merupakan tantangan tersendiri, dan pembangkitan listrik di lokasi dengan panel surya dan penyimpanan di lokasi atau di tingkat jaringan, seperti baterai, semakin mempersulitnya.

Untuk merespons lonjakan permintaan yang tiba-tiba, perusahaan listrik diharuskan menyediakan daya tambahan ketika diminta oleh operator jaringan, biasanya dalam waktu singkat. Generator mampu menyediakan daya ini karena "spinning reserve" – yaitu, kapasitas pembangkit ekstra yang tersedia bagi generator yang sudah terhubung ke jaringan listrik. Kemampuan model AI untuk membuat prakiraan yang lebih akurat telah mulai memengaruhi area ini. Sensor dan meter pintar mengumpulkan data waktu nyata di seluruh jaringan energi, dan sistem kendali pengawasan dan akuisisi data (SCADA) memungkinkan bisnis untuk mengontrol proses secara lokal atau jarak jauh, mengumpulkan dan memantau data meter pintar, berinteraksi dengan perangkat keras seperti sensor dan motor, dan merekam informasi yang dihasilkan. Data juga diperoleh dari prakiraan pola cuaca dan kondisi atmosfer serta pemantauan jumlah energi yang diproduksi di pembangkit listrik dan yang dikonsumsi.

Terdapat juga data yang memadai terkait konsumsi pelanggan, perilaku pelanggan, dan data pihak ketiga untuk memungkinkan algoritma AI membuat prediksi penggunaan yang berhasil. Hal ini memungkinkan untuk memiliki pemahaman yang lebih baik tentang apa yang terjadi antara gardu induk dan meter penggunaan. Akibatnya, perusahaan listrik kini memiliki gambaran yang lebih terperinci tentang jaringan listrik setiap jam. Prakiraan menjadi lebih akurat dan stabil, dan perangkat pemantauan mandiri yang menandai kesalahan memberikan informasi yang diperlukan operator untuk membuat keputusan yang lebih baik.

Contoh potensi semua ini datang dari Google DeepMind, yang pada awal 2017 mengungkapkan bahwa mereka sedang berdiskusi dengan National Grid Inggris untuk membantu menyeimbangkan pasokan dan permintaan listrik di seluruh Britania Raya. Idennya adalah untuk memprediksi lonjakan permintaan listrik dengan menggunakan data dari prakiraan cuaca dan pencarian internet yang dikombinasikan dengan informasi tentang data energi pelanggan untuk menciptakan model yang dapat digunakan untuk membuat jaringan

listrik Inggris lebih efisien, menghemat energi, dan menghindari kelebihan beban. Google memperkirakan biaya dapat dikurangi sebesar 10%.

Perusahaan energi juga telah menjajaki solusi respons permintaan: kemampuan agar pelanggan lokal sedikit mengurangi penggunaan listrik pada jam sibuk, sehingga penyedia tidak perlu menyalakan apa yang dikenal sebagai "peaker" – yaitu, pembangkit listrik tenaga gas yang beroperasi untuk mengisi situasi permintaan puncak dan menjual listrik dengan tarif lebih tinggi. Ketika permintaan bahkan lebih tinggi daripada yang dapat dipenuhi oleh peaker, perusahaan energi sering kali mengatasi situasi ini dengan membeli dan menyalurkan listrik dari wilayah lain dengan biaya lebih tinggi. Namun, solusi respons permintaan berbasis AI yang potensial sudah tersedia dalam bentuk termostat pintar, Nest. Setelah Nest dipasang di rumah, unit tersebut mulai mengumpulkan data, mempelajari kapan pemilik rumah membutuhkan lebih banyak atau lebih sedikit pemanas dan pendingin, serta menyesuaikan sistem agar sesuai. Hal ini tidak hanya menghemat uang pengguna Nest untuk tagihan pemanas; tetapi juga memungkinkan perusahaan utilitas untuk menginstruksikan Nest untuk, misalnya, menaikkan termostat sedikit selama gelombang panas, ketika listrik paling mahal.

Austin Energy adalah salah satu yang pertama melakukan hal ini, menawarkan potongan harga kecil satu kali kepada pelanggan Nest untuk hak menyediakan layanan tersebut. Solusi respons permintaan yang ringkas ini telah membantu Austin menghindari penggunaan generator tambahan, sehingga menghemat energi dan uang. Perubahan yang dilakukan Nest begitu halus sehingga tidak menimbulkan keluhan pelanggan, sekaligus mengurangi permintaan daya. Solusi yang sama atau serupa juga memungkinkan perusahaan energi menghindari pengeluaran ratusan juta dolar yang diperlukan untuk membangun pembangkit listrik tambahan. Jenis optimasi mikro ini akan menjadi lebih penting dan bahkan mungkin secara aktif dicari di dunia yang lebih sadar lingkungan.

## 5.6 PRODUKSI MINYAK

Perusahaan-perusahaan energi, seperti Shell dan BP, saat ini menginvestasikan miliaran dolar untuk memungkinkan kilang-kilang baru, ladang-ladang minyak, dan anjungan pengeboran air dalam memanfaatkan algoritma AI baru. Model AI digunakan dalam eksplorasi minyak dan gas di berbagai bidang, termasuk pencitraan dan interpretasi seismik, analisis log sumur, prakiraan produksi, dan analisis properti reservoir.

Pencitraan seismik adalah teknik yang digunakan perusahaan energi untuk menemukan cadangan minyak dan gas. Perusahaan-perusahaan minyak menggunakan peralatan untuk menghasilkan gelombang seismik dan kemudian menangkap perambatan gelombang ini melalui bumi menggunakan sensor: pada dasarnya, menggunakan pencitraan seismik untuk mengambil sonogram dari apa yang ada di bawah tanah. Ahli geofisika mempelajari bagaimana gelombang-gelombang ini bergerak melalui tanah dan menafsirkan data ini untuk memprediksi apa yang ada jauh di bawah permukaan. Setelah seorang ahli manusia memberi label pada data ini, sistem AI menerapkan pembelajaran terawasi untuk mendeteksi kondisi serupa dalam citra seismik tiga dimensi. Hal ini lebih baik dalam menargetkan peluang pengeboran dan memungkinkan analisis berkualitas lebih tinggi, karena

model AI dapat mempertimbangkan semua data, sedangkan geosains hanya dapat secara manual menggunakan sebagian kecil data yang tersedia bagi mereka.

Ketika cadangan minyak atau gas ini ditemukan, perusahaan mengebor lubang di lapangan untuk menentukan sifat-sifat reservoir tersebut, termasuk informasi penting tentang formasi batuan bawah permukaan. Inti lubang bor (bagian sumur sedalam 1.000 – 1.500 kaki) diekstraksi dan dianalisis di laboratorium inti untuk mengumpulkan sifat-sifat batuan pada resolusi skala milimeter. Analisis laboratorium inti didasarkan pada berbagai sensor, termasuk pemindaian CT, spektrometri fluoresensi sinar-X, dan lainnya. Kemudian, saat sumur dibor, sensor dikirim bersama bor ke dalam tanah, mengumpulkan data setiap beberapa inci. Data ini disimpan dalam log sumur. Algoritma pembelajaran mesin menggunakan data ini dari analisis inti bersama dengan log sumur untuk memprediksi sifat-sifat batuan untuk area minat baru. Dengan menggunakan informasi ini, ahli geofisika menentukan jenis dan sifat batuan di bawah permukaan, termasuk karakteristik seperti kepadatan, porositas, dan radioaktivitasnya. Informasi ini memungkinkan para ilmuwan untuk menentukan keberadaan dan nilai hidrokarbon dalam batuan, dan pada akhirnya, fraksi minyak atau gas yang dapat diambil dari setiap reservoir lapangan.

Data yang disimpan dalam log sumur juga digunakan untuk menentukan arah pengeboran. Menetapkan arah pengeboran dapat menghasilkan sumur vertikal sederhana, tetapi juga memungkinkan keberhasilan eksploitasi prospek minyak non-konvensional. Non-konvensional dalam hal ini mengacu pada jenis hidrokarbon tertentu yang diperoleh dengan cara non-konvensional, seperti fracking. Hidrokarbon ini membutuhkan metode ekstraksi yang berbeda untuk membebaskan minyak. Pengeboran sumur non-konvensional menghadirkan beberapa tantangan. Untuk mengembangkan reservoir ini diperlukan pengetahuan tentang karakteristik penyelesaian yang optimal dan metode stimulasi untuk reservoir permeabilitas rendah, serta pemahaman tentang peran rekahan alami dalam aliran fluida.

Fisika aliran fluida melalui berbagai media bisa sangat rumit dan sulit dimodelkan. Untuk membantu mengeluarkan minyak dari dalam tanah, model AI digunakan untuk setiap langkah ini. Kumpulan data yang diperoleh dari sumur-sumur yang telah dibor sebelumnya, serta data dari pihak ketiga yang mencakup header sumur, log, penyelesaian, dan produksi bulanan, digunakan untuk melatih model AI yang memprediksi kinerja sumur-sumur yang ada di masa mendatang menggunakan riwayatnya, serta data dari sumur-sumur terdekat yang dibor di lokasi geologis yang serupa. Untungnya, sensor menangkap semakin banyak jenis data pada frekuensi yang lebih tinggi, termasuk karakteristik seperti kepadatan batuan, hambatan listrik batuan yang dikenal sebagai resistivitas, dan radioaktivitas alami batuan, yang terdiri dari sinar gamma. Data tambahan ini memungkinkan model AI yang menggunakan pembelajaran mendalam untuk secara signifikan membantu kemampuan bisnis dalam memperkirakan prospek minyak.

Meskipun log sumur menyediakan lebih banyak data, seringkali terdapat masalah dengan pengumpulan data itu sendiri karena sensor yang tidak berfungsi atau kondisi yang merugikan di dalam lubang bor. Dalam kasus ini, model AI mengisi data yang hilang atau menghapus data yang buruk dengan mengidentifikasi anomali berdasarkan apa yang telah

dipelajari algoritma dari set data sebelumnya. Algoritma juga menggunakan set data dari sumur terdekat atau serupa untuk menyediakan lebih banyak data kepada ahli geofisika dengan relatif cepat, membantu mereka menentukan di mana dan bagaimana mengebor. Untuk menentukan di mana akan mengebor, sifat batuan dari analisis inti dan log sumur digunakan untuk mensimulasikan rekahan hidrolik, memperkirakan jumlah hidrokarbon, dan menghitung lintasan sumur. Model-model ini, misalnya, dapat membantu merekomendasikan agar kru pengeboran mengebor pada sudut tertentu ke bawah, kemudian berbalik dan mengebor secara horizontal untuk menghindari batuan yang lebih keras, dan kemudian mengebor ke bawah lagi untuk menemukan lokasi yang paling produktif. Diaktifkan oleh AI, pengetahuan tentang sifat-sifat reservoir tertentu ini memungkinkan tim eksplorasi dan produksi untuk meningkatkan perkiraan produksi, baik dalam hal menentukan perkiraan produksi umum maupun bagaimana parameter pompa minyak dapat memengaruhi perkiraan tersebut.

BP memanfaatkan kemampuan ini dalam upayanya untuk mengeksploitasi ladang Wyoming yang menua. Bermitra dengan perusahaan rintisan Kelvin yang berbasis di San Francisco, BP memasang ribuan sensor di ratusan sumur. Algoritma AI Kelvin memantau aliran data besar yang berasal dari sumur-sumur ini, yang memungkinkan mereka membangun kembaran digital lapangan. Kembaran ini cukup canggih untuk memprediksi dampak pembukaan katup di satu sisi lapangan terhadap sisi yang berlawanan, dan menentukan bagaimana pembacaan tekanan akan berubah di sisi yang lain. Ketika sistem beroperasi, BP memutuskan untuk membiarkan sistem AI berjalan secara independen, tanpa campur tangan manusia. Perusahaan memperkirakan bahwa metana yang dikeluarkan dari lapangan turun 74% karena peningkatan pemantauan dan pemeliharaan yang dimungkinkan oleh model AI. Selain itu, produksi gas naik 20% dan biaya turun 22%.

## **5.7 PERDAGANGAN ENERGI**

Perusahaan-perusahaan di bidang perdagangan energi dan komoditas terus meningkatkan kemampuan mereka untuk mendeteksi peristiwa pasar dengan cepat dan mengevaluasi dampaknya terhadap operasi pasar jangka pendek guna menemukan peluang perdagangan baru. Mereka melakukan ini dengan mengembangkan pemahaman kualitatif dan kuantitatif yang lebih baik tentang dampak peristiwa-peristiwa ini. Tujuannya adalah untuk mewujudkan peningkatan pendapatan dan margin tanpa perlu meningkatkan jejak organisasi atau aset secara signifikan.

Pertimbangkan contoh sederhana dari pedagang minyak bumi global, yang ditunjukkan pada Gambar 5.3. Misalkan perusahaan memiliki hak kepemilikan atas kapal bensin di Pantai Teluk AS dan komitmen untuk mengirimkannya di Pelabuhan New York, dengan mengabaikan kompleksitas beberapa kendala fisik dan perbedaan spesifikasi di berbagai wilayah. Perusahaan tersebut mendengar tentang suatu peristiwa: sebuah kilang di Pantai Barat telah ditutup karena kebakaran.

Analisis perusahaan percaya bahwa hal ini akan menyebabkan kenaikan harga sementara di California dan ingin memanfaatkannya dengan mengalihkan kapal tersebut ke Pantai Barat. Namun, sebelum membuat keputusan ini, para analis ingin memastikan bahwa kemungkinan akan ada kenaikan harga di California dan bahwa perusahaan masih akan mampu memenuhi kewajibannya di Pelabuhan New York tanpa kehilangan semua keuntungan dari Pantai Barat.



**Gambar 5.3 Skenario perdagangan energi.**

Untuk mengevaluasi pertimbangan ini secara cepat, perusahaan menggunakan sistem AI untuk melihat apakah ada kapal lain yang ditujukan ke California, atau apakah kapal dengan jenis bensin yang sama dapat dialihkan ke California dalam jangka waktu tiga hari. Jika tidak, berapa kisaran kemungkinan kenaikan harga sementara? Perusahaan juga menggunakan AI untuk mengetahui apakah ada kapal yang tersedia untuk diperoleh dan diarahkan menuju Pelabuhan New York dengan rentang pengiriman, rentang harga, dan biaya demurrage yang wajar, serta berapa total dampak biayanya. (Biaya demurrage mengacu pada biaya yang harus dibayarkan kepada pemilik kapal sewaan jika terjadi kegagalan pemuatan atau pembongkaran kapal tersebut dalam waktu yang disepakati.) Model AI menyediakan tanggal dan rentang harga spesifik yang akan berfungsi untuk memonetisasi situasi tersebut.

Model AI digunakan untuk menyeleksi perusahaan atau grup yang akan dihubungi secara eksternal dengan permintaan yang relevan, serta memungkinkan perusahaan untuk mendapatkan pemahaman kualitatif dan kuantitatif yang lebih baik tentang pasar jangka pendek dan bagaimana perusahaan harus meresponsnya. Tanpa analisis informasi yang cepat dan relevan, perusahaan tidak akan dapat memanfaatkan peristiwa di California, karena sebagian besar peristiwa seperti ini berumur pendek dan memerlukan tindakan segera.

Asumsi sebelumnya adalah bahwa sebagian besar peristiwa, seperti kebakaran kilang dalam contoh ini, sangat terlihat. Namun, ketika perusahaan mulai mengumpulkan data intelijen energi mereka, mereka menyadari bahwa situasi seperti ini, baik besar maupun kecil, terjadi setiap hari. Sistem AI telah memungkinkan bisnis untuk mengidentifikasi peluang mikro baru, menilai dengan cepat dampak dari peluang makro yang lebih terlihat serta peluang mikro yang baru terlihat ini, dan memutuskan apakah akan memanfaatkannya.

Meskipun bab ini membahas banyak contoh bagaimana AI digunakan dalam industri manufaktur dan energi, ini adalah area di mana penggunaan AI masih sangat baru dan jauh dari jenuh. Dalam Bab 12, kami membahas pola teknis penginderaan ambien dan kontrol fisik yang relevan untuk kasus penggunaan ini.

Pada bab selanjutnya, kita akan membahas bagaimana kecepatan dan fleksibilitas AI telah membantu industri kesehatan.

## **BAB 6**

### **AI DALAM LAYANAN KESEHATAN**

Kondisi AI dalam layanan kesehatan saat ini merupakan kombinasi dari kesuksesan yang gemilang, potensi yang sangat besar, dan tingkat frustrasi yang cukup tinggi. Jumlah data yang tersedia untuk membantu dokter dan peneliti medis mendiagnosis dan mengobati penyakit sangat besar, tetapi sistem yang ada dapat terfragmentasi dan sulit digunakan. Sistem komputerisasi yang diterapkan pada perawatan pasien dimaksudkan untuk membebaskan dokter dari pekerjaan yang tidak perlu. Sayangnya, sistem ini terlalu sering menambah beban administratif yang mengurangi efektivitas interaksi pasien-dokter, alih-alih meningkatkannya. Beberapa dokter melaporkan bahwa mereka menghabiskan hingga separuh waktu mereka untuk mencoba mengoordinasikan solusi kesehatan yang tidak terhubung.

Meskipun demikian, sistem berbasis AI telah berhasil digunakan untuk memenuhi berbagai kebutuhan perawatan kesehatan di berbagai bidang, termasuk penemuan obat farmasi, diagnosis penyakit, dan perawatan rumah sakit.

#### **6.1 PENEMUAN OBAT FARMASI**

Penemuan obat merupakan salah satu bentuk penelitian paling signifikan yang saat ini dilakukan dengan bantuan algoritma AI. Penggunaan ilmu data dan statistik dalam penemuan obat bukanlah hal baru, tetapi kimia yang digunakan untuk membuat molekul obat cukup rumit. Salah satu alasannya adalah molekul seringkali mengandung serangkaian gugus fungsi, yang terdiri dari satu atau lebih atom yang, ketika digabungkan, menghasilkan reaksi kimia yang sama atau serupa, terlepas dari molekul tempat mereka berada.

Untuk merancang molekul obat, gugus fungsi penting harus berada di tempat yang tepat agar tubuh dapat menyerap molekul tersebut, bergerak ke lokasi yang diinginkan (misalnya, lokasi tumor atau infeksi), berinteraksi dengan lokasi yang diinginkan, dan kemudian dikeluarkan dari tubuh. Namun, karena molekul obat cenderung besar dan rumit serta biasanya mengandung beberapa gugus fungsi, menentukan gugus mana yang akan dimasukkan serta di mana menambahkannya dapat menjadi tantangan.

Membuat obat di laboratorium bahkan lebih sulit lagi. Mensintesis zat kimia apa pun dengan tingkat kerumitan seperti ini mendorong sains hingga batasnya saat ini. Hal ini karena untuk mensintesis obat, Anda menggabungkan molekul yang lebih kecil, yang dikenal sebagai prekursor, dengan molekul lain. Selangkah demi selangkah, saat Anda menambahkan lebih banyak, molekul-molekul ini tumbuh semakin besar. Dengan menggunakan teknik yang tersedia saat ini, biasanya rumit untuk menemukan cara menambahkan gugus fungsi di tempat yang Anda inginkan tanpa merusak apa yang sudah Anda miliki.

Akan menghemat waktu, uang, dan upaya untuk mengetahui secara teoritis apakah suatu obat dapat dibuat sebelum mencoba membuatnya di laboratorium. Itulah sebabnya perusahaan saat ini menggunakan algoritma AI yang dilatih pada data reaksi kimia bervolume besar untuk menyarankan rute terbaik untuk membuat molekul obat yang lebih rumit ini.

Meskipun penggunaan AI belum menunjukkan banyak hasil praktis, bidang ini tetap aktif dan ada beberapa hasil awal yang menjanjikan. Salah satu contohnya adalah bagaimana model AI membantu perusahaan farmasi secara radikal mempersingkat proses pengembangan obat, yang biasanya memakan waktu tujuh hingga delapan tahun, termasuk tiga hingga empat tahun penelitian senyawa. Dengan menggunakan model AI *jaringan adversarial generatif* (GAN), perusahaan tersebut berhasil menguji 30.000 senyawa berbeda dan menentukan molekul tepat yang dapat digunakan untuk pengujian pada hewan dan manusia – sehingga memperpendek periode penelitian dari tiga hingga empat tahun menjadi hanya 46 hari.

## 6.2 UJI KLINIS

Uji klinis membutuhkan biaya yang cukup besar, dan biaya ini terus meningkat. Pembuatan obat dan peluncurannya di pasaran bisa mencapai Rp. 10 hingga Rp. 20 Triliun. Uji klinis juga memakan waktu. Hal ini terutama berlaku untuk pengobatan gangguan yang perkembangannya lambat dan sulit dipantau, termasuk penyakit neurologis, neurodegeneratif, psikiatrik, degeneratif, dan yang berkaitan dengan penuaan. Rata-rata, penemuan obat awal dan pengujian praklinis membutuhkan waktu tiga hingga enam tahun; kemudian, dibutuhkan enam hingga tujuh tahun untuk tiga fase uji klinis. Setelah itu, Anda perlu menyerahkan data yang diperlukan kepada regulator, seperti Badan Pengawas Obat dan Makanan (FDA) di Amerika Serikat atau Badan Pengawas Obat dan Makanan Eropa (EMA) di Eropa, agar obat tersebut disetujui untuk dijual. Pemantauan pasca-persetujuan perlu dilakukan.

Dua faktor yang meningkatkan biaya uji klinis adalah durasi dan jumlah pasien yang dibutuhkan untuk berpartisipasi. Faktor lainnya adalah fakta bahwa biomarker yang diukur selama uji klinis untuk menentukan apakah suatu obat bekerja mungkin melibatkan penilaian perilaku oleh dokter yang dilakukan melalui observasi dan penggunaan stopwatch. Misalnya, suatu tes mungkin mencakup penghitungan berapa kali Anda dapat menyentuh hidung dan kemudian telapak tangan Anda yang lain secara bergantian dalam satu menit. Tes lain mungkin melibatkan pengukuran seberapa jauh Anda dapat berjalan dalam enam menit. Tes biomarker yang agak primitif ini membutuhkan banyak waktu untuk dilakukan dan memerlukan kohort uji coba yang besar. Karena diukur secara subjektif, sulit untuk dievaluasi tanpa sampel yang besar. Karena banyak pengukuran dilakukan secara manual dengan mengamati dan menggunakan stopwatch, hasilnya juga tidak presisi – sehingga membutuhkan durasi uji coba yang lebih lama untuk mengukur efek obat. Tes-tes ini juga mengharuskan pasien untuk datang ke klinik agar dapat diukur, alih-alih menilai dirinya di lingkungan rumah aslinya. Itu berarti mendapatkan lebih banyak pasien dan lebih banyak dokter, yang meningkatkan biaya.

AI secara dramatis meningkatkan situasi ini dengan meminta subjek uji coba menggunakan teknologi yang dapat dikenakan seperti jam tangan pintar atau sensor lainnya. Perangkat ini mengumpulkan data objektif dan terukur yang dapat dilakukan selama aktivitas sehari-hari. Hasilnya adalah variabel pasien yang lebih baik dan lebih produktif dalam bentuk data yang dapat diekstraksi dan digabungkan dengan data lain tentang penyakit tertentu, baik dalam studi klinis yang lebih kecil maupun lebih ekstensif. Pendekatan berbasis AI ini tidak

hanya dapat mengurangi waktu uji coba obat dan populasi pasien uji coba; pendekatan ini juga mengurangi waktu pemasaran, sehingga berpotensi menghemat ratusan juta dolar. Hal ini dilakukan dengan mengurangi jumlah pasien yang dibutuhkan untuk uji coba serta memantau secara tepat bagaimana pengobatan bekerja. Selain itu, kemudahan bagi pasien mengurangi angka putus asa, sehingga uji coba tetap tepat waktu dan sesuai anggaran.

Aspek lain yang mahal dari proses uji klinis adalah menemukan dan merekrut pasien untuk penelitian. Menemukan pasien yang cukup merupakan proses yang melelahkan. Banyak uji klinis dibatalkan karena tidak dapat merekrut pasien yang cukup. Untuk mengatasi hal ini, direncanakan lebih banyak lokasi klinis di lokasi yang berbeda, sehingga meningkatkan biaya uji klinis. Dalam Bab 3, kami membahas beberapa teknik yang digunakan untuk mendapatkan pelanggan bagi bisnis e-commerce dengan menargetkan audiens yang relevan untuk mengiklankan produk tertentu. Dengan menggunakan metode yang sama, menggunakan data dari media sosial, tayangan iklan, perilaku penelusuran, dan informasi terkait lainnya, keberhasilan rekrutmen pasien meningkat secara signifikan untuk banyak uji klinis. Hal ini khususnya relevan untuk uji klinis penyakit langka di mana lebih sulit untuk menemukan dan mempertahankan pasien dalam uji klinis.

### **6.3 DIAGNOSIS PENYAKIT**

Bagi ahli patologi, citra merupakan sumber penting informasi diagnostik dan prognostik. Di sini, model AI juga digunakan. Saat ini, terdapat banyak penelitian yang sedang dilakukan tentang bagaimana teknik pembelajaran mesin dapat digunakan untuk mengklasifikasikan citra-citra ini dan memberikan informasi kuantitatif. Salah satu penggunaannya adalah dalam klasifikasi citra onkologi yang mengidentifikasi kanker.

Teknologi pemindai seluruh slide sudah mampu mendigitalkan slide patologi pada resolusi mikroskopis, sehingga memungkinkan penerapan model AI pada gambar. Saat ini, para peneliti di Universitas Stanford sedang melatih algoritma untuk mengidentifikasi kanker kulit, salah satu jenis kanker paling umum pada manusia. Untuk melakukannya, mereka memberi label "hampir 130.000 gambar lesi kulit yang mewakili lebih dari 2.000 penyakit untuk menguji apakah komputer dapat membedakan tahi lalat yang tidak berbahaya dari melanoma dan karsinoma ganas." Hasilnya patut dicatat: algoritma mereka berkinerja sebaik panel yang terdiri dari 21 dokter kulit bersertifikat. Para peneliti berencana untuk menyediakan alat ini kepada pengguna ponsel pintar di masa mendatang.

Untuk banyak kasus penyakit lainnya, gambar belum dikumpulkan dan diberi label agar dapat digunakan sebagai data pelatihan, karena waktu yang dibutuhkan dokter untuk memberi label serta hambatan yang tak terelakkan dalam industri yang sangat teregulasi dan terfragmentasi. Bahkan setelah pelabelan selesai, dokter akan membutuhkan waktu untuk dapat menggunakan solusi AI ini mengingat keamanan dan kemanjuran harus dibuktikan kepada regulator sebelum dapat diimplementasikan.

Salah satu bidang penting dalam perawatan kesehatan saat ini adalah menentukan faktor risiko, diagnosis, dan pengobatan demensia. Dalam populasi di mana hidup lebih lama menjadi norma, demensia merupakan masalah yang semakin serius: saat ini merupakan

penyebab utama kecacatan dan ketergantungan pada lansia di seluruh dunia. Sebuah studi terbaru yang dilakukan oleh para ilmuwan di Universitas McGill menggunakan AI untuk mengembangkan algoritma yang mampu mengenali tanda-tanda demensia dua tahun penuh sebelum onsetnya menggunakan satu pemindaian PET amiloid pada otak pasien yang berisiko terkena penyakit Alzheimer. Algoritma ini diharapkan dapat meningkatkan cara dokter menangani pasien serta mempercepat dan menghemat biaya uji klinis, sehingga memungkinkan pengobatan dipasarkan lebih cepat.

Studi demensia lain dilakukan pada tahun 2018 oleh para peneliti di Universitas Boston, yang memanfaatkan AI pada data dari Studi Jantung Framingham yang terkenal. Para peneliti menggunakan AI untuk menemukan faktor risiko yang meningkatkan kemungkinan terkena demensia. Tidak mengherankan, usia merupakan faktor risiko yang signifikan untuk penyakit tersebut. Namun, para penulis menemukan hubungan bermakna lainnya dalam data tersebut. Profesor neurobiologi Rhoda Au menulis, "Analisis tersebut juga mengidentifikasi status perkawinan 'janda', BMI yang lebih rendah, dan kurang tidur di usia paruh baya sebagai faktor risiko demensia." Harapannya adalah bahwa hasilnya akan bermanfaat bagi individu maupun dokter.

Penggunaan AI dan big data lainnya adalah upaya Google untuk memprediksi penyakit jantung dan tekanan darah tinggi. Alasan pengujian kedua penanda tersebut adalah untuk dapat memprediksi sesuatu yang bahkan lebih parah: apakah seorang pasien akan menderita serangan jantung atau stroke. Untuk melakukan ini, tim Google memeriksa retina subjek, karena peneliti medis sebelumnya telah mencatat "beberapa korelasi antara pembuluh retina dan risiko episode kardiovaskular mayor." Dengan menggunakan gambar retina, Google mengatakan dapat secara akurat memprediksi pasien mana yang akan mengalami serangan jantung atau kejadian penting lainnya dalam waktu lima tahun, dan itu dilakukan secara akurat 70% dari waktu. Hasil ini sejalan dengan tes yang mengukur kadar kolesterol darah pasien.

Algoritme Google dilatih menggunakan data dari 284.335 pasien dan diuji pada dua set data independen yang masing-masing terdiri dari 12.026 dan 999 pasien. Kepala peneliti, Dr. Lily Peng, melaporkan bahwa set data pelatihan memang kecil. Namun, ia menambahkan, "Kami pikir akurasi prediksi ini akan sedikit meningkat seiring dengan semakin lengkapnya data yang kami peroleh."

AI juga saat ini digunakan untuk melacak dan memahami gejala penyakit Alzheimer. Penelitian untuk merawat pasien Alzheimer berdasarkan akumulasi data tentang perilaku mereka saat ini sedang dilakukan oleh para peneliti di Laboratorium Ilmu Komputer dan Kecerdasan Buatan (CSAIL) MIT. Untuk mengumpulkan data ini, tim telah merekatkan kotak putih datar dengan Velcro ke dinding kamar pasien yang dapat menentukan gerakan pasien, mulai dari yang terbesar hingga yang paling halus. Mesin ini melacak ribuan gerakan setiap hari menggunakan sinyal nirkabel berdaya rendah untuk memetakan hal-hal seperti kecepatan berjalan, pola tidur, lokasi, dan pola pernapasan.

Setelah informasi ini diunggah ke cloud, algoritma pembelajaran mesin (ML) menemukan pola dalam pergerakan harian pasien, membantu para peneliti mendiagnosis sekaligus melacak perjalanan penyakit. Perubahan kecil pada otak dapat bermanifestasi

sebagai perubahan halus dalam perilaku dan pola tidur bertahun-tahun sebelum gejala Alzheimer muncul. Hal ini memungkinkan para peneliti menggunakan model AI untuk mengidentifikasi pasien yang berisiko mengembangkan bentuk penyakit yang paling parah, yang berarti perawatan lebih dini, mengidentifikasi pasien yang dapat memperoleh manfaat dari terapi eksperimental, dan bahkan membantu anggota keluarga merencanakan kebutuhan perawatan di masa mendatang.

#### **6.4 PERSIAPAN UNTUK PERAWATAN PALIATIF**

Kemampuan menentukan kapan seorang pasien akan meninggal adalah hal yang sangat berharga untuk diketahui, tetapi sangat sulit bagi dokter untuk memutuskannya. Itulah sebabnya AI saat ini digunakan untuk memprediksi akhir hidup seseorang. Sebuah studi yang dilakukan pada tahun 2016 oleh tim di Departemen Penelitian Perawatan Paliatif Marie Curie di University College, London, menunjukkan bahwa tingkat kelangsungan hidup yang diprediksi oleh dokter menunjukkan variasi kesalahan yang luas, mulai dari "perkiraan yang kurang dari 86 hari hingga perkiraan yang lebih dari 93 hari." Mengapa para dokter begitu sering salah? Memprediksi mortalitas berarti memperhitungkan berbagai faktor kompleks, termasuk riwayat keluarga, usia, sifat penyakit itu sendiri, dan reaksi pasien terhadap berbagai obat. Bias dokter seperti keengganan alami mereka untuk mengakui bahwa akhir sudah dekat, semakin membingungkan situasi. Dokter dan penulis peraih penghargaan Pulitzer, Siddhartha Mukherjee, menggambarkan situasi tersebut sebagai berikut: "Dokter memiliki rekam jejak yang sangat buruk dalam memprediksi pasien mana yang akan meninggal. Kematian adalah kotak hitam utama kita."

Prediksi waktu kematian yang akurat dapat memberikan perbedaan yang luar biasa bagi pasien dan fasilitas medis. Perawatan paliatif dapat diberikan pada waktu yang tepat sehingga pasien dapat memanfaatkan berbagai layanan yang tersedia saat mereka membutuhkannya, termasuk manajemen nyeri yang tepat, dukungan psikologis, dan dukungan untuk kebutuhan sosial, budaya, dan spiritual. Prediksi yang baik juga berarti menghindari beban yang tidak perlu pada sistem perawatan kesehatan itu sendiri, yang timbul ketika pasien memasuki perawatan terlalu dini.

Bisakah AI digunakan untuk menentukan secara tepat kapan seorang pasien harus dirawat di rumah sakit? Terhubung dengan keluarga dan teman? Apakah rejimen obat telah berubah? Itulah yang ingin diketahui oleh tim peneliti dari Universitas Stanford. Dengan menggunakan algoritma AI untuk memprediksi mortalitas pasien, tim tersebut berharap dapat meningkatkan waktu perawatan akhir hayat bagi pasien yang sakit kritis.

Tim mencapai tujuannya dengan menggunakan rekam medis rumah sakit (data yang ada sebelum kematian pasien) untuk mengetahui kapan waktu terbaik untuk dapat membuat prediksi akurat tentang kapan pasien tersebut akan meninggal. Dokter di rumah sakit telah mengkodekan informasi ini, dan mencakup semuanya, mulai dari diagnosis pasien hingga jumlah pemindaian yang dilakukan dan resep medis yang ditulis. Meskipun terbatas (misalnya, tidak ada kuesioner atau percakapan), informasi tersebut dimasukkan ke dalam jaringan saraf dalam untuk melihat apakah data tersebut dapat memungkinkan algoritma menghasilkan skor

probabilitas bahwa pasien tertentu akan meninggal dalam waktu tiga hingga 12 bulan. Studi ini menggunakan data pelatihan dari sekitar 175.000 pasien dan kemudian diuji pada 45.000 pasien lainnya. Selain rekam medis rumah sakit, tim menambahkan data tambahan dari pasien yang sudah ada sebelum waktu tersebut untuk melihat apakah data tersebut dapat membantu memprediksi perkiraan tanggal kematian, serta jenis input apa yang akan mengajarkan algoritma untuk membuat prediksi tersebut. Sistem ini terbukti cukup akurat untuk memprediksi kemungkinan kematian dengan tepat pada 90% kasus historis. Model ini sekarang digunakan untuk memberikan laporan harian kepada dokter tentang pasien baru yang dirawat di rumah sakit dengan kemungkinan kematian yang tinggi, untuk intervensi atau perencanaan yang tepat.

Dalam studi lain, yang dilakukan oleh Universitas Yale, tim peneliti yang menggunakan AI menganalisis data kesehatan dari registri lebih dari 40.000 pasien untuk mencoba meningkatkan prediksi kelangsungan hidup pasien gagal jantung. Studi ini memungkinkan tim tidak hanya untuk memprediksi kelangsungan hidup tetapi juga untuk meningkatkan pilihan pengobatan dengan mengelompokkan pasien ke dalam empat kelompok, yang masing-masing memiliki respons berbeda terhadap obat-obatan yang umum digunakan. Pendekatan inovatif ini dapat menghasilkan perawatan yang lebih baik untuk kondisi kronis yang tidak dapat disembuhkan ini.

Karena masalah kotak hitam, saat ini mustahil untuk sepenuhnya memahami mengapa AI pandai dalam prediksi ini, sehingga sulit bagi dokter untuk belajar darinya. Namun, sistem AI kemungkinan akan terus bekerja lebih baik daripada dokter dalam beberapa tugas. Model AI jenis ini dapat membantu dokter meningkatkan perawatan pasien, sekaligus mengurangi tekanan yang memakan waktu bagi tenaga medis sehingga mereka dapat memanfaatkan keahlian mereka yang mengesankan.

## **6.5 PERAWATAN DI RUMAH SAKIT**

Pada tanggal 24 Januari 2018, Google menerbitkan sebuah makalah penelitian yang ditulis oleh 34 peneliti AI-nya yang mengklaim telah memprediksi hasil rawat inap pasien dengan akurasi yang lebih baik dibandingkan dengan perangkat lunak yang ada. Area yang diperiksa meliputi apakah seorang pasien akan dipulangkan dan kemudian dirawat kembali; apa diagnosis akhirnya; dan apakah ia akan meninggal di rumah sakit. Dengan melakukan hal tersebut, Google berhasil mencapai sesuatu yang sangat signifikan: untuk melakukan penelitian ini, Google berhasil memperoleh data anonim dari 216.221 orang dewasa, dengan lebih dari 46 miliar titik data di antara mereka.

Google menggunakan tiga jaringan saraf untuk belajar dari data dan mencari tahu data mana yang paling berguna untuk memprediksi hasil pasien, mengidentifikasi kata-kata dan peristiwa yang paling dekat dengan tiga hasil. Algoritme AI menggunakan proyek Google sebelumnya yang disebut Vizier dan belajar untuk mengabaikan data yang tidak relevan daripada mengandalkan ilmuwan AI untuk memilih sendiri data yang akan dimasukkan dalam analisis. Meskipun hasilnya belum dievaluasi oleh rekan sejawat, Google menunjukkan bahwa

jaringan sarafnya, antara lain, mampu memprediksi kematian pasien satu hingga dua hari sebelum metode yang tersedia saat ini.

AI juga dapat membantu menentukan kapan pasien dapat meninggalkan rumah sakit. Sebuah tim di Rumah Sakit Anak Los Angeles (CHLA), sekarang menggunakan pembelajaran mendalam di Unit Perawatan Intensif Pediatrik untuk memprediksi keadaan yang dapat diterima saat keluar. Mereka melakukannya dengan data yang diperoleh dari pertemuan unit perawatan intensif (ICU), yang masing-masing berisi banyak informasi tentang perkembangan penyakit, perawatan, dan hasil pasien. "AI memungkinkan kami mengeksplorasi informasi yang tersedia dari pertemuan di ICU dan menghasilkan model yang memahami hubungan kompleks antara obat, intervensi, dan kesejahteraan pasien," kata David Ledbetter, ilmuwan data senior di CHLA. "Intinya, kami menggunakan AI untuk belajar dari pengalaman tersebut dan menerapkan apa yang telah dipelajari untuk memberikan perawatan terbaik bagi setiap anak yang datang ke klinik kami."

Algoritma pembelajaran mesin membutuhkan data dalam jumlah yang sangat besar, dan landasan kerja dengan Epic, sebuah perusahaan rekam medis elektronik yang berbasis di Madison, Wisconsin, menyediakan hal tersebut. Pada tahun 2018, Ochsner Health System, salah satu sistem kesehatan pertama di negara ini yang mengintegrasikan AI ke dalam alur kerja perawatan pasien, memanfaatkan data yang melimpah ini untuk meluncurkan alat kecerdasan buatan baru yang menarik. Platform ML, yang dibangun oleh Epic dan didukung oleh Microsoft Azure, dimaksudkan untuk memprediksi dan dengan demikian mencegah "kode" – yaitu, saat-saat ketika seorang pasien mengalami henti jantung atau pernapasan dan membutuhkan intervensi medis segera. Ochsner menjalankan studi percontohan selama 90 hari dengan sistem tersebut pada akhir tahun 2017, dan hasilnya "mengurangi jumlah kode tipikal rumah sakit sebesar 44 persen."

Platform tersebut melacak ribuan titik data untuk memprediksi pasien mana yang akan segera memburuk. Kemudian, platform tersebut menghasilkan peringatan "prakode" yang membantu tim perawatan Ochsner merawat pasien lebih cepat dan lebih proaktif, menghemat waktu dan nyawa. Peringatan ini telah disempurnakan untuk memberikan peringatan empat jam sebelumnya. "Ini seperti alat triase," kata Michael Truxillo, direktur medis tim respons cepat dan resusitasi di Ochsner Medical Center. "Seorang dokter mungkin mengawasi 16 hingga 20 pasien di satu unit dan mengetahui siapa yang paling membutuhkan perhatian Anda selalu menjadi tantangan. Alat itu mengatakan, 'Hei, berdasarkan nilai lab, tanda-tanda vital, dan data lainnya, lihat pasien ini sekarang.'"

Rumah Sakit Royal Free di Inggris bekerja sama dengan DeepMind milik Google untuk membangun aplikasi bernama Streams guna meningkatkan efektivitas tenaga kerja. Mereka berfokus pada masalah "kegagalan untuk menyelamatkan" – yaitu, tidak mendapatkan dokter yang tepat untuk pasien yang tepat pada waktunya. Streams sejauh ini telah digunakan untuk membantu dokter mengidentifikasi dan menangani cedera ginjal akut dengan lebih baik – suatu kondisi yang terkait dengan 40.000 kematian di Inggris setiap tahun, seperempatnya diperkirakan dapat dicegah oleh NHS Inggris. Sulit untuk mengidentifikasi gejalanya, karena gejala tersebut merupakan kombinasi dari sinyal data kecil yang mengindikasikan hal tersebut,

dan titik-titik data tersebut dapat berarti banyak hal lain, sehingga sering terlewatkan. Streams menggabungkan berbagai sumber data sehingga dapat mendeteksinya. Sarah Stanley, perawat di Royal Free yang memimpin tim resusitasi, mengatakan: "Streams menghemat banyak waktu kami setiap hari. Peringatan instan tentang beberapa pasien kami yang paling rentan berarti kami dapat memberikan perawatan yang tepat kepada pasien yang tepat jauh lebih cepat."

Contoh lain peningkatan efektivitas tenaga kerja adalah Rumah Sakit Presbyterian New York, salah satu rumah sakit tersibuk di Amerika Serikat. Mereka telah mendirikan pusat komando di luar lokasi bagi perawat untuk memantau pasien. Pusat Operasi Klinis (CLOC) memantau data fisiologis yang berasal dari berbagai sumber secara waktu nyata (real-time) dari tempat tidur pintar dan teknologi berbasis sensor lainnya. Pemantauan berbantuan AI memungkinkan perawat CLOC untuk mengarahkan bagaimana perawat di lokasi sebaiknya menghabiskan waktu mereka. Hal ini membantu mengurangi kelelahan kewaspadaan bagi perawat di rumah sakit dan memungkinkan mereka untuk menghabiskan lebih banyak waktu dengan pasien. Leo Bodden, kepala bagian teknologi rumah sakit, mengatakan, "Kami telah berhasil memanfaatkan sistem otomatis yang sangat kompleks yang secara signifikan mengurangi redundansi dalam tugas yang dilakukan oleh perawat terdaftar, dokter, dan staf lainnya, mengurangi jumlah anggota tim yang secara fisik diperlukan untuk memantau pasien, dan secara signifikan mengurangi jumlah waktu yang dihabiskan staf untuk memasukkan data pasien."

Bagaimana tanggapan dokter dan pasien terhadap pengenalan AI dalam dunia kedokteran? Memang, ada beberapa penolakan. Pasien, misalnya, khawatir mesin akan menggantikan dokter mereka, atau manusia pada akhirnya akan disingkirkan dari semua keputusan perawatan kesehatan. Namun, David Ledbetter dari CHLA yakin dokter akan selalu menjadi penentu akhir. Begitu pula direktur medis Ochsner, Tuxillo. "Para dokter pada dasarnya skeptis," kata Tuxillo. "Mereka bertanya-tanya, 'Apakah kecerdasan buatan akan menggantikan dokter?' Jawabannya tidak; dokter tetap harus mendiagnosis pasien. Alat ini membantu mereka memprioritaskan perawatan dan mensintesis aliran informasi yang terus berubah yang harus dilacak oleh dokter. Teknologi ini membantu menyelamatkan nyawa. Jika kita dapat membuat perbedaan bagi pasien karena peringatan ini, kita telah melakukan pelayanan yang luar biasa bagi mereka, keluarga mereka, dan masyarakat."

Sebagaimana tes darah dan pemindai MRI tidak menggantikan dokter ketika diperkenalkan ke dunia medis, AI pun tidak akan menggantikannya. AI hanyalah alat lain yang dapat digunakan dokter untuk menyelidiki, menguji, dan menganalisis pasien mereka, sehingga memungkinkan mereka memberikan perawatan yang lebih baik.

## BAB 7

### MENGEMBANGKAN STRATEGI AI

Seperti yang dicatat Andrew Ng, banyak pemimpin bisnis yang tidak menganggap serius fakta bahwa teknologi internet merupakan salah satu peluang utama dalam bisnis mereka. Saat ini, banyak bisnis menyerahkan keputusan tentang bagaimana, atau bahkan apakah, akan menggunakan solusi AI kepada unit bisnis masing-masing. Seringkali hanya sedikit tindakan yang diambil selain mengembangkan bukti konsep. Ketika tindakan diambil, masalah besar dapat muncul ketika unit-unit ini tidak berkolaborasi dalam strategi AI, meninggalkan perusahaan dengan sejumlah implementasi AI yang tidak kompatibel atau bersaing. Hal ini tidak hanya dapat menghabiskan waktu dan uang, tetapi kurangnya standar yang dihasilkan dapat melemahkan kemampuan untuk bertransformasi secara skala besar.

Untungnya, banyak perusahaan dan jajaran direksinya kini menjadikan transformasi digital berbasis AI sebagai prioritas tinggi. Hal ini penting untuk dilakukan sedini mungkin, karena mudah untuk meremehkan lamanya waktu yang dibutuhkan untuk perubahan serta banyaknya tantangan yang ada saat menerapkan strategi AI. Hanya karena ada tren di pasar untuk menggunakan AI, bukan berarti langkah-langkah yang harus diambil perusahaan sudah jelas.

Wajar bagi para eksekutif untuk khawatir tentang kesulitan yang melekat dalam mengintegrasikan kapabilitas AI ke dalam bisnis dan budayanya, atau tentang tidak memiliki infrastruktur yang tepat untuk menuai manfaat algoritma AI. Bahkan ketika integrasi dan infrastruktur tidak dipertanyakan, ada masalah tentang kasus penggunaan dan model AI mana yang paling berguna dan hemat biaya untuk diimplementasikan oleh bisnis. Ada juga pertanyaan tentang bagaimana mengumpulkan data yang tepat. Memiliki "big data" tidak membantu implementasi AI jika data tersebut tidak relevan dengan strategi AI perusahaan. Tanpa mengetahui aplikasi AI mana yang dapat digunakan bisnis, kebutuhan data tetap tidak jelas.

Terlepas dari semua kesulitan ini, satu hal yang jelas: organisasi yang secara tepat meningkatkan skala penggunaan AI mereka memperoleh keuntungan besar dibandingkan pesaing dan bisnis yang tidak kesulitan bersaing. Bab ini membahas cara mengembangkan strategi AI yang dapat diimplementasikan untuk perusahaan, dengan fokus pada mengatasi tantangan yang muncul ketika tidak ada strategi yang diterapkan.

#### 7.1 TUJUAN SISTEM KECERDASAN TERHUBUNG

Salah satu cara untuk memahami bagaimana AI dapat menguntungkan bisnis adalah dengan mengelompokkan penggunaannya ke dalam tiga kategori tingkat tinggi: menghilangkan tugas berulang yang dilakukan manusia, menghasilkan wawasan dan prediksi, dan memperkuat kecerdasan manusia.

**Menghilangkan tugas berulang:** Mengotomatiskan tugas-tugas sederhana dan berulang terkadang disebut sebagai "mengeluarkan robot dari manusia," dan komputer telah

berhasil melakukannya dengan baik. Hasilnya, teknologi otomatisasi proses robotik (RPA) dapat membebaskan karyawan dari pekerjaan robotik yang hanya membutuhkan sedikit upaya kognitif. Otomatisasi mengurangi biaya, mengurangi kesalahan manusia, dan mengurangi ketidakpuasan karyawan yang disebabkan oleh pekerjaan yang tidak memerlukan banyak pikiran pada pekerja pengetahuan. Sistem RPA, yang telah mendapatkan daya tarik sejak 2009, menciptakan lapisan orkestrasi: program yang menghubungkan proses bisnis yang sudah dilakukan oleh komputer tanpa harus menulis ulang semua perangkat lunak atau memerlukan perubahan substansial apa pun pada aplikasi yang ada. Tidak seperti sistem pembelajaran mesin, sistem RPA ini bersifat ahli, yaitu berbasis aturan. Artinya, sistem RPA dapat "diajarkan" penekanan tombol yang digunakan manusia untuk menjalankan tugas mereka serta logika di balik apa yang dilakukan manusia. Pengaturan ini memungkinkan bisnis untuk menggantikan manusia yang melakukan tugas-tugas yang membosankan dan berulang (bukan pekerjaan) dengan sistem RPA otomatis. RPA saja tidak selalu cukup untuk menyelesaikan suatu tugas, karena masih ada keputusan operasional harian sederhana yang harus dibuat yang saat ini tidak dapat ditangani oleh aturan logis RPA. Untuk memungkinkan RPA membuat keputusan yang lebih cerdas atau lebih kompleks, RPA dikombinasikan dengan sistem AI yang lebih cerdas, sehingga membangun fondasi bagi tenaga kerja digital dan otomatisasi robotik yang cerdas. Jenis solusi pengambilan keputusan yang cerdas ini memiliki keuntungan karena tidak memerlukan perubahan apa pun pada aplikasi yang mendasarinya.

**Menghasilkan wawasan dan prediksi:** Sistem pembelajaran mesin dan penalaran semantik memberi kita kemampuan untuk secara otomatis mengekstrak wawasan yang sebelumnya belum ditemukan dari data terstruktur dan tidak terstruktur, memungkinkan kita untuk mengidentifikasi pola dalam data dan membuat prediksi atau menemukan hubungan antara dan di antara peristiwa. Dari sini, komputer dapat merekomendasikan dan bahkan mungkin menindaklanjuti rekomendasi ini. Pemanfaatan data ini meningkatkan keunggulan kompetitif perusahaan, memungkinkannya untuk melakukan hal-hal seperti memperoleh pelanggan baru, yang mengarah pada pendapatan baru, dan mengurangi atrisi pelanggan sehingga meningkatkan margin dan memperbaiki rantai pasokan.

**Memperkuat kecerdasan manusia:** Sistem yang memberikan rekomendasi kontekstual bagi karyawan atau pelanggan memanfaatkan penilaian dan kreativitas manusia, bekerja sama dengan kekuatan analitis AI. Kita sudah menggunakan aplikasi untuk mencapai kecerdasan tertambah semacam ini dengan cara yang terlokalisasi, misalnya, di mana sistem memberikan saran atau pengetahuan dalam konteks proses bisnis atau alur kerja pengguna (seperti membaca gambar radiologi) dan membuat tugas berbasis pengetahuan yang saat ini dilakukan oleh karyawan menjadi lebih produktif. Manfaat dari peningkatan kecerdasan manusia adalah bahwa hal itu mentransisikan bisnis menuju suatu bentuk kewaskitaan bisnis: ketika suatu situasi terjadi, para manajer dan eksekutif telah memperkirakan bahwa hal itu mungkin terjadi dengan beberapa probabilitas dan siap untuk menanganinya dengan tepat.

## 7.2 TANTANGAN IMPLEMENTASI AI

Sebagaimana telah dijelaskan sebelumnya, tentu ada tantangan dalam upaya implementasi inisiatif AI. Para pemimpin bisnis seringkali tidak dapat dengan mudah memutuskan secara tepat bagaimana dan di mana AI akan digunakan dalam organisasi mereka. Terkadang, kasus penggunaan yang mereka butuhkan untuk menciptakan nilai bagi perusahaan mereka tidak terlihat jelas. Terlalu sering, perusahaan terjebak dalam tahap pembuktian konsep (POC). Terkadang, studi mereka menunjukkan potensi nilai implementasi kasus penggunaan AI, tetapi manajemen khawatir untuk menerapkan AI ke tahap produksi karena khawatir AI akan "menjadi liar" dan membuat keputusan buruk yang dapat merugikan bisnis dari perspektif regulasi, komersial, atau merek. Di lain waktu, terdapat tantangan kolaboratif, di mana tim AI bekerja secara terpisah dan tidak memiliki komitmen untuk mengintegrasikan modelnya ke dalam sistem dan alur bisnis yang ada.

Perusahaan lain melaporkan bahwa setelah menugaskan proyek, butuh waktu terlalu lama untuk menyiapkan data dan lingkungan yang tepat. Seringkali, data berada dalam sistem yang berbeda, dan manajer tidak memiliki anggaran atau keterampilan (atau keduanya) untuk menyiapkan lingkungan komputasi berkinerja tinggi yang sering kali diperlukan untuk AI. Selain itu, dalam hal profitabilitas, perusahaan yang telah berhasil mengeksekusi proyek AI telah melaporkan hasil yang beragam. Menurut survei tahun 2016, peluang rata-rata untuk menghasilkan keuntungan dari implementasi strategi big data hanya 27%. Enrique Dans berkata, "Ternyata ada kelemahan fatal dalam pendekatan sebagian besar perusahaan terhadap pembelajaran mesin, alat analisis masa depan: 87% proyek tidak melewati fase eksperimental dan karenanya tidak pernah berhasil diproduksi." Beberapa masalah yang menghalangi kesuksesan dijelaskan lebih rinci dalam subbagian berikut. Setiap strategi AI yang sukses perlu mengatasi tantangan-tantangan ini.

**Hype versus kenyataan:** Dalam AI, jurang pemisah antara hype dan kenyataan cukup besar. Istilah *kecerdasan buatan* telah menjadi semacam frasa yang umum, dan tanpa pemahaman mendalam tentang berbagai jenis kecerdasan buatan dan cara kerjanya, sulit untuk membedakan mana yang nyata dan mana yang tidak. Ketidaktepatan ini dapat menyulitkan penerapan strategi yang sukses. Perusahaan mungkin langsung terjun tanpa mengetahui jenis AI apa yang harus digunakan dan kapan harus menggunakannya. Selain itu, jika orang-orang yang membuat keputusan tentang AI dalam suatu organisasi merasa tidak yakin tentang kemampuan, nilai, atau ROI-nya yang sebenarnya, adopsi AI dalam bisnis mereka mungkin terbatas atau bahkan tidak ada sama sekali.

**Data buruk:** Bahkan setelah memulai proyek, banyak perusahaan melaporkan masalah dengan data "kotor". Masalah ini sering muncul sebagai data yang hilang, atau nilai yang salah, atau hubungan yang hilang di antara entitas data. Tim tidak hanya mengetahui bahwa data mereka tidak cukup bersih, mereka mungkin menemukan bahwa mereka tidak memiliki data yang mereka butuhkan untuk bergerak maju. Ada juga tantangan data duplikat, yang membuatnya sulit untuk memiliki gambaran yang akurat tentang, misalnya, pelanggan – menghabiskan waktu, uang, dan terkadang, reputasi. Jika pembelajaran mesin yang diawasi menggunakan terlalu banyak data duplikat, prediksi mungkin terlalu bergantung pada fitur-

fitur yang paling umum dalam duplikat. Data yang buruk ini, serta berbagai faktor lainnya, dapat menyebabkan masalah yang lebih signifikan. Misalnya, setelah sebuah bisnis menugaskan proyek AI, mungkin butuh waktu terlalu lama untuk menyiapkan data dan membersihkan serta mengintegrasikannya dengan lingkungan pemodelan AI. Penundaan ini dapat mengakibatkan proyek AI membutuhkan waktu terlalu lama untuk diluncurkan dan menunjukkan nilai, yang menyebabkan banyak proyek terhenti di tengah jalan. Terlalu banyak ide bagus yang terbengkalai pada tahap ini.

**Pola penggunaan:** Mungkin ada banyak area fungsional di perusahaan yang membutuhkan penggunaan kecerdasan buatan, tetapi setiap area mungkin memiliki kebutuhan yang berbeda. Apa yang diinginkan SDM dari kecerdasan buatan, yang mungkin berupa prediksi tentang seberapa sukses calon karyawan baru nantinya, bisa sangat berbeda dari apa yang dianggap berharga oleh tenaga penjualan. Di sisi lain, mungkin ada area yang tampaknya memiliki kebutuhan yang sangat berbeda tetapi sebenarnya dapat dilayani oleh algoritma yang sama atau serupa. Misalnya, memprediksi henti jantung pada pasien dari data monitor jantung historis dan memprediksi kebocoran pada pipa transportasi gas adalah kasus penggunaan yang sangat berbeda, tetapi keduanya mungkin menggunakan jenis model dan struktur data yang sama. Kurangnya pemahaman tentang pola semacam ini dapat menyebabkan perusahaan menggunakan solusi yang berbeda dan terputus-putus. Tanpa memahami kesamaan di antara berbagai metode AI dan potensi standarisasi serta pemanfaatan silang, manajer mungkin kesulitan mengembangkan rencana yang komprehensif. Hasilnya seringkali berupa upaya duplikasi dan biaya yang lebih tinggi.

**Kompleksitas dan teknologi yang sedang berkembang:** Pemodelan matematika dan manajemen data tidak dapat diakses oleh semua orang. Banyak model dan algoritma dalam AI yang pada dasarnya kompleks. Penyempurnaan model untuk variabel keputusan yang relevan membutuhkan keterampilan, waktu, dan kesabaran. Yang dapat memperburuk situasi adalah teknologi AI yang terus berubah, menghadapi tantangan baru dan merintis jalan baru. Saat ini tidak hanya beragam pendekatan dan alat yang tersedia, tetapi juga terus bermunculan. Proliferasi ini dapat menyulitkan untuk mengetahui mana yang harus digunakan dan mana yang sudah usang. Kompleksitas AI dan pengenalan teknologi baru yang terus-menerus dapat menyulitkan perusahaan untuk secara proaktif menerapkan strategi AI yang berkelanjutan, sehingga memaksa banyak perusahaan untuk menerapkannya dalam silo organisasi atau menunda solusi AI sepenuhnya.

**Regulasi yang terus berkembang:** Regulasi yang ada dan yang akan datang mulai mengubah lanskap pemasaran digital. Pada tahun 2018, Peraturan Perlindungan Data Umum (GDPR) mulai berlaku. Undang-undang ini merupakan hasil dari hampir empat tahun kerja yang dimulai pada Januari 2012, ketika gagasan reformasi perlindungan data di seluruh Uni Eropa diperkenalkan. Pada tahun 2020, Undang-Undang Privasi Konsumen California (CCPA) menjadi undang-undang. Kedua rangkaian peraturan tersebut bertujuan untuk memastikan bahwa perlindungan data dan privasi ditegakkan bagi warga negaranya masing-masing. Dalam kasus GDPR, undang-undang ini melindungi warga negara Uni Eropa (UE) dan Wilayah Ekonomi Eropa (EEA) dan memengaruhi perusahaan yang melakukan bisnis di wilayah tersebut atau

dengan warga negara di wilayah tersebut. Dalam kasus CCPA, undang-undang ini melindungi penduduk California. Individu yang dilindungi berhak untuk mengetahui data apa yang dikumpulkan tentang mereka dan apakah dan kapan data tersebut dijual atau ditransfer. Mereka juga dapat mengakses data mereka, mewajibkan bisnis untuk menghapus informasi pribadi tentang mereka, dan melarang penjualan data pribadi mereka. Peraturan ini, serta peraturan-peraturan mendatang, tidak hanya akan memengaruhi mereka yang dilindungi tetapi juga bisnis di seluruh dunia, dan perusahaan-perusahaan sudah berupaya untuk mencapai transparansi yang lebih besar dan kontrol yang lebih ketat atas data pribadi. Kami membahas peraturan dan kebijakan yang muncul lebih lanjut di Bab 14.

**Bakat langka:** Bakat AI saat ini sedang langka. Meskipun perkiraannya bervariasi, pada Desember 2017, raksasa internet Tiongkok Tencent Holdings Ltd. memperkirakan bahwa dunia memiliki lowongan pekerjaan untuk sekitar 200.000 hingga 300.000 praktisi dan peneliti AI. Secara anekdot, statistik ini tercermin dalam gaji tinggi yang diterima oleh ilmuwan AI dan waktu rekrutmen yang lama. Mengingat banyaknya pilihan yang dimiliki kandidat potensial, mereka bisa pilih-pilih, dan mereka sering kali lebih suka bekerja di tempat di mana mereka dapat menjadi bagian dari tim yang kuat dan memiliki beragam masalah untuk dikerjakan; di mana data bersih tersedia dengan mudah; dan di mana platform yang kuat untuk pemodelan sudah ada. Tanpa faktor-faktor ini, perusahaan merasa bahwa merekrut bakat papan atas sangat sulit.

**Eksekusi yang tidak selaras:** Bahkan ketika sebuah perusahaan telah berhasil merekrut anggota tim AI yang tepat, sebuah tim dapat membangun model AI yang memecahkan masalah yang tidak dianggap prioritas oleh bisnis, tidak tahu cara menggunakannya, atau tidak dipercaya. Hal ini dapat terjadi karena berbagai alasan. Departemen TI – bagian penting dari penskalaan solusi AI apa pun – mungkin tidak mendukung pengembangan dan penerapannya atau mungkin mengalokasikan dana yang tidak memadai. Tim infrastruktur mungkin tidak memiliki cukup landasan bisnis untuk meminta penggunaan server atau infrastruktur yang diperlukan untuk mendukungnya. Tata kelola data mungkin telah membatasi penggunaan data yang diperlukan karena peraturan yang membingungkan, atau rekayasa data mungkin telah menunda pembangunan jalur data yang memadai karena proses yang rumit. Atau hal ini dapat terjadi karena pengguna bisnis tidak cukup terlibat dalam definisi model dan proses pengembangan. Yang lebih mendasar lagi, tim sains AI mungkin tidak memahami bisnis dengan cukup baik untuk menciptakan algoritma yang tepat atau tidak dapat mengartikulasikan solusi dengan cara yang menarik minat pengguna. Manajemen sendiri mungkin juga lalai mengartikulasikan kebutuhan dan tujuan spesifik, menyerahkan semuanya kepada tim AI, atau menolak untuk memberikan pengaruh yang cukup untuk memaksa pengguna dan TI mendukung tim. Sebuah bisnis juga mungkin gagal menerapkan solusi yang baik, menundanya ke masa depan yang tidak pasti, jika solusi ini terlalu banyak mengubah proses bisnis atau membutuhkan investasi tambahan.

### 7.3 KOMPONEN STRATEGI AI

Akar penyebab banyak masalah implementasi AI adalah belum dikembangkannya strategi dan peta jalan AI yang komprehensif, kohesif, dan berkelanjutan. Ketika para eksekutif mulai berpikir untuk memanfaatkan AI di perusahaan mereka, mereka harus memutuskan apa yang ingin mereka dapatkan darinya. Memikirkan hasil-hasil berikut dapat membantu membentuk strategi tersebut.

Pertama, bisnis harus menetapkan tujuan yang mendefinisikan tujuan dan visi AI dalam perusahaan. Tujuan tersebut dapat mencakup keunggulan kompetitif yang berkelanjutan, peluang pendapatan tambahan, atau pengurangan biaya. Seringkali tujuannya adalah transformasi AI holistik yang dapat mendukung transformasi bisnis digital dan inovasi bisnis melalui produk, layanan, atau model bisnis baru.

Selanjutnya, bisnis harus mengidentifikasi *kasus penggunaan* yang mendefinisikan potensi cara jangka pendek dan jangka panjang untuk menggunakan AI guna mendorong tujuannya. Kasus penggunaan membantu menentukan bagaimana data perusahaan dimonetisasi. Katalog kasus penggunaan ini memungkinkan bisnis untuk mencapai berbagai tujuan, memungkinkannya memahami keseluruhan kasus bisnis untuk AI dan menunjukkan berbagai cara AI dapat memberikan nilai. Mengidentifikasi kasus penggunaan membantu menentukan persyaratan holistik untuk jenis solusi AI yang diinginkan serta manajemen informasi terkait untuk mendukung solusi tersebut. Hal ini juga membantu bisnis memahami kelompok pengguna dan proses bisnis yang mungkin terdampak. Informasi ini diperlukan untuk membuat peta jalan dan rencana manajemen perubahan.

Bisnis kemudian harus membuat arsitektur yang mencakup komponen teknologi dan platform yang akan digunakan perusahaan untuk mendukung kebutuhan AI dan datanya. Arsitektur ini tidak hanya melibatkan komponen AI, tetapi juga bagaimana menjadikan arsitektur perusahaan dan aplikasi yang ada berpusat pada AI. Arsitektur ini harus mencakup pengumpulan, penyimpanan, dan pemrosesan informasi, pemodelan AI, visualisasi, pengalaman pengguna, manajemen dan penerapan model, serta integrasi dengan sistem lini bisnis atau proses bisnis untuk aktivasi.

Perusahaan juga memerlukan *strategi data* dan rencana kesiapan data. Harus ada cara untuk menyediakan data yang relevan untuk kasus penggunaan, dan data tersebut harus memiliki kualitas dan ketepatan waktu yang memadai. Kesiapan data dapat mencakup aktivitas seperti penilaian dan remediasi kualitas data serta tata kelola siklus hidup data.

Bisnis harus membangun *kapabilitas organisasi*, termasuk struktur, bakat, dan proses yang diperlukan untuk menjalankan proyek AI dalam skala besar. Kapabilitas organisasi mencakup keputusan seputar struktur; misalnya, apakah AI harus berada dalam satu kelompok, tertanam dalam kelompok bisnis yang relevan, atau memelihara komunitas praktik. Ini juga harus mencakup peran dan proses untuk mempertahankan siklus hidup AI, keahlian yang diperlukan untuk memelihara dan mengembangkannya, dan cara menjaga koneksi dengan ekosistem AI eksternal.

Terakhir, perusahaan harus membangun sistem *tata kelola dan manajemen perubahan* yang mendefinisikan bagaimana keputusan dibuat, bagaimana "keamanan" model AI

dipertahankan, dan bagaimana perusahaan menerapkan dan memanfaatkan kapabilitas baru ini di seluruh perusahaan. Struktur tata kelola menentukan hal-hal seperti pemilik proyek, standar, pengukuran nilai, serta persetujuan dan prioritas proyek. Tata kelola juga mendefinisikan agenda penelitian dalam hal analisis pasar, penilaian kompetitif, dan penilaian vendor. Manajemen perubahan berupaya untuk melibatkan pimpinan dan perusahaan dalam perjalanan AI guna meningkatkan keterlibatan aktif dan mengurangi resistensi dalam organisasi.

#### **7.4 LANGKAH-LANGKAH UNTUK MENGEMBANGKAN STRATEGI AI**

Setelah hasil yang diharapkan ditentukan, saatnya untuk mengambil langkah-langkah penting menuju perluasan dan penguatan visi yang akan mewujudkan rencana AI organisasi. Pertama, anggota tim eksekutif harus disegarkan, sehingga mereka secara kolektif memiliki tingkat pemahaman dan apresiasi yang sama terhadap AI. Mereka perlu memahami apa yang esensial dan dapat dicapai serta membangun konsensus tentang area fokus dan tujuan program AI. Memiliki ekspektasi yang tidak realistis dapat menenggelamkan upaya AI yang bertujuan terbaik. Briefing eksekutif harus mencakup investigasi kredibilitas AI: tinjauan kritis antara sensasi dan kenyataan, yang didasari oleh bagaimana AI telah menguntungkan perusahaan di berbagai industri. Briefing tersebut harus memberikan visi yang jelas kepada setiap anggota tim tentang seperti apa keberhasilan penggunaan AI.

Setelah briefing, merupakan waktu yang tepat untuk memikirkan pembentukan tim AI, karena setiap langkah selanjutnya membutuhkan upaya yang lebih besar daripada sebelumnya. Pembentukan tim dapat dilakukan dengan awalnya mengumpulkan orang-orang yang memiliki pemikiran yang sama dari dalam organisasi, kemudian melengkapi kelompok ini dengan kolaborator eksternal dan karyawan baru. Keputusan yang harus diambil oleh tim ini mencakup kepemilikan atas aktivitas seperti mendorong penemuan kasus penggunaan dan mendefinisikan kebutuhan arsitektur sistem AI.

Selanjutnya, bisnis harus mengidentifikasi area peluang yang spesifik. Penyelarasan perlu dicapai pada pendorong penting untuk melibatkan (atau meningkatkan) teknologi AI di perusahaan. Di mana teknologi ini akan bekerja paling baik? Apakah sepadan dengan investasi waktu dan uang? Mengidentifikasi peluang harus mencakup pemeriksaan kasus penggunaan potensial bagi perusahaan dan memahami bagaimana AI di area tersebut dapat meningkatkan bisnis secara keseluruhan. Untuk setiap area, tentukan tujuan bisnis, imbal hasil yang diantisipasi, dampak pada proses dan sistem saat ini, serta ketercapaian tujuan. Prioritaskan area peluang tersebut. Tentukan kasus bisnis mana yang paling menarik, berdampak, dan layak bagi perusahaan.

Setelah ini, bisnis harus menentukan kesiapannya untuk mengimplementasikan AI di area peluang yang telah diidentifikasi sebelumnya. Penilaian harus mencakup data, teknologi, struktur organisasi, tata kelola, merek, serta rantai pasokan dan mitra. Para pemimpin bisnis perlu memahami seberapa siap mereka untuk memanfaatkan teknologi AI di berbagai dimensi teknis dan nonteknis. Mereka harus mencari tahu kesenjangan antara kondisi saat ini dan kemampuan yang dibutuhkan untuk sistem AI berdasarkan penilaian ini dan pemahaman

tentang perusahaan sejenis yang melakukan transisi serupa, meskipun mereka berasal dari industri yang berbeda. Memahami kesenjangan tersebut membantu bisnis mengidentifikasi area yang perlu ditingkatkan dan menentukan jalur transformasi.

Setelah semua langkah ini dilakukan, bisnis dapat mengembangkan peta jalan AI yang strategis dan operasional untuk membantu organisasi bergerak menuju penggunaan AI dalam skala besar. Perusahaan dapat menentukan aktivitas yang diperlukan beserta pendekatan, prioritas, rencana tingkat tinggi, tonggak pencapaian, dan jadwal berdasarkan profil risiko, urgensi, tingkat kesiapan, serta pertimbangan anggaran dan jadwal. Peta jalan ini harus mencakup setiap kesenjangan dalam penilaian kesiapan yang dibahas sebelumnya dan memastikan bahwa setiap kekurangan mencakup rencana perbaikan.

Rencana pada tingkat ini tidak memungkinkan penilaian terperinci atas dampak keuangan terhadap bisnis, tetapi bisnis dapat mulai memetakan apa yang mereka butuhkan untuk membuat kasus bisnis tingkat tinggi. Kasus bisnis tersebut mencakup penilaian kualitatif atau skala besar atas biaya dan manfaat. Potensi biaya bagi organisasi dapat berupa kebutuhan untuk memprioritaskan atau mengalokasikan kembali sumber daya, serta melakukan investasi yang diperlukan dalam hal staf, pelatihan, perangkat lunak, perangkat keras, cloud/hosting, konsultasi, dan operasional. Jika memungkinkan, trade-off dan manfaat serta biaya terkait dari peluang tersebut harus diidentifikasi. Manfaat potensial bagi perusahaan dari penggunaan kecerdasan buatan dapat mencakup peningkatan efisiensi operasional, pengurangan biaya, perolehan pendapatan, peningkatan waktu pemasaran, nilai IP, keunggulan kompetitif, kepuasan pelanggan, dan peningkatan pangsa pasar.

Pada titik ini, sudah saatnya untuk mulai membiasakan orang-orang di dalam organisasi dengan AI. Dengan memanfaatkan perangkat yang sama seperti yang sebelumnya digunakan oleh tim eksekutif, organisasi yang lebih luas dapat diajarkan untuk memahami apa yang esensial dan dapat dicapai. Para pemimpin harus merencanakan roadshow edukasi di seluruh perusahaan untuk mempercepat proses dan menggerakkan perusahaan menuju budaya yang berbasis AI dan data.

## **7.5 DIPERLUKAN BEBERAPA PERAKITAN**

Sisa buku ini membahas secara lebih rinci beberapa topik yang diperlukan untuk mendefinisikan strategi AI dan peta jalan transformasi. Ada beberapa elemen fundamental yang penting untuk memastikan bahwa perusahaan AI baru dapat ditingkatkan skalanya, memberikan nilai transformatif sekaligus sepadan dengan waktu, upaya, dan uang yang dibutuhkan untuk mencapainya. Ini termasuk strategi data transformatif, platform AI, dan pusat keunggulan AI untuk mempertahankan program.

### **Membangun Pusat Keunggulan AI**

Cara paling efektif untuk meningkatkan skala AI di seluruh perusahaan adalah dengan mendirikan pusat keunggulan AI (AI CoE). AI CoE adalah organisasi di perusahaan yang menjalankan rencana AI. Ini membantu mendefinisikan dan memelihara strategi untuk platform, data, dan talenta di seluruh unit bisnis. Ini juga merupakan tempat di mana keputusan tentang tata kelola dan manajemen platform AI dibuat.

Ada tiga model eksekusi yang berbeda untuk menentukan posisi tim AI (lihat Gambar 11.1 di Bab 11). Sebagian besar perusahaan kemungkinan akan memutuskan dan beralih ke model terfederasi. Dalam model terfederasi, talenta dan ilmuwan AI tidak semuanya berada di dalam CoE. Sebaliknya, karyawan tersebut bekerja di dalam unit bisnis (BU) masing-masing, dengan CoE mendukung mereka sesuai kebutuhan. Pengaturan ini memungkinkan BU tertentu untuk berinovasi secara lokal jika diperlukan, menghasilkan kasus penggunaan khusus mereka sekaligus memanfaatkan CoE untuk berkolaborasi lintas BU jika diperlukan. Jika suatu model terlalu tersentralisasi (ditunjukkan di sebelah kanan Gambar 11.1), maka kasus penggunaan lokal tidak mendapatkan perhatian yang semestinya. Jika suatu model terlalu terdesentralisasi (ditunjukkan di sebelah kiri Gambar 11.1), maka mungkin terdapat terlalu sedikit kerja sama antar BU, yang memaksa setiap unit untuk menetapkan strateginya sendiri, yang mengakibatkan redundansi dan inefisiensi. Akibatnya, segala sesuatu yang masuk ke dalam produksi aktual membutuhkan biaya yang relatif lebih tinggi, memberikan imbal hasil yang lebih rendah, dan membutuhkan waktu lebih lama untuk mencapai kasus bisnis.

Membangun dan memelihara hubungan dengan rekan sejawat, universitas, lembaga penelitian, dan vendor produk sangat penting dalam ekosistem AI. Hubungan ini memungkinkan CoE untuk mengikuti perkembangan yang dilakukan pihak lain di industri ini dan membuka jalan bagi pemanfaatan riset dan rekrutmen akademis. Hubungan ini sangat penting karena bidang ini berkembang begitu cepat. Mitra teknologi tepat guna yang meningkatkan ekosistem juga penting. Mitra-mitra ini dapat berdampak langsung pada kecepatan dan skala adopsi solusi AI dalam organisasi. Selain itu, mitra teknologi yang tepat dapat membantu bisnis memahami ke mana arah industri ini. Lanskap AI berkembang pesat; penting untuk selalu mengikuti perkembangan industri guna memilih pemasok, vendor, dan kolaborator terbaik bagi perusahaan. Fungsi AI CoE dan interaksinya dengan bagian lain perusahaan dibahas lebih lengkap di Bab 11.

### **Membangun Platform AI**

Platform AI adalah kerangka kerja perangkat keras, perangkat lunak, dan alat yang memungkinkan bisnis untuk mempercepat siklus hidup penuh proyek AI perusahaan dalam skala besar. Ketika perusahaan membuat proyek AI tanpa platform yang tangguh, mereka membuang-buang waktu dan uang sebelum mendapatkan hasil yang signifikan. Bagian terburuknya bukanlah proyek tersebut membutuhkan waktu lebih lama dan memiliki pengembalian investasi yang lebih rendah; melainkan orang-orang kehilangan semangat dan meninggalkan proyek. Angka-angka tidak berfungsi, sehingga proyek tersebut dihentikan, yang pasti akan menyebabkan pemborosan uang dan penurunan kredibilitas untuk proyek-proyek mendatang. Memiliki platform dan proses terkaitnya memungkinkan siklus yang lebih cepat dari eksperimen hingga produksi dan mengotomatiskan sebagian besar pekerjaan kasar bagi pengguna platform.

Ada beberapa hal penting yang harus disediakan oleh platform AI. Sangat penting untuk mendasarkan desain sistem AI pada persyaratan dari seluruh bisnis, baik saat ini maupun yang diantisipasi di masa mendatang. Desain tersebut harus mendukung siklus hidup proyek AI dan pembelajaran mesin, mulai dari manajemen data hingga eksperimen model hingga

memindahkan model kerja ke produksi dan memantaunya secara berkelanjutan. Selain itu, desain tersebut harus menciptakan layanan mandiri bagi ilmuwan AI tanpa bergantung pada TI: data dan proses harus dapat diakses oleh mereka. Lebih lanjut, desain tersebut harus mencakup luasnya alat, kerangka kerja, dan metode AI dan pembelajaran mesin yang kemungkinan akan digunakan oleh ilmuwan AI sehingga menunggu infrastruktur atau pengadaan tidak menghambat kemajuan mereka. Platform AI dibahas lebih lengkap di Bab 9 dan 12.

### **Mendefinisikan Strategi Data**

Di masa lalu, data dianggap hanya sebagai produk sampingan dari prosedur bisnis. Meskipun fungsi seperti layanan pelanggan dan pelaporan keuangan mungkin perlu mengakses data untuk ditindaklanjuti, data tersebut memiliki nilai yang kecil dan seringkali tidak mudah diakses setelah penggunaan awalnya. Namun, dalam 20 tahun terakhir, banyak hal telah berubah secara dramatis. Hingga saat ini, para ahli di bidangnya menanamkan pengetahuan mereka dalam logika yang mereka bangun, menciptakan perangkat lunak untuk sebagian besar aplikasi pendukung keputusan. Dengan pembelajaran mesin, komputer memperoleh logika dari data tanpa campur tangan manusia secara eksplisit. Ini berarti data telah menjadi sangat penting dalam organisasi.

Saat ini, semua orang sepakat bahwa data adalah aset yang sangat berharga bagi suatu organisasi, fundamental tidak hanya untuk pengambilan keputusan, tetapi juga untuk menciptakan atau menyempurnakan produk dan layanan. Faktanya, sebagian besar bisnis terkemuka saat ini bergerak menuju strategi berbasis data dan algoritmik. Namun, meskipun data mungkin merupakan sumber wawasan yang vital, segelintir perusahaan besar mengendalikannya dalam jumlah besar, terutama data pribadi. Monopoli yang nyaris terjadi ini tidak terjadi pada algoritma; tidak ada yang memiliki matematika AI. Kerangka kerja sumber terbuka kini lebih baik dari sebelumnya, dan banyak di antaranya mendukung beragam inisiatif AI. Meskipun paket pembelajaran mesin berkualitas tinggi juga semakin menjamur, data berkualitas tinggi masih langka. Namun, terlepas dari ini, hanya sedikit perusahaan yang benar-benar menyesuaikan cara mereka menangkap, menyimpan, dan mengelola aset data mereka. Strategi data – seperangkat aturan, kebijakan, dan standar terkait pengelolaan data di seluruh perusahaan – sangat penting untuk memastikan nilai maksimum dapat diekstraksi dari data tersebut.

Strategi data modern memungkinkan perusahaan mengembangkan keseimbangan antara *kontrol* yang lebih ketat dan *fleksibilitas* yang lebih besar. Dengan meningkatnya kekhawatiran akan regulasi dan reputasi, banyak perusahaan menyusun strategi data seputar standarisasi dan kepatuhan. Kontrol dan standarisasi data secara historis diperlukan untuk kepatuhan dan pelaporan regulasi dan terus diharapkan. Namun kini, fleksibilitas dan berbagi data jauh lebih penting untuk mendorong keputusan bisnis dan pengembangan produk serta layanan baru. Meskipun sebagian besar perusahaan memahami pendekatan untuk pengendalian, standar, dan kualitas data melalui tata kelola data, banyak yang baru mulai memikirkan fleksibilitas, berbagi, penggunaan kembali, dan monetisasi yang diperlukan untuk mengubah data menjadi aset strategis.

Sebagai bagian dari strategi data mereka, perusahaan perlu memutuskan strategi *monetisasi data* dan strategi *komersialisasi data* mereka. Monetisasi data menggunakan data internal perusahaan untuk menghasilkan nilai, termasuk konsumsi data pihak ketiga. Monetisasi dilakukan melalui penerapan algoritma untuk kasus penggunaan yang disebutkan sebelumnya, dengan menggunakan data internal perusahaan. Komersialisasi data menciptakan produk dan layanan untuk perusahaan lain berdasarkan data yang dihasilkan atau dikumpulkan di dalam perusahaan. Ini menjadi aliran pendapatan baru. Semakin banyak perusahaan yang memanfaatkan komersialisasi data sebagai bagian dari strategi data mereka.

Data harus berkualitas tinggi dan selengkap mungkin untuk menciptakan keunggulan kompetitif berbasis data. Bisnis harus melihat data dari perspektif menciptakan kembar digital bisnis – pelanggan, produk, proses, dan lingkungan bisnis. Jika data berasal dari perusahaan itu sendiri (data pihak pertama), data tersebut harus diintegrasikan dan dibersihkan. Penentuan juga harus dibuat tentang data tambahan apa yang diperlukan untuk pemodelan dan di mana mendapatkannya. Terkadang, bisnis dapat mengumpulkan data tambahan ini tanpa intervensi dari luar. Misalnya, bank yang menggunakan profil sosiodemografi nasabah untuk membuat rekomendasi penawaran dapat meningkatkan respons secara signifikan dengan menggunakan data perilaku yang sudah dimilikinya tetapi tidak digunakan dalam proses penawaran. Data ini dikumpulkan saat nasabah berinteraksi dengan situs web, aplikasi, konten, dan medianya.

Mengutip kutipan yang dikaitkan dengan Bill Joy, salah satu pendiri Sun Microsystems, tidak peduli seberapa banyak data bermanfaat yang Anda miliki di perusahaan, ada lebih banyak data bermanfaat di luar perusahaan Anda. Ini dalam bentuk data pihak kedua dan ketiga. Data pihak kedua adalah apa yang disetujui bisnis untuk diperoleh dari mitra, pemasok, dan vendor mereka. Sebuah perusahaan mungkin, misalnya, meminta mitra untuk data penjualan yang lebih terperinci untuk memungkinkan perkiraan permintaan mereka yang lebih baik. Data pihak ketiga adalah data yang dapat dibeli dari sumber independen. Data ini mencakup hal-hal seperti data geolokasi atau citra satelit untuk memungkinkan pemahaman tentang, misalnya, lalu lintas pejalan kaki atau penggunaan tempat parkir di toko ritel. Data yang diperoleh dari masing-masing sumber ini meningkatkan kemampuan perusahaan untuk membuat model di area baru dan dengan akurasi yang lebih baik. Dengan cara ini dan lainnya, data dan pembelajaran mesin dapat digunakan untuk meningkatkan pemahaman tentang bisnis, meningkatkan dan mempertahankan keunggulan kompetitif. Gambar 7.1 menunjukkan berbagai jenis data pihak kedua yang dikumpulkan oleh banyak perusahaan. Mereka melakukan beberapa hal ini secara langsung dengan mengikis situs web. Selain itu, ratusan perusahaan telah muncul yang menghasilkan atau mengumpulkan informasi semacam ini pada tingkat kualitas yang mudah dibeli dan terintegrasi dengan lebih mulus.

Untuk memanfaatkan semua data ini dengan sukses, tata kelola perlu dibangun. Tata kelola data memastikan bahwa data berkualitas tinggi tersedia di seluruh perusahaan saat dibutuhkan. Apakah data tersebut dapat digunakan? Apakah akurat? Siapa yang bertanggung jawab dan siapa yang diizinkan untuk melihatnya? Apakah aman? Tata kelola data, keamanan,

dan privasi merupakan topik relevan yang telah dibahas secara ekstensif di buku-buku lain, sehingga kami tidak membahasnya secara mendalam di sini.

## 7.6 MELANGKAH MAJU

Setelah tujuan ditetapkan, dukungan eksekutif diperoleh, bakat serta data dikumpulkan dari seluruh organisasi dan eksternal, karyawan terlibat, dan Pusat Kewirausahaan AI sedang dibangun, saatnya untuk melihat lebih dekat bagaimana segala sesuatunya akan berjalan. Di bab berikutnya, kami membahas alur kerja untuk mengimplementasikan proyek AI dalam suatu organisasi. Memahami alur kerja ini akan membantu menjelaskan sistem, kebijakan manajemen risiko, dan struktur organisasi apa yang paling dapat mendukung siklus hidup implementasi AI.

Firmografi	Demografis	teknografi	Hubungan	Satelit	Berita dan Blog	Perangkat	Perilaku
Pendapatan	Biografi	infrastruktur TI	Jejaring sosial	Lalu lintas pejalan kaki	Situs berita	Alamat IP terbalik	Konsumsi konten
Karyawan	Judul/peran	Aplikasi (digunakan atau diinstal)	Hubungan pelaporan	Penggunaan tempat parkir	Laporan analisis dan broker	Geolokasi	Riwayat pencarian
Kode SIC/NAICS	Waktu dalam judul/peran	Kontrak layanan profesional	Kontak sosial eksternal	Inventaris dan pemindahan	Blog	Penggunaan	Kunjungan situs, halaman, dan aplikasi
Lokasi kantor	Lokasi kantor	Peralatan	Kontak kolega dan karyawan internal				Postingan sosial
Departemen dan unit bisnis	Departemen	Teknologi lainnya	Peringkat pengaruh (Klout atau skor kepemilikan)				
Siaran pers	Pendidikan	Tanggal perpanjangan kontrak					
Postingan pekerjaan							

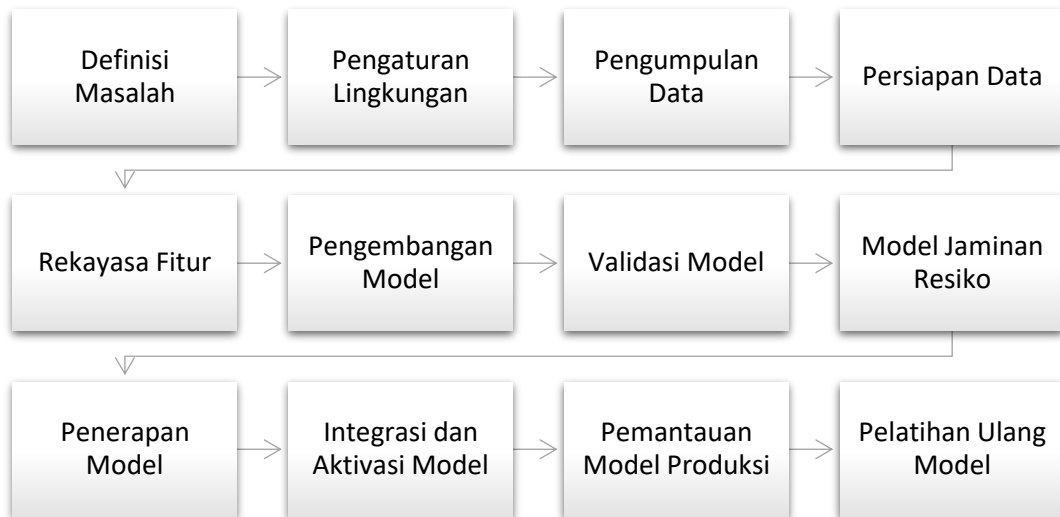
**Gambar 7.1** Berbagai jenis data pihak ketiga yang tersedia secara komersial atau melalui web.

## **BAB 8**

### **SIKLUS HIDUP AI**

Bab ini menyajikan alur kerja pembuatan model AI tingkat tinggi: deskripsi langkah-langkah yang diperlukan untuk menjalankan proyek AI dalam skala besar, mulai dari menemukan kasus penggunaan hingga menerapkan dan menggunakan model dalam produksi. Beberapa langkah bersifat iteratif dan mungkin tumpang tindih, tetapi secara umum, langkah-langkah tersebut berlangsung sesuai urutan yang dijelaskan. Memahami perbedaan antara berbagai solusi pembelajaran mesin dapat menjadi tantangan bagi mereka yang bukan ahli AI, dan mungkin ada banyak algoritma dan beragam cara yang dapat digunakan untuk memenuhi kebutuhan bisnis. Ini tidak berarti Anda harus menjadi ahli dalam setiap nuansa pemodelan untuk mengelola tim atau proyek AI dengan sukses. Namun, penting untuk memahami tidak hanya langkah-langkah dalam pembuatan setiap model, tetapi juga mengapa langkah-langkah tersebut diperlukan, yang memungkinkan manajer tim untuk membedakan antara apa yang berhasil dan apa yang tidak. Selain itu, memahami siklus hidup ini memperjelas mengapa fitur-fitur platform AI yang dijelaskan di bab berikutnya diperlukan.

Gambar 8.1 menunjukkan proses end-to-end tingkat tinggi untuk pemodelan. Langkah pertama dalam proses ini adalah mengidentifikasi dan mendefinisikan kasus penggunaan. Untuk setiap kasus penggunaan, tim memutuskan pertanyaan spesifik apa yang ingin dijawab dan menentukan bagaimana menjawab pertanyaan-pertanyaan ini bermanfaat bagi bisnis. Sangat penting untuk mengambil keputusan sejak awal tentang cara meringkas kasus penggunaan dan mendefinisikan manfaat kualitatifnya, metrik target, dan kebutuhan data. Langkah selanjutnya adalah mengumpulkan dan menilai data. Data apa yang tersedia? Berapa banyak? Seberapa baik data tersebut? Di mana data tersebut saat ini disimpan? Seberapa mudah diakses? Apakah ada sinyal dalam data? Apakah ada kebutuhan untuk memperoleh data tambahan dari sumber internal atau eksternal, dan seberapa mudah mendapatkannya? Tim kemudian memperbaiki masalah apa pun dengan data, membersihkannya, dan mempersiapkannya untuk membuat model pembelajaran dasar.



**Gambar 8.1** alur kerja untuk proyek AI, pembelajaran mesin, dan ilmu data.

Tim menciptakan model atau beberapa model terbaik untuk kebutuhannya, melatih algoritma AI yang paling sesuai untuk kasus penggunaan, data yang tersedia, dan pertanyaan yang ingin dijawab oleh bisnis. Hasilnya bisa berupa satu model atau beberapa model yang dapat digunakan bersama. Akhirnya, tim menguji dan mempersiapkan penerapan model tersebut untuk digunakan dalam produksi. Mari kita bahas langkah-langkah ini lebih detail.

### 8.1 MENDEFINISIKAN KASUS PENGGUNAAN

Mengidentifikasi kasus penggunaan umumnya dimulai dengan memahami para aktor dalam suatu organisasi. Aktor seringkali adalah karyawan, tetapi bisa juga pelanggan, mitra, atau pemasok. Misalnya, di perusahaan ritel, karyawan adalah orang-orang yang bekerja di toko atau di bagian keuangan; pelanggan adalah orang-orang yang datang ke toko atau daring untuk membeli; pemasok adalah orang-orang yang membuat atau mengirimkan produk ke toko untuk dijual; dan mitra dapat berupa orang-orang yang mengambil produk di toko untuk diantar (seperti Instacart). Semua individu ini dianggap sebagai aktor. Membuat peta aktor – sebuah dokumen tingkat tinggi yang menjelaskan siapa para aktor dan apa fungsi atau tujuan utama mereka – merupakan titik awal yang baik.

Selanjutnya, tim mengamati proses bisnis yang digunakan setiap tim aktor untuk mencapai fungsi atau tujuan mereka. Sifat setiap tugas diperiksa dan aktivitas dikelompokkan ke dalam kategori, seperti apa yang dilakukan aktor setiap hari atau secara teratur versus hal-hal yang mereka lakukan secara sporadis; pekerjaan yang tidak memerlukan keputusan yang rumit versus pekerjaan yang memerlukan pertimbangan yang matang; dan tindakan yang merupakan bagian dari proses kolaboratif versus hal-hal yang dilakukan masing-masing aktor secara independen. Apakah informasi atau pengetahuan tambahan dapat meningkatkan fungsi atau tujuan, atau meningkatkan sebagian dari proses, atau bahkan menghilangkannya sama sekali? Apakah termasuk memprediksi sesuatu? Apakah memprediksi sesuatu akan membantu aktor melakukan suatu tugas? Apa yang perlu diantisipasi? Apakah tujuan prediksi yang ada terdefinisi dan terukur dengan baik? Misalnya, jika perusahaan telekomunikasi dapat

memprediksi pelanggan mana yang paling mungkin membatalkan layanan telepon mereka, mereka akan dapat memfokuskan tindakan pencegahan mereka hanya pada kelompok pelanggan tersebut.

Cara terbaik untuk mengeksplorasi praktik bisnis adalah dengan mengajukan pertanyaan kunci kepada para aktor yang terlibat. Ini mungkin termasuk pertanyaan seperti:

- Apa yang Anda harap tidak perlu Anda lakukan karena sifatnya yang berulang dan tidak menarik?
- Sesuatu telah terjadi. Apa yang Anda harap Anda ketahui tentang hal itu yang sulit Anda teliti?
- Apa yang Anda harap dapat Anda prediksi – dengan kata lain, apa yang mungkin terjadi meskipun belum terjadi?

Jawaban atas pertanyaan-pertanyaan ini dan pertanyaan lainnya membantu menentukan jenis tugas AI yang paling mungkin memenuhi kebutuhan para aktor. Contohnya meliputi ekstraksi informasi dari teks tak terstruktur menggunakan pemrosesan bahasa alami (NLP), prediksi menggunakan pembelajaran mesin atau pembelajaran mendalam, otomatisasi tindakan menggunakan RPA, atau kombinasi ketiganya.

Setelah melalui proses ini, tim telah menentukan daftar luas kasus penggunaan potensial (lihat Gambar 8.2 untuk contoh "ruang keputusan"): apa tujuan kasus penggunaan, apa tugas pemodelannya, data apa yang dibutuhkan, proses bisnis atau fungsi apa yang sesuai dengannya, aktor mana yang terdampak, dan sebagainya. Langkah selanjutnya adalah memprioritaskan hal-hal ini. Informasi tambahan diperlukan untuk setiap kasus penggunaan potensial agar dapat dicantumkan dan ditindaklanjuti dengan tepat.

Perlu dicatat bahwa ketika sebuah tim bersiap untuk memulai sebuah proyek, seringkali tidak jelas apakah kasus penggunaan tertentu dapat diimplementasikan. AI disebut sains karena tim membuat hipotesis dan kemudian mencoba memvalidasi atau membatalkan hipotesis tersebut. Mantra yang sering terdengar di sirkuit konferensi "gagal cepat" sebenarnya tentang seberapa cepat dan dengan sedikit usaha sebuah tim dapat menentukan apakah sebuah "hipotesis" tidak akan berhasil. Semakin banyak eksperimen yang dapat dilakukan tim, semakin cepat mereka menemukan model yang berguna, dan semakin tinggi nilai yang mereka sadari. Inilah yang dirujuk Jeff Bezos dalam kutipan di awal bab ini.

Kasus penggunaan yang dipilih harus, pertama-tama, relevan dengan kesuksesan organisasi. Untuk menentukan hal ini, kami biasanya melihat dua atribut – kompleksitas dan nilai. Hal ini dapat dilakukan dalam dua tahap: pertama, tahap yang lebih detail untuk memahami prioritas relatif, dan kemudian, untuk rangkaian kasus penggunaan yang dipilih, tahap yang lebih detail pada nilai dan biaya bisnis, termasuk kasus bisnis dan ROI. Aspek-aspek detailnya akan kami bahas nanti.

Sangat penting untuk memahami nilai relatif dari setiap kasus penggunaan. Untuk memahami nilainya, tim melihat bagaimana tugas AI (misalnya, otomatisasi, ekstraksi pengetahuan, atau prediksi) dapat meningkatkan hasil bisnis. Misalnya, jika AI digunakan untuk membuat prediksi, tetapi prediksi tersebut tidak dapat ditindaklanjuti – jika tidak ada cara untuk menggunakan informasi tersebut guna memengaruhi proses bisnis – maka tidak

ada nilai yang dapat dihasilkan darinya. Jika tugas AI dapat ditindaklanjuti, maka tim akan mempertimbangkan dampaknya: misalnya, jika itu adalah otomatisasi, apakah kita akan mengotomatiskan pekerjaan untuk lima orang atau 300 orang, dan manajemen perubahan apa yang diperlukan untuk menerapkannya?



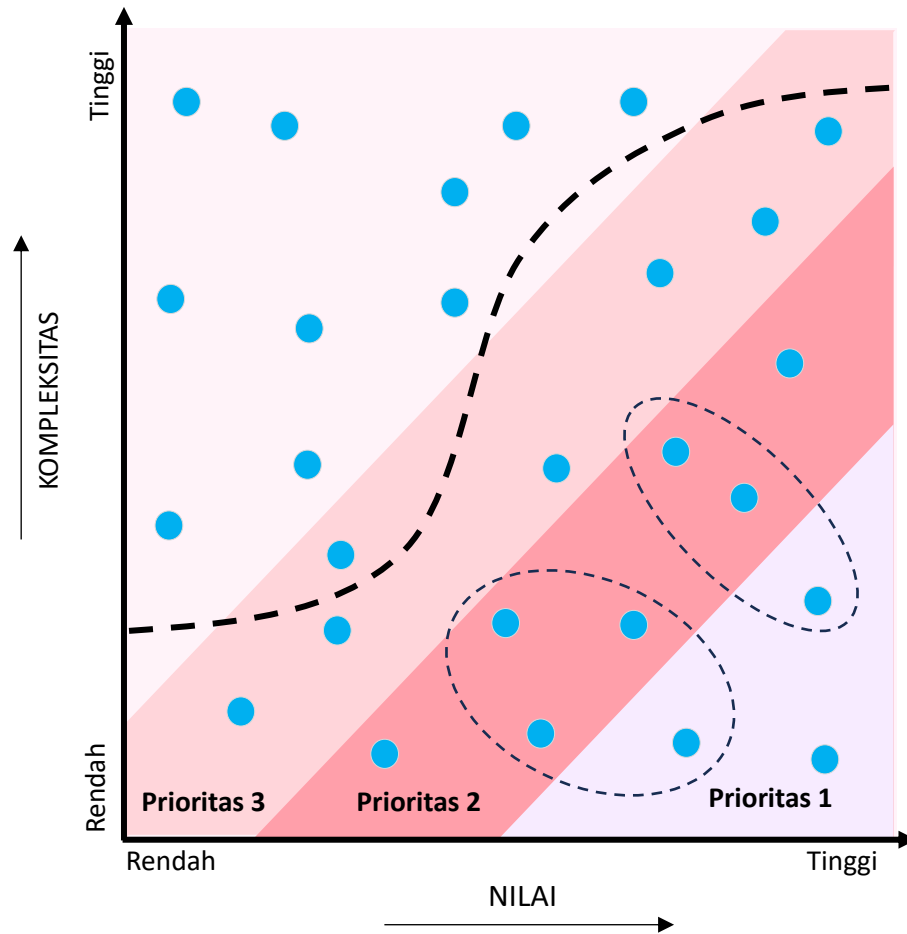
**Gambar 8.2** Contoh peta tujuan kasus penggunaan, tugas pemodelan untuk mendukung keputusan, dan data yang diperlukan untuk mendukung pemodelan.

Untuk memahami kompleksitas suatu kasus penggunaan, tim mengevaluasi jenis AI yang dibutuhkan untuk mengimplementasikannya, misalnya pembelajaran mesin, otomatisasi pemrosesan robotik (RPA), atau yang lainnya. Selain itu, untuk tugas pembelajaran mesin atau pembelajaran mendalam, tim melihat apakah mereka dapat memperoleh data yang diperlukan untuk setiap kasus penggunaan. Kuantitas dan kualitas data ini sangat penting. Jika informasi tidak akurat, penggunaannya menghasilkan keluaran yang tidak dapat diandalkan. Ada beberapa potensi kompleksitas lain yang perlu dipertimbangkan juga, seperti seberapa besar perubahan organisasi yang diperlukan untuk setiap kasus penggunaan. Apakah pengujian dan penyetelan kinerja dapat dieksekusi? Apakah pembelajaran atau aktivasi harus dilakukan secara real-time, atau dapat dilakukan secara batch? Apakah keluaran akan dihasilkan sebagai, misalnya, laporan atau spreadsheet untuk pengguna atau diberikan kepada program perangkat lunak lain melalui panggilan API? Berdasarkan pertimbangan ini, tim

membuat grafik nilai relatif dan kompleksitas (atau biaya) kasus penggunaan untuk mengevaluasi mana yang merupakan prioritas yang lebih tinggi.

Dimensi lain yang perlu dipertimbangkan adalah kasus penggunaan mana yang akan dikelompokkan bersama. Misalnya, sebuah perusahaan mungkin mendasarkan keputusannya pada pemanfaatan jenis algoritma atau kumpulan data yang sama. Sekalipun manajer tidak mengetahui jawaban atas pertanyaan-pertanyaan ini secara tepat, penting untuk memikirkannya, setidaknya secara relatif, agar dapat memprioritaskan kasus penggunaan. Selama proses penentuan prioritas, tim sering kali mengelola pertimbangan yang mempertimbangkan, misalnya, apakah akan memulai kasus penggunaan baru dengan data baru atau apakah akan mengiterasi model dengan algoritma yang lebih canggih tetapi dengan data yang sudah ada. Pilihan terakhir mungkin bernilai lebih rendah tetapi dapat diimplementasikan lebih cepat daripada model baru. Gambar 8.3 menunjukkan contoh distribusi kasus penggunaan berdasarkan nilai dan kompleksitas. Gambar tersebut menunjukkan pengelompokan kasus penggunaan yang mencakup prioritas berdasarkan jenis data atau jenis model. Kurva-s menunjukkan batas ROI di mana kasus penggunaan tidak layak diimplementasikan.

Satu hal lagi yang perlu diingat oleh tim adalah bahwa saat mengembangkan kasus penggunaan pertama dan mempelajari lebih lanjut tentang data melalui iterasi, banyak ide lain untuk kasus penggunaan akan ditemukan. Sangat penting untuk menambahkannya kembali ke dalam katalog kasus penggunaan di backlog. Kasus penggunaan ini membantu menciptakan nilai tambahan; hal ini juga menjadi salah satu alasan mengapa membangun kasus bisnis apriori terkadang bisa menantang. Tim sering kali menemukan sekitar empat kasus penggunaan baru untuk setiap kasus yang mereka implementasikan.



**Gambar 8.3** Grafik yang menunjukkan kasus penggunaan berdasarkan nilai dan kompleksitas.

## 8.2 MENGUMPULKAN, MENILAI, DAN MEMULIHKAN DATA

Mengumpulkan, menilai, dan memulihkan data merupakan salah satu langkah krusial menuju implementasi kasus penggunaan AI yang sukses. Sangat penting untuk mengumpulkan sebanyak mungkin data relevan tentang pelanggan, produk, transaksi, dan proses organisasi. Seringkali terdapat volume data yang sangat besar yang telah dikumpulkan, tetapi apakah data tersebut cukup relevan dengan apa yang ingin dicapai bisnis? Kumpulan data yang tepat merupakan salah satu aset perusahaan yang paling berharga. Kumpulan data yang bermanfaat inilah yang memungkinkan kemajuan terbaru dalam pembelajaran mesin. Di sinilah sinyal ditemukan yang akan dicari dan dimanfaatkan oleh algoritma.

### Instrumentasi Data

Sebagian besar perusahaan telah mencatat transaksi mereka sehingga mereka dapat menjalankan bisnis, menagih pelanggan, dan menerima pembayaran. Transaksi biasanya mengacu pada aktivitas seperti jual beli, atau ketika seseorang online untuk membayar tagihan. Selain mencatat transaksi ini, banyak perusahaan menyimpan informasi pencatatan. Pencatatan adalah cara mereka melacak aktivitas atau interaksi, baik antar sistem maupun antara pengguna dan sistem. Contoh interaksi yang dicatat adalah saat pelanggan mengunjungi situs web perusahaan atau menghubungi layanan pelanggannya. Data yang

dihasilkan pengguna ini dapat membantu bisnis memahami, misalnya, apakah pelanggan daring mengisi keranjang belanja mereka tetapi meninggalkannya dan tidak melakukan pembayaran. Cara lain bagi perusahaan untuk menginstrumentasikan lebih banyak data adalah melalui sensor. Sensor merupakan sumber data yang berkembang pesat. Banyak perusahaan menggunakan sensor pada truk mereka, misalnya, untuk mengukur posisi truk-truk ini dalam armada mereka.

Untuk setiap kasus penggunaan yang dipertimbangkan, tim melihat data yang telah dikumpulkan dan bagaimana data tersebut dikumpulkan. Jika aplikasi menghadap pengguna, anggota tim mencari tahu apakah sistem mencatat semua interaksi pengguna yang relevan. Ada baiknya juga untuk menentukan seberapa mudah dan bermanfaat untuk mencatat interaksi tambahan yang mungkin berguna tetapi belum diinstrumentasikan. *Instrumentasi data* mengacu pada pengambilan dan pengukuran data tentang proses yang terjadi. Semakin banyak instrumentasi yang dilakukan, semakin banyak data yang ada. Dalam contoh belanja daring sebelumnya, instrumentasi tambahan mungkin digunakan untuk menangkap dinamika penekanan tombol dan pergerakan tetikus dan panel sentuh (yang harus diinstrumentasikan di sisi klien atau dalam aplikasi).

Dengan menggunakan data transaksional, model AI memberikan rekomendasi kepada pengguna. Dengan menggunakan data tambahan tentang alur klik pengguna, apa yang sedang dijelajahnya, apa yang ditambahkan ke keranjang belanja, atau apa yang dihapusnya, model AI "memahami" mengapa pelanggan mungkin meninggalkan keranjang belanjanya – misalnya, untuk menjelajahi produk lain atau karena ia kehilangan minat. Ini membantu menyempurnakan pesan yang tepat untuk membuatnya kembali, alih-alih mencoba kampanye penargetan ulang generik. Dengan data tambahan tentang penekanan tombol dan gerakan tetikus pelanggan, model AI mendapatkan pemahaman awal tentang apakah dan mengapa pelanggan tertentu cenderung meninggalkan keranjangnya sehingga intervensi yang tepat dapat dicoba. Dengan setiap lapisan data tambahan yang diinstrumentasikan dan digunakan, berbagai jenis prediksi dan respons menjadi mungkin, dan konversi keseluruhan terus meningkat di setiap langkah.

Sebagaimana dibahas dalam bab-bab sebelumnya, tidak semua data perlu dikumpulkan secara internal. Data pihak ketiga yang berasal dari luar organisasi dapat dibeli. Salah satu contoh data pihak ketiga yang sering digunakan adalah tayangan iklan digital, yang digunakan dalam penargetan dan personalisasi pelanggan. Contoh lain termasuk citra satelit tempat parkir untuk memahami tren pelanggan, informasi geolokasi, dan tren persaingan dari berita. Selama beberapa tahun terakhir, sejumlah besar perusahaan telah bermunculan yang menawarkan berbagai jenis data alternatif. Saat mengevaluasi kasus penggunaan AI, tim tidak hanya melihat kumpulan data yang ada, tetapi juga menilai kumpulan data lain apa yang harus diinstrumentasikan atau diperoleh dari pihak ketiga untuk memenuhi kebutuhan mereka sebaik mungkin.

### **Pembersihan Data**

Data dalam bentuk mentah jarang siap digunakan dalam pemodelan. Data perlu dipersiapkan dengan cermat untuk memastikan tidak ada yang salah atau hilang. Jika ada yang

salah atau hilang, seseorang perlu memperbaiki data tersebut melalui pembersihan data, salah satu bagian terpenting dan kurang dihargai dari pemodelan berbasis data. Pembersihan data sangat penting untuk memulai model AI dengan benar. Ini adalah cara untuk menemukan dan memperbaiki masalah data seperti data buruk yang dikumpulkan karena sensor yang tidak andal, data yang tidak tersedia karena perubahan perangkat lunak yang merusak mekanisme pencatatan, atau data yang disalahartikan. Data yang hilang dapat mengakibatkan prediksi yang lemah, dan memutuskan cara menangani data yang hilang merupakan bagian dari proses pembersihan. Model yang berbeda mungkin lebih atau kurang sensitif terhadap data yang hilang, dan bahkan model yang biasanya menangani data yang hilang dengan baik terkadang dapat sensitif terhadapnya. Ketika perlu mengisi data yang hilang, tim dapat menggunakan model untuk imputasi nilai yang hilang.

Pengorganisasian data juga penting. Sebuah perusahaan mungkin telah mengumpulkan banyak data, tetapi alur data yang mudah digunakan dan andal diperlukan saat menerapkan platform AI. Alur tersebut mencakup tempat penyimpanan data, cara mendapatkannya, dan kemudahan akses serta analisis setelah data tersedia. Tim juga harus memeriksa keandalan aliran data. Apakah sumber eksternal akurat? Apakah sensor dikalibrasi dengan benar? Ada uji statistik yang dijalankan pada data untuk memastikan keandalannya. Kemudian, tim perlu menggabungkan data dari berbagai aliran data untuk mempersiapkannya bagi penggunaan spesifik dalam pemodelan AI. Hal ini mengharuskan data diekstraksi, ditransformasi, dan dimuat (ETL). Kami membahas arsitektur untuk manajemen data di Bab 9.

Selain itu, organisasi perlu menetapkan proses tata kelola data dan pemantauan kualitas yang berkelanjutan dan teratur yang diperlukan untuk memastikan data bersih dan untuk memperbaiki setiap celah.

### **Pelabelan Data**

Langkah selanjutnya adalah mulai mempersiapkan data pelatihan. Jika sebuah organisasi menggunakan pembelajaran terawasi, tim perlu memberi label pada data yang digunakan untuk pelatihan. Label adalah apa yang coba diprediksi oleh perusahaan dari data. Sebagai contoh, mungkin ada informasi tentang transaksi pelanggan dan aktivitas lainnya seperti jumlah panggilan ke pusat panggilan, dan kolom atau label yang menunjukkan apakah pelanggan tersebut membatalkan layanannya atau tidak. Model AI belajar dari data ini dan memprediksi label untuk pelanggan baru yang datanya tidak ada dalam set data pelatihan. Label-label ini mungkin melekat pada informasi yang telah diperoleh: misalnya, apakah seseorang gagal bayar pinjaman. Jika tidak, pelabelan memerlukan intervensi ahli, sebuah proses di mana pakar bisnis atau ilmuwan AI menandai data yang tidak berlabel dengan informasi yang bermakna.

Pertimbangkan contoh deteksi otomatis kerusakan pada pipa minyak dari citra drone. Perusahaan memiliki citra historis pipa tersebut, dan citra tersebut telah ditandai secara manual untuk menentukan mana yang menunjukkan kerusakan dan mana yang tidak. Pelabelan tambahan yang diperlukan adalah menggambar batas atau mengarsir bagian yang rusak pada citra yang ada. Pelabelan yang tepat ini kemudian dapat digunakan sebagai data pelatihan.

Penerapan label pada data harus dilakukan secara akurat. Jika tidak, setiap kesalahan akan mengurangi daya prediktif model AI apa pun yang telah dilatih pada kumpulan data ini. Jika data tidak diberi label secara inheren, penting untuk memutuskan siapa yang melakukan pelabelan, bagaimana ia akan melakukannya, dan seberapa besar upaya yang diperlukan. Jenis data yang perlu diberi label menentukan jenis alat yang digunakan untuk melakukan pelabelan. Untuk data gambar – misalnya, untuk mengetahui bagian mana dari pipa industri yang terkorosi atau tidak – terdapat alat seperti Annotorious dan LabelMe. Untuk data audio, misalnya, untuk mengenali lagu apa yang sedang diputar, seseorang dapat menggunakan Praat. Untuk teks tak terstruktur atau bahasa alami, terdapat beberapa alat sumber terbuka dan komersial.

Jika data yang diberi label memiliki kebutuhan privasi data, kemungkinan besar data tersebut harus diberi label secara internal, atau mungkin oleh perusahaan yang bergerak di bidang pelabelan data. Untuk data yang kurang privat, seperti gambar publik untuk melatih pengenalan gambar, platform crowdsourcing seperti Amazon Mechanical Turk dapat digunakan. Alih daya seringkali lebih murah dan lebih cepat daripada pelabelan internal, tetapi harus diperhatikan agar label data yang dihasilkan tidak berkualitas rendah.

Pelabelan otomatis juga dapat digunakan. Salah satu pendekatannya adalah dengan menetapkan label berdasarkan aturan bisnis secara terprogram. Metode lain adalah dengan menggunakan pembelajaran semisupervised, dimulai dengan hanya beberapa set yang diberi label. Terkadang juga dimungkinkan untuk menghasilkan data sintetis, misalnya, menggunakan Jaringan Adversarial Generatif (GAN) dalam pembelajaran mendalam. Dalam pendekatan ini, algoritma pembelajaran mendalam menghasilkan data baru agar selaras dengan statistik set data asli. Pelabelan otomatis dan data sintetis lebih murah dan biasanya memiliki masalah privasi yang lebih rendah, tetapi seringkali kurang akurat dibandingkan pendekatan lain yang dijelaskan sebelumnya.

### **8.3 REKAYASA ALAM**

Dengan data bersih dan berlabel yang tersedia, salah satu cara bagi ilmuwan AI untuk mulai memahami data sebagai persiapan untuk pemodelan AI adalah dengan melakukan *analisis data eksploratori* (EDA), seperti mencari korelasi dan anomali dalam data. Apakah dua variabel hampir sepenuhnya bergantung satu sama lain? Adakah titik data yang merupakan outlier – kasus di mana, misalnya, data yang telah dikumpulkan memiliki nilai yang berada di luar rentang normal? Menentukan apa itu outlier bisa agak subjektif. Outlier, baik outlier global, outlier kontekstual, maupun outlier kolektif, dapat disebabkan oleh data yang buruk, tetapi bisa juga valid. Dalam beberapa kasus, outlier mungkin justru merupakan apa yang dicari oleh sebuah bisnis: titik data menarik yang akan berguna bagi bisnis tersebut untuk dipahami jika berfokus pada, misalnya, deteksi penipuan. Di lain waktu, outlier menghalangi. Tim perlu memastikan bahwa mereka tahu mana outlier yang mana. Deteksi anomali, meskipun dapat diterapkan di banyak domain, merupakan bagian penting dari prapemrosesan data untuk mengidentifikasi dan menangani outlier.

Ilmuwan AI juga melihat statistik ringkasan yaitu, karakterisasi sederhana dari data. Persentil dapat membantu mengidentifikasi rentang untuk sebagian besar data. Rata-rata dan median dapat memberikan gambaran umum tentang distribusi data. Korelasi dapat menunjukkan hubungan antar elemen data. Visualisasi seringkali bermanfaat. Ilmuwan AI dapat menghasilkan visualisasi data ini dengan memasukkannya ke dalam keranjang dan memplotnya. Faktor-faktor penting yang perlu dipertimbangkan pada tahap ini meliputi rata-rata, atau mean, dari setiap kumpulan data dan deviasi standarnya serta korelasi orde pertama antar variabel. Diagram kotak membantu mengidentifikasi outlier. Histogram dan diagram kepadatan memberikan gambaran tentang penyebaran data. Diagram sebar dapat menggambarkan hubungan antara dua variabel, yang juga dikenal sebagai hubungan bivariat. Semua ini memberikan pemahaman yang lebih baik kepada bisnis dan tim AI tentang data yang dimilikinya.

Setelah data dipahami dengan baik, ilmuwan AI melakukan rekayasa fitur. Ini mengekstraksi fitur-fitur yang melakukan pekerjaan terbaik dalam memprediksi apa yang ingin diketahui perusahaan. Untuk melihat churn pelanggan, misalnya, bank ritel memeriksa semua fitur yang tersedia dari sampel pelanggannya dan mencari tahu fitur mana yang memiliki nilai prediktif. Ini mungkin termasuk berapa lama seseorang telah menjadi pelanggan, saldo rata-ratanya, seberapa aktif akunnya, jumlah interaksi yang telah ia lakukan dengan staf dukungan pelanggan, bahkan usianya. Data pelatihan bank menggunakan semua fitur tersebut, serta apakah pelanggan tertentu ini churn (label), untuk memungkinkan sistem AI bank mempelajari model yang dapat memprediksi pelanggan lain mana yang paling mungkin churn.

Tim perlu menghasilkan dan memilih fitur terbaik untuk kebutuhan mereka. Fitur yang dihasilkan adalah fitur yang merupakan kombinasi dari data yang ada misalnya, durasi, bukan waktu mulai dan waktu berakhir. Mungkin ada fitur yang dihasilkan yang berguna dalam lebih dari satu kasus penggunaan; jika demikian, tim menyimpan dan membagikan fitur-fitur ini di pasar fitur daripada membuatnya kembali setiap kali dibutuhkan. Pasar fitur dijelaskan di Bab 9. Jika pendapatan adalah data mentah dan telah dikelompokkan ke dalam wadah rentang pendapatan, tim memastikan bahwa jika orang lain perlu menggunakan rentang yang sama, mereka dapat dengan mudah melakukannya.

Untuk pemodelan AI, tidak hanya penting untuk menentukan fitur mana yang harus digunakan, tetapi fitur-fitur ini harus ditransformasikan agar bekerja lebih efisien dalam suatu model, membuat model lebih alami untuk ditafsirkan, menangkap hubungan yang lebih kompleks, mengurangi redundansi dan dimensionalitas data, dan menstandarisasi atau menormalkan variabel. Transformasi fitur dapat mengambil berbagai bentuk. Dalam konteks pembelajaran mesin, normalisasi berarti menskalakan ulang variabel, sehingga nilainya berkisar antara nol dan satu. Normalisasi dapat membantu perhitungan karena model umumnya menggunakan beberapa fitur, dan normalisasi membawa nilai berbagai fitur ke skala yang sama tanpa memberi bobot lebih pada satu fitur di atas yang lain.

Misalnya, jika usia populasi sampel berkisar antara 22 hingga 92, usia tersebut dapat dipetakan, dengan 22 dipetakan ke 0 dan 90 dipetakan ke 1, dengan mengurangi 22 dari fitur usia dan membagi hasilnya dengan 70 (92 dikurangi 22). Sebagai alternatif, untuk membuat

penskalaan lebih umum sehingga berfungsi untuk semua usia yang diketahui, usia nol dapat dipetakan ke 0 dan usia 125 dapat ditetapkan ke 1. Jika fitur lainnya adalah gaji, maka penskalaan keduanya akan mencegah pembobotan yang berlebihan pada salah satu dibandingkan yang lain berdasarkan angka-angka di dalamnya. Standardisasi memiliki hasil yang serupa, tetapi dicapai dengan mengurangi rata-rata setiap variabel dari setiap sampel dan membaginya dengan deviasi standar variabel tersebut.

Binning dan bucketing mengubah variabel dengan kardinalitas tinggi yaitu, banyak kemungkinan nilai menjadi variabel dengan kardinalitas rendah. Ini mungkin termasuk mengambil seluruh rentang usia pelanggan tersebut, yang merupakan angka tinggi katakanlah, 18–92 – dan mengelompokkannya ke dalam bin atau bucket: misalnya, orang yang berusia 18–29, 30–39, 40–49, dan seterusnya. Teknik-teknik ini dapat diterapkan pada fitur numerik dan kategoris, yang berarti mereka bekerja bahkan jika data yang dimaksud tidak menyertakan angka apa pun. Seseorang dapat, misalnya, mengambil semua warna pelangi dan mengelompokkannya ke dalam warna dingin dan warna hangat, atau warna cerah dan warna pastel. Diskritisasi lagi melibatkan penyederhanaan nilai fitur yang sedang dikerjakan tim untuk membuat perhitungan lebih mudah, dalam hal ini dengan proses yang mirip dengan pembulatan ke atas dan ke bawah. Misalnya, tim mungkin mengambil orang-orang yang terdaftar sebagai usia 18.0000, 18.9999, dan semua orang di antaranya dan menyebut mereka semua "18."

Ada juga cara untuk menggabungkan beberapa fitur yang saling bergantung atau melakukan pekerjaan dengan baik dalam memprediksi hal yang sama. Menghilangkan beberapa di antaranya dapat mengurangi redundansi dalam data. Jika tidak dihilangkan, prediksi mungkin kurang akurat karena pengindeksan berlebih pada data yang berlebihan. Pengindeksan berlebih terjadi ketika satu fitur berdampak berlebihan pada keseluruhan model. Proses ini umumnya disebut reduksi dimensionalitas dan umumnya dilakukan dengan menggunakan teknik yang dikenal sebagai analisis komponen utama (PCA), di mana seorang ilmuwan AI melihat semua data dan mengelompokkan variabel sehingga variabel dalam setiap kelompok memiliki korelasi tertinggi satu sama lain dalam kelompok tersebut, dan korelasi terendah dengan variabel dalam kelompok lain. Variabel dalam setiap kelompok ditransformasikan menjadi satu fitur. Sekarang setiap fitur cukup independen satu sama lain. Namun, hal ini dapat mengurangi interpretasi manusia terhadap setiap fitur, karena merupakan kombinasi dari berbagai fitur. Misalnya, melalui PCA, tiga fitur tinggi badan, usia, dan ukuran sepatu dapat digabungkan menjadi satu.

Salah satu aspek penting dari rekayasa fitur adalah pembuatan embedding. Embedding adalah vektor variabel kontinu yang merepresentasikan beberapa variabel kategoris diskrit (misalnya, pelanggan atau produk). Misalnya, alih-alih kategori "parfum", Anda dapat menggunakan vektor enam dimensi (0,01359, 0,00075997, 0,24608, 0,2524, 1,0048, 0,06259) yang dibuat dari data sumber menggunakan pembelajaran mendalam. Embedding ini kemudian dapat disimpan di pasar fitur dan digunakan dalam model pembelajaran mesin dan pembelajaran mendalam klasik lainnya sebagai fitur input. Meskipun ini merupakan konsep yang agak abstrak, penggunaan embedding seringkali meningkatkan akurasi model hingga

10%. Seringkali, dalam metode yang lebih sederhana seperti pengodean one-hot, variabel dummy dan pengodean label digunakan untuk mengonversi tipe data kategoris menjadi biner dari pelatihan model dan wawasan data.

Model yang berbeda mungkin memiliki persyaratan rekayasa fitur yang berbeda atau bahkan rekayasa fitur bawaan, sehingga mentransformasi ulang data masukan membantu menciptakan masukan terbaik sehingga model AI dapat mempelajari data tersebut secara maksimal.

#### 8.4 MEMILIH DAN MELATIH MODEL

Selanjutnya, ilmuwan AI memilih algoritma yang akan digunakan sebagai dasar model AI. Algoritma dapat dianggap sebagai fungsi matematika dasar, misalnya, fungsi linear  $f(x) = ax + b$ . Sebuah model dibuat dengan melatih algoritma pada sekumpulan data tertentu. Melanjutkan contoh sebelumnya, satu model mungkin  $m1(x) = 2x + 3$ , atau model lain yang didasarkan pada algoritma yang sama tetapi menggunakan data masukan yang berbeda dapat berupa  $m2(x) = 7x + 2$ . Dalam model tersebut, parameter  $a$  dan  $b$  dari algoritma telah dilatih, berdasarkan data masukan yang berbeda.

Agar model prediktif dianggap baik, empat kriteria perlu dipertimbangkan. Pertama, model tersebut harus memiliki kinerja model yang tinggi. Model tersebut harus akurat model tersebut harus membuat prediksi yang bermanfaat. Biasanya, model terbaik didefinisikan sebagai model yang memiliki kesalahan terkecil yang diukur dengan fungsi kesalahan matematis, yang lebih umum disebut fungsi kerugian. Terkadang fungsi kerugian dapat berupa kombinasi kesalahan terendah dan bias terendah, jika kewajaran merupakan aspek penting dari model ini.

Kedua, model tersebut harus dapat diinterpretasikan: relatif mudah untuk menjelaskan bagaimana prediksi dibuat. Ini mungkin terdengar jelas, tetapi tidak selalu demikian. Jaringan saraf dalam, pada dasarnya adalah kotak hitam, dan kotak hitam tidak dapat dijelaskan. Kita tahu apa yang masuk dan apa yang keluar, tetapi apa yang terjadi di dalamnya seringkali tidak dapat direpresentasikan dalam istilah yang dapat dipahami manusia. Di sisi lain, pohon keputusan umumnya mudah diinterpretasikan secara prinsip, tetapi jika memiliki ribuan cabang, mungkin lebih sulit untuk memahaminya secara praktis. Mengingat kebutuhan perusahaan, tim mengukur kelebihan dan kekurangan setiap model.

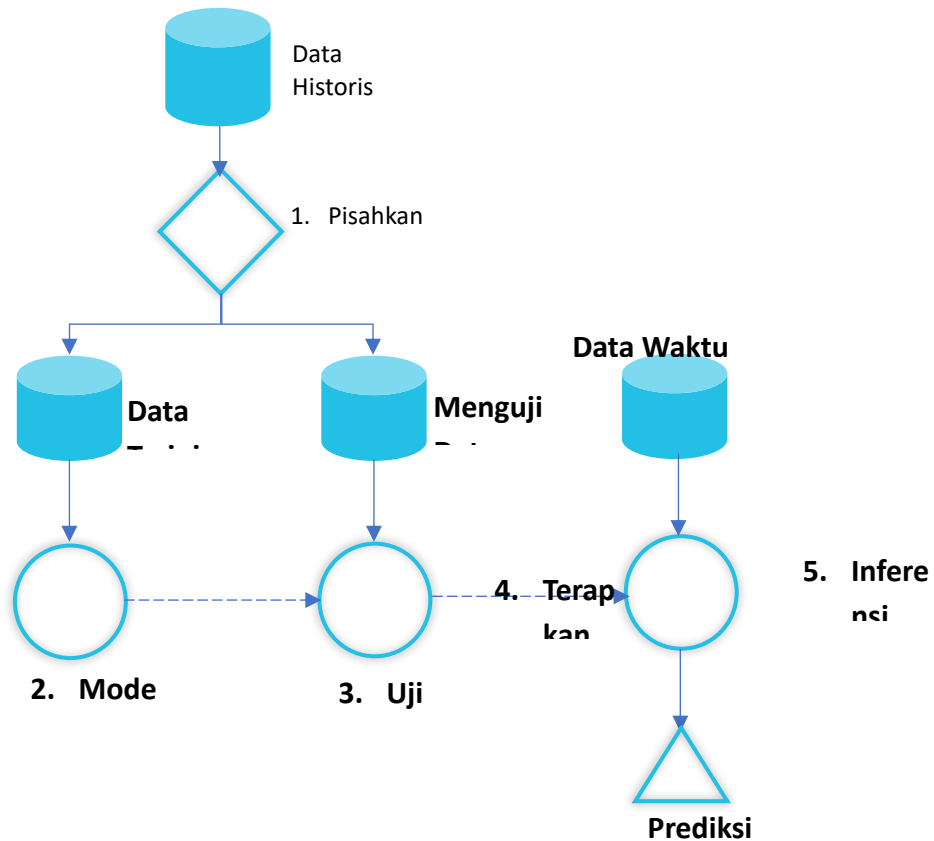
Ketiga, model tersebut harus cepat dan skalabel. Pelatihannya tidak boleh memakan waktu terlalu lama, dan tidak boleh terlalu lama untuk membuat prediksi pada waktu inferensi. Model berubah seiring dengan perubahan kebutuhan bisnis, pengumpulan data, dan lingkungan produksi, sehingga perusahaan membutuhkan model yang fleksibel. Terakhir, model tersebut harus tangguh, memberikan hasil yang konsisten ketika input yang sama atau serupa digunakan.

Setelah tim mengetahui apa yang ingin diprediksi atau dipelajari dan mulai memeriksa data yang berisi fitur-fitur yang akan digunakan untuk melatih modelnya, serta menyiapkan data pelatihan dengan membuat label, anggota tim harus membuat keputusan. Algoritma AI atau pembelajaran mesin seperti apa yang harus dipilih? Seperti yang telah disebutkan di Bab

2, ada banyak pilihan, termasuk jaringan saraf dalam (deep neural nets), pohon keputusan (decision tree), regresi logistik (logistic regression), dan mesin vektor pendukung (support vector machines), dan masih banyak lagi. Penting untuk disadari bahwa berbagai model, betapa pun bagusnya, dapat unggul dalam berbagai hal. Masing-masing model lebih baik dalam berbagai jenis tugas, meskipun beberapa di antaranya hanya merupakan varian kecil dari yang lain. Satu model mungkin berkinerja baik untuk banyak kasus penggunaan, tetapi tidak semua kasus penggunaan. Itulah sebabnya seorang ilmuwan AI dapat mempertimbangkan banyak opsi model, dan terkadang dapat menggunakan lebih dari satu. Saat mempertimbangkan opsi-opsi tersebut, ia harus dapat memandu tim bisnis dalam mengambil keputusan tentang apa yang optimal dan apa saja trade-off yang perlu dilakukan untuk mendapatkan hasil yang dibutuhkan oleh kasus penggunaan tersebut. Seringkali, pertukaran ini memiliki implikasi bisnis.

Saat seorang ilmuwan AI memulai proses pemodelan, ia perlu mempertimbangkan dengan cermat pendekatannya dalam mengambil sampel data untuk pelatihan, pengujian, dan validasi model. Pengambilan sampel data adalah metode statistik untuk memilih observasi (baris) dari kumpulan data agar dapat memperkirakan parameter tentang populasi. Proses ini terasa sederhana, tetapi seringkali rumit. Pengambilan sampel perlu dilakukan dengan mempertimbangkan distribusi, ukuran dan jenis data, serta daya komputasi yang tersedia. Secara umum, ada beberapa jenis utama, tergantung pada kasus penggunaan dan algoritma yang digunakan: pengambilan sampel acak, pengambilan sampel bertingkat, pengambilan sampel klaster, pengambilan sampel multistage, dan pengambilan sampel sistematis.

Untuk membuat model pembelajaran terbaik, ilmuwan AI umumnya membagi data menjadi dua kelompok. Sekitar 60–80% data digunakan untuk melatih model, dan sisanya digunakan untuk menguji model. Tujuan pembelajaran mesin adalah untuk memprediksi secara akurat berdasarkan data yang belum pernah dilihat sebelumnya oleh model. Oleh karena itu, data dibagi menjadi dua kelompok pertama untuk pelatihan dan kelompok kedua untuk digunakan sebagai data "yang sebelumnya tidak terlihat" yang digunakan untuk mengukur akurasi prediksi. Hal ini menunjukkan apakah model cukup tergeneralisasi untuk bersifat prediktif dengan data yang belum pernah terlihat, alih-alih hanya bersifat deskriptif dari data sebelumnya.

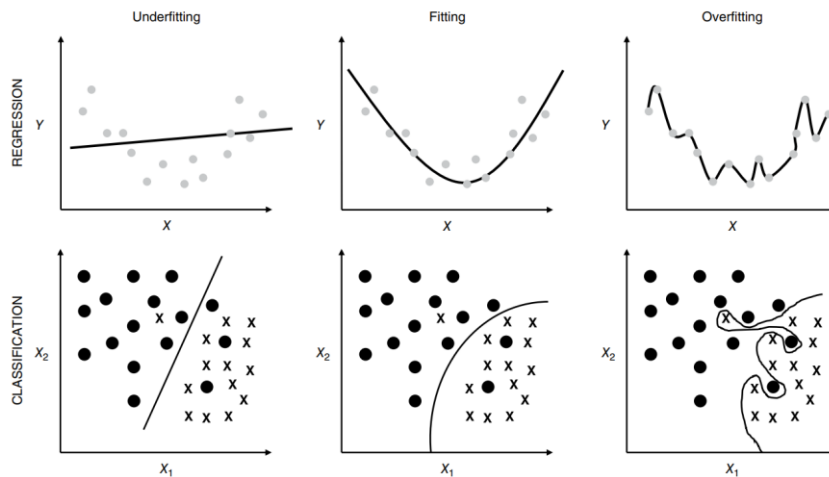


**Gambar 8.4** Proses untuk pelatihan dan validasi model.

Beberapa konsep penting untuk memahami kinerja model atau akurasi model. Pertama, sebuah model didefinisikan oleh serangkaian parameter model. Parameter-parameter ini adalah variabel-variabel dalam model yang dipelajari melalui proses pelatihan. Proses pelatihan mengeksplorasi berbagai kombinasi parameter model dan memilih yang memberikan kesalahan terendah (atau kerugian optimal). Semakin banyak parameter yang dimiliki model, semakin kompleks model tersebut artinya, semakin banyak parameter yang harus dipelajarinya. Hal ini terkadang disebut sebagai kompleksitas model atau kapasitas model. Jika Anda memiliki terlalu sedikit parameter untuk data pelatihan, Anda tidak akan mendapatkan model yang sangat akurat. Ini disebut *underfitting*. Di sisi lain, jika Anda memiliki terlalu banyak parameter, model Anda mulai menjadi sangat spesifik terhadap data pelatihan dan tidak dapat digeneralisasi dengan baik ke kumpulan data baru artinya, model tersebut tidak terlalu akurat untuk data baru. Ini disebut *overfitting*. Gambar 8.5 menunjukkan contoh sederhana *overfitting* dan *underfitting* untuk model regresi dan model klasifikasi.

*Underfitting* terjadi ketika model tidak cukup kompleks untuk merepresentasikan apa yang sebenarnya terjadi dalam data. Contoh ekstremnya adalah ketika seseorang mencoba menentukan harga yang wajar untuk sesuatu yang ingin dijualnya misalnya, rumah berusia 10 tahun. Ia membangun model menggunakan data yang telah dikumpulkannya tentang penjualan rumah di kota tersebut dengan luas persegi dan harga jual setiap rumah. Ia mengatakan bahwa model tersebut harus memiliki dua parameter, yaitu harga  $P$  adalah kelipatan ( $A$ ) dari luas persegi  $S$  ditambah parameter lain ( $B$ ). Jadi,  $P = A \times S + B$ . Model ini

akan berupa garis lurus, dan ia menemukan bahwa model tersebut berfungsi untuk rumah-rumah dengan harga rendah hingga titik tertentu, tetapi rumah-rumah dengan harga tinggi tidak sesuai dengan garis lurus tersebut sama sekali (seperti gambar pertama pada Gambar 8.5). Ini disebut underfitting.



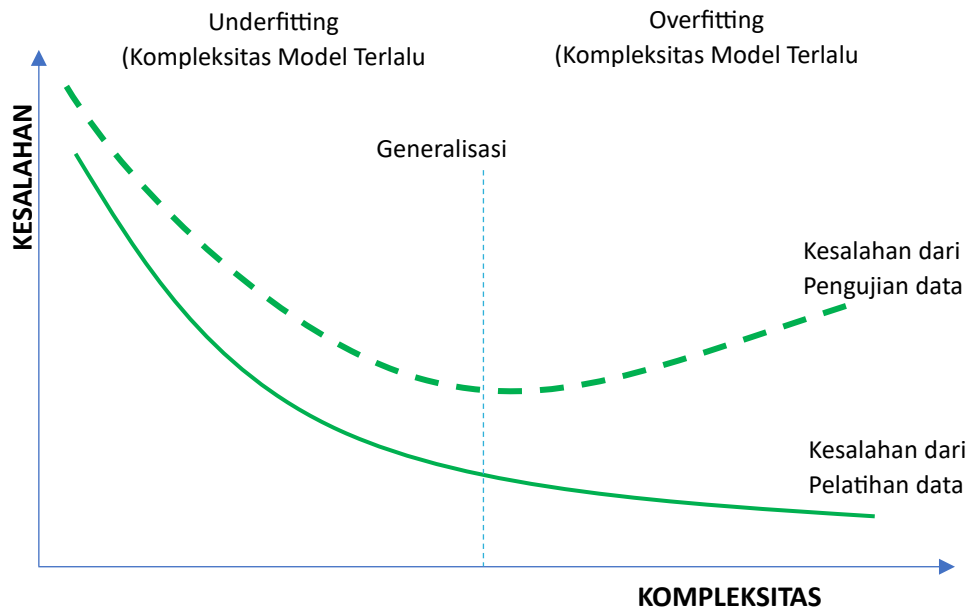
**Gambar 8.5** Underfitting dan overfitting untuk model regresi (atas) dan untuk model klasifikasi (bawah).

Di sisi lain, overfitting terjadi ketika model Anda tidak berkinerja baik pada data selain data pelatihan Anda. Melanjutkan contoh ekstrem kita, misalkan penjual kita mengatakan bahwa model tersebut harus memiliki 100 parameter, dengan mempertimbangkan kelipatan luas persegi, seperti  $S, S^2, S^3$ , dan seterusnya. Model yang dihasilkannya terlalu detail untuk memprediksi dengan baik berapa banyak uang yang seharusnya ia kenakan (seperti gambar ketiga pada Gambar 8.5). Ini adalah overfitting. Ini memodelkan data pelatihan terlalu baik. Ini terjadi ketika detail dan noise dalam data pelatihan diambil dan dipelajari sebagai konsep oleh model. Masalahnya, detail dan noise ini mungkin tidak berlaku untuk data baru dan oleh karena itu akan berdampak negatif pada kemampuan model untuk menggeneralisasi.

Ada berbagai cara untuk mengelola overfitting atau underfitting. Mereka biasanya ditangani dalam pengujian, yang merupakan bagian dari proses pemodelan ilmuwan AI. Proses pengujian umumnya dimulai dengan model underfitting, yang tidak memberikan hasil seakurat mungkin. Kompleksitas kemudian terus ditambahkan ke model hingga ada bukti overfitting. Pada saat itu, ilmuwan AI kemudian berhenti menguji dan kembali ke model sebelumnya yang memiliki tingkat kesalahan pengujian terendah.

Gambar 8.6 memberikan contoh bagaimana kesalahan pelatihan dan pengujian digunakan untuk memilih kompleksitas model yang optimal. Seiring bertambahnya kompleksitas model, model tersebut semakin baik dalam mempelajari data pelatihan. Namun, sebagian pembelajaran tersebut hanya spesifik untuk data tempat model dilatih dan tidak mewakili karakteristik data secara umum. Masalah muncul setelah pelatihan, ketika model diberikan data yang belum pernah dilihat sebelumnya. Model tersebut mungkin mencoba menjelaskan keistimewaan yang ada dalam proses pelatihan dengan membuat prediksi yang

tidak seakurat mungkin. Dalam situasi ini, model tersebut mungkin tidak akan sukses yang seharusnya dalam memprediksi dari data baru. Model yang paling umum untuk data yang tidak terlihat adalah model dengan kesalahan set pengujian terendah, ditunjukkan sebagai "generalisasi terbaik" pada Gambar 8.6.



**Gambar 8.6** Kesalahan pelatihan versus kesalahan pengujian.

Seperti yang disebutkan sebelumnya, data dipecah sehingga sebagian dapat digunakan untuk pelatihan dan sebagian lagi untuk pengujian. Memecah data memungkinkan pengujian model pada data yang berbeda dari data tempat model dilatih. Biasanya, model dilatih pada sekitar 60% hingga 80% data, menyisakan sisanya untuk pengujian. Pendekatan lain yang umum digunakan disebut *k-fold cross validation*. Ini terjadi ketika data dibagi menjadi  $k$  kelompok (misalnya,  $k = 5$ ) dan kemudian empat kelompok pertama digunakan untuk pelatihan dan yang terakhir untuk pengujian. Selanjutnya, kelompok 1, 2, 3, dan 5 digunakan untuk pelatihan, dan kelompok 4 digunakan untuk pengujian, dan seterusnya. (Perhatikan bahwa jika data adalah deret waktu, pendekatan  $k - fold$  ini tidak akan berfungsi.) Tingkat kesalahan rata-rata dari  $k$  runs digunakan sebagai tingkat kesalahan model. Oleh karena itu,  $k - fold$  cross validation meningkatkan kinerja dan generalisasi model dan sering digunakan ketika ukuran total dataset kecil. Metode populer lainnya adalah stratified  $k - fold$  cross validation dan leave-one-out cross validation.

Dalam masalah klasifikasi pembelajaran mesin, ada langkah tambahan yang diperlukan untuk menguji dan menyetel kinerja model. Ini adalah optimasi *hiperparameter*. Hiperparameter adalah parameter eksternal model, yang ditetapkan oleh ilmuwan AI sebelum proses pelatihan dimulai dan tidak dipelajari melalui proses pelatihan model. Seringkali keluaran model klasifikasi adalah probabilitas untuk kelas tersebut. Ilmuwan AI kemudian dapat menentukan hiperparameter yang menetapkan batas di atas yang mana kelas akan diprediksi menjadi "Ya/Positif" dan di bawahnya kelas akan diprediksi menjadi "Tidak/Negatif."

Misalnya, dalam model "apakah pasien menderita kanker", batas ini awalnya dapat ditetapkan pada 50%. Ilmuwan AI kemudian dapat melihat tidak hanya akurasi model tetapi juga positif palsu dan negatif palsu yang diberikannya dan menyesuaikan batas 50% ke atas atau ke bawah untuk meningkatkan prediksi model.

Salah satu cara untuk melakukan ini adalah dengan menggunakan matriks kebingungan, yang memiliki jumlah kelas yang diprediksi oleh model berdasarkan baris dan jumlah kelas aktual berdasarkan kolom (lihat Gambar 8.7). Dalam contoh kita, kita memiliki dua kategori positif dan negatif untuk diagnosis kanker tetapi hal yang sama berlaku untuk lebih banyak kelas; misalnya, Anda akan memiliki matriks kebingungan  $5 \times 5$  untuk masalah dengan lima kategori. Empat kuadran dari matriks dua kelas tersebut adalah:

1. Positif benar (TP) – jumlah pasien dengan kanker dan diprediksi secara akurat menderita kanker.
2. Negatif benar (TN) – jumlah pasien tanpa kanker dan diprediksi secara akurat tidak menderita kanker.
3. Positif salah (FP) – jumlah pasien tanpa kanker tetapi diprediksi secara salah menderita kanker (juga disebut kesalahan Tipe 1).
4. Negatif salah (FN) – jumlah pasien dengan kanker tetapi diprediksi secara salah tidak menderita kanker (juga disebut kesalahan Tipe 2).

Dalam contoh ini, hasil negatif palsu dapat berpotensi mengancam jiwa, sedangkan hasil positif palsu tidak terlalu buruk, meskipun kemungkinan akan memberikan pengalaman buruk bagi pasien dan menimbulkan biaya yang tidak perlu. Oleh karena itu, penting untuk mengurangi rasio negatif palsu sekaligus meningkatkan rasio positif-benar dan negatif-benar. Dua metrik yang sering digunakan adalah rasio positif-benar (TPR), yang mengukur persentase positif-benar (TP) dari semua yang diprediksi positif ( $TP + FN$ ); dan rasio positif-salah (FPR), yang mengukur persentase positif-salah (FP) dari semua yang sebenarnya negatif ( $FP + TN$ ).

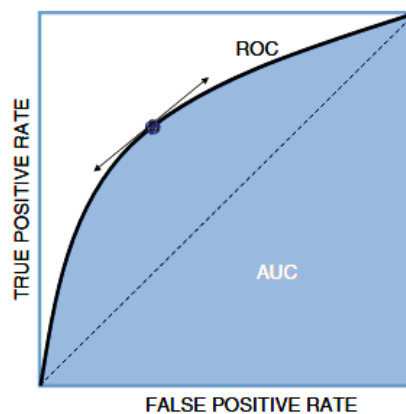
		Nilai Aktual	
		Positif	Negatif
Nilai Prediksi	Positif	True Positif	False Positif
	Negatif	False Negatif	True Negatif

**Gambar 8.7** pengaturan Confusion Matrix

Kurva karakteristik operasi penerima (ROC) pada Gambar 8.8 menunjukkan rasio positif-benar terhadap rasio positif-salah model. Idealnya, TPR harus setinggi mungkin, dan FPR harus serendah mungkin. Agar hal ini terjadi, area di bawah kurva (AUC) harus sebesar

mungkin, sehingga mendorong kurva ke kiri atas. Pemodel memilih titik pada kurva yang paling dekat dengan sumbu kiri atas, sehingga memilih keseimbangan antara positif palsu dan negatif palsu.

Melalui proses ini, ilmuwan AI memilih model setelah bereksperimen dengan beberapa algoritma dan mengatasi masalah underfitting atau overfitting. Pemilihan model memastikan bahwa kurva ROC berada sejauh mungkin ke kiri dan atas dan area di bawah kurva adalah yang terbesar. Ini menunjukkan bahwa model tersebut sebaik yang akan didapat.



**Gambar 8.8** Kurva karakteristik operasi penerima (ROC) dan area di bawah kurva (AUC).

Ilmuwan AI kemudian menyetel hiperparameter, memilih lokasi optimal pada kurva ROC spesifik ini untuk memecahkan masalah bisnis (ditunjukkan oleh panah pada Gambar 8.8). Untuk akurasi tertinggi, terlepas dari tingkat negatif palsu dan positif palsu, yang terbaik adalah berada pada titik pada kurva ROC yang paling dekat dengan sudut kiri atas. Dalam contoh diagnosis kanker, ilmuwan AI ingin mengurangi rasio negatif palsu dan mengimbangi rasio positif palsu jika diperlukan, jadi ia menetapkan hiperparameter sedemikian rupa sehingga ambang batas (hiperparameter) berada lebih ke kanan pada kurva ROC.

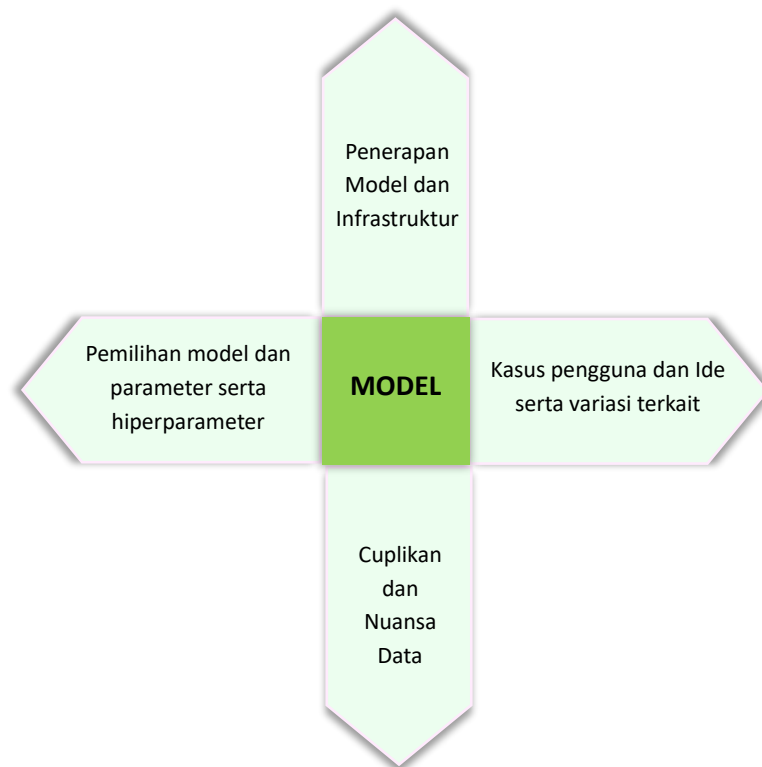
## 8.5 MENGELOLA MODEL

Membangun algoritma pembelajaran mesin di dunia nyata itu rumit dan sangat iteratif. Seorang ilmuwan AI dapat membangun puluhan atau bahkan ratusan model sebelum sampai pada satu model yang paling memenuhi kriteria penerimaan. Melacak model-model tersebut ibarat menjalankan misi pencarian beberapa pendaki yang hilang di kaki bukit Himalaya. Area yang ditelusuri oleh tim penyelamat, bahkan area di mana para pendaki tidak ditemukan, harus ditandai dengan jelas. Jika tidak, para pencari akan membuang waktu berharga untuk berulang kali menyalisir area yang sama. Demikian pula, manajemen model sangat penting untuk mencakup semua konfigurasi yang memengaruhi kinerja model dan melacak area mana yang telah tercakup sebelumnya atau yang belum. Ini termasuk pelacakan (lihat Gambar 8.9):

- Data apa yang digunakan untuk melatih model
- Kasus penggunaan atau variasi apa yang ingin diimplementasikan oleh model ini

- Konfigurasi infrastruktur apa yang diperlukan untuk melatih model ini atau digunakan selama inferensi, serta informasi lain yang diperlukan untuk penerapan
- Eksperimen model apa yang telah dicoba, seperti pemilihan model atau optimasi hiperparameter

Tidak melacak model dapat menyebabkan penumpukan utang teknis dan hilangnya pengulangan model. Jika konfigurasi tidak dilacak dengan benar, berbagai hal dapat menjadi salah. Tanpa catatan terorganisir dari eksperimen yang dilakukan, mustahil untuk memastikan apakah seseorang telah mencoba sesuatu yang identik atau sangat mirip dengan apa yang akan dilakukan orang lain. Tidak melacak eksperimen tidak hanya tidak efisien dan mahal; tetapi juga mempersulit kolaborasi, peninjauan, revisi, dan pengembangan dari karya orang lain. Tanpa proses yang jelas, akan lebih sulit untuk melakukan meta-analisis lintas model, dan tim akan kehilangan waktu untuk melakukan penelusuran kembali ketika tidak ada yang dapat dengan mudah mengetahui, misalnya, parameter mana yang digunakan untuk hasil sebelumnya. Melacak versi model, misalnya perubahan apa yang terjadi antara versi model 233 dan versi 237, juga akan lebih sulit.



**Gambar 8.9** Manajemen model komprehensif mencakup empat jenis konfigurasi.

Masalah potensial dapat muncul saat menambahkan, mengubah, atau menghapus fitur dalam model AI seiring perkembangannya. Perubahan umumnya mengharuskan model divalidasi ulang, yaitu, dijalankan kembali melalui pembelajaran dan pengujian, karena konfigurasi yang telah berubah dapat memengaruhi validitas prediksi. Ketika ada masalah yang sangat rumit untuk dipecahkan, sebuah tim mau tidak mau harus mengubah berbagai hal di sepanjang prosesnya. Melacak apa yang berubah, kapan, dan bagaimana perubahan itu

berkembang merupakan suatu keharusan. Mungkin beberapa sensor tidak berfungsi pada hari tertentu atau peristiwa cuaca yang signifikan dan tak terduga memengaruhi jadwal pengiriman. Ketika hal-hal ini bertambah seiring waktu, model menjadi kurang andal kecuali perubahan ini dilacak dan ditangani dengan tepat dalam proses pemodelan.

Fitur juga dapat terpengaruh oleh berbagai masalah operasional. Satu fitur mungkin telah dicatat secara tidak benar selama minggu tertentu. Fitur lain mungkin belum tersedia sebelum sensor baru diaktifkan. Perubahan dalam format akuisisi data untuk fitur ketiga mungkin memerlukan penulisan ulang kode yang melakukan prapemrosesan. Jika satu jenis model digantikan dengan yang lain, lingkungan komputasi mungkin perlu dikonfigurasi ulang untuk menyediakan memori atau daya pemrosesan tambahan.

Ketergantungan data yang tidak diperlukan merupakan masalah lain yang mengancam keandalan suatu model. Ini berarti mempertahankan fitur yang tidak digunakan dalam model: fitur yang tidak digunakan dan, oleh karena itu, dapat dihapus. Hal ini sering terjadi karena tidak ada waktu untuk menghapusnya, atau tim tidak cukup terorganisir untuk menyadari bahwa fitur tersebut tidak lagi digunakan. Hal ini tidak masalah jika fitur-fitur ini tidak pernah memengaruhi hasil, tetapi jika data berubah, hal tersebut dapat berdampak negatif pada hasil.

Masalah potensial lainnya adalah hutan pipeline. Hal ini terjadi ketika tidak ada arsitektur data yang jelas atau standar yang terdefinisi dengan baik untuk membuat semua elemen berurutan dalam proses koheren dari ujung ke ujung sehingga data dapat mengalir dengan lancar dalam proses produksi. Jika terdapat hutan pipeline, tim DevOps menemukan bahwa persiapan data kurang konsisten dari waktu ke waktu, sehingga sulit untuk melacak apa yang bergantung pada apa. Ini berarti bahwa hal-hal seperti mendeteksi kesalahan dan memulihkan dari kegagalan menjadi sangat rumit, membuat produksi menjadi tidak efisien dan tidak praktis.

Ketika sebuah organisasi mengeksplorasi peluang untuk menggunakan AI, diperlukan pendekatan formal untuk melacak berbagai hal: menguji eksperimen, menangkap apa yang berhasil, dan memelihara "kuburan ide" untuk konsep-konsep yang telah diuji dan dinyatakan tidak dapat dipertahankan. Proses seperti ini memudahkan berbagai hal bagi individu dan organisasi. Hal ini memungkinkan ilmuwan AI untuk melacak pekerjaan mereka secara detail, memberi mereka catatan eksperimen mereka. Hal ini juga memungkinkan mereka untuk menangkap wawasan yang bermakna di sepanjang jalan, mulai dari bagaimana normalisasi memengaruhi hasil hingga bagaimana fitur granular tampaknya memengaruhi kinerja untuk subset data tertentu. Tidak ada organisasi yang mencapai tingkat visibilitas dan analisis yang diperlukan saat mengelola model melalui spreadsheet. Sistem yang jauh lebih canggih diperlukan untuk mendokumentasikan apa yang telah dilakukan dan apa hasilnya. Untuk berhasil dalam skala perusahaan, sebuah organisasi harus dapat menyimpan, melacak, membuat versi, dan mengindeks model serta jalur data. Hal ini memungkinkan model untuk ditanyakan, direproduksi, dianalisis, dan dibagikan, bahkan jika model-model tersebut akhirnya tidak digunakan. Manajemen model yang baik memberdayakan ilmuwan AI untuk meninjau, merevisi, dan mengembangkan karya satu sama lain, membantu mempercepat kemajuan dan menghindari pemborosan waktu.

Selain itu, tim dapat kehilangan wawasan dan mengorbankan kesempatan untuk memahami apa yang mereka lihat. Ini mungkin termasuk bagaimana parameter individual yang berkontribusi terhadap kompleksitas model telah memengaruhi hasil, atau kemampuan untuk memperhatikan bahwa fitur tertentu tampaknya meningkatkan kinerja untuk subset data tertentu. Potensi kuantitas ide dan nuansa dapat dengan cepat menjadi berlebihan. Untuk menghindari kompleksitas tersebut, tim harus merancang dan menerapkan proses manajemen model otomatis untuk melacak dan mengelola siklus hidup kasus penggunaan, manajemen model, dan eksperimen. Melakukan hal tersebut membantu dalam melacak kinerja ide dan memastikan kualitas ide-ide tersebut. Juga jauh lebih efisien untuk menyediakan informasi kepada tim tentang dan akses mudah ke ide-ide yang berhasil dan tidak berhasil untuk menghindari upaya duplikasi dan potensi konflik. Hal ini juga memungkinkan organisasi untuk melakukan meta-analisis lintas model untuk menjawab pertanyaan yang lebih luas, seperti pengaturan hiperparameter apa yang paling sesuai untuk berbagai jenis model atau fitur. Semua ini mempercepat efisiensi penerapan layanan AI.

## **8.6 MENGUJI, MENYEBARKAN, DAN MENGAKTIFKAN MODEL**

Bahkan setelah seorang ilmuwan AI menghasilkan model yang skalabel, masih banyak hal yang perlu dilakukan dalam menyebarkan proyek AI yang belum ditangani pada tahap awal. Model tersebut harus diuji. Tim tata kelola risiko model harus mengeluarkan persetujuan akhir. Model tersebut harus terintegrasi dengan aplikasi yang menggunakannya. Aplikasi harus diperluas untuk memberikan hasil dari model kepada orang-orang yang membutuhkannya, dan mekanisme harus tersedia bagi para pengguna untuk memberikan umpan balik.

### ***Pengujian***

Pengujian berkelanjutan sangat penting untuk menciptakan dan menjalankan model yang tangguh dalam jangka waktu yang wajar. Sebagaimana dibahas sebelumnya, selama pembuatannya, model AI harus diverifikasi keakuratannya, yang harus berada pada ambang batas tertentu agar praktis untuk kasus penggunaan. Setelah model dibangun dan diintegrasikan ke dalam data dan aplikasi yang menggunakannya, model tersebut harus melalui bentuk pengujian lain, seperti halnya aplikasi TI atau modul teknologi baru yang sedang diterapkan. Salah satunya adalah pengujian fungsional menyeluruh (end-to-end). Apakah model tersebut bekerja dalam lingkungan seperti produksi, mampu menerima data untuk inferensi? Apakah model tersebut menghasilkan keluaran atau prediksi pada tingkat akurasi yang diharapkan? Apakah aplikasi yang menggunakannya menerima keluaran dan menampilkan atau menindaklanjutinya?

Model menyeluruh juga perlu melalui pengujian kinerja. Apakah model tersebut memberikan hasil dengan segera, dengan throughput yang diharapkan mengingat beban data yang diharapkan? Dalam beberapa kasus, kode yang ditulis oleh ilmuwan AI mungkin tidak berskala produksi, dan mungkin harus ditulis ulang untuk skalabilitas yang lebih tinggi. Pengujian lain yang mungkin perlu dijalankan oleh tim adalah untuk kaskade model. Kaskade model terjadi ketika satu model memasukkan keluaran ke model berikutnya, yang menggunakan keluaran ini sebagai masukannya. Dalam situasi ini, kehati-hatian harus

diperhatikan dalam cara penanganan data. Apa yang terjadi jika keluaran dari model kedua salah? Bagaimana tim dapat memastikan di mana masalah muncul? Apakah di dalam model pertama atau kedua?

Pengujian, seperti pengujian integrasi, pengujian penerimaan pengguna, staging, dan terakhir, pengujian produksi, perlu dilanjutkan seiring perkembangan model melalui berbagai lingkungan. Beberapa jenis pengujian tambahan terkadang dilakukan di lingkungan produksi. Salah satunya disebut uji canary atau penerapan canary. Pengujian canary terjadi ketika model yang diterapkan hanya diekspos ke sekelompok kecil pengguna (biasanya kurang dari 5%) dan dipantau untuk memastikannya berfungsi dengan baik. Pengujian ini bertindak sebagai burung kenari di tambang batu bara. Jika berhasil, model dapat diterapkan ke semua pengguna. Uji produksi lainnya adalah ketika dua model atau lebih diterapkan untuk melihat model mana yang paling efektif di dunia nyata. Pengujian A/B menerapkan berbagai versi model secara paralel, dan kelompok pengguna yang berbeda dikonfigurasi untuk menggunakan model yang berbeda. Pengujian A/B sering digunakan ketika suatu model sedang disempurnakan atau ditingkatkan. Model lama dan baru diuji A/B, dan model yang berkinerja lebih baik tetap dipertahankan. Dalam beberapa kasus, pengujian A/B digunakan untuk menguji hipotesis di mana A adalah kelompok uji dan B adalah kelompok kontrol.

### ***Mengelola Risiko Model***

Tim harus memanfaatkan tata kelola risiko model untuk mengurangi risiko yang diakibatkan oleh penggunaan model AI. Tata kelola model diterapkan sebelum, selama, dan sesudah model dikembangkan. Sebelum pemodelan, fokusnya adalah pada kualitas dan bias data. Selama pemodelan, fokusnya adalah pada bias dan interpretabilitas model, selain akurasi model. Setelah pemodelan, tim melakukan pengujian kepatuhan, seperti analisis sensitivitas dan pengujian bias. Selain itu, mekanisme pengaman harus diterapkan jika terjadi kesalahan pada model. Memanfaatkan praktik terbaik industri seperti ini memastikan bahwa perusahaan, dan tim tata kelola risiko modelnya, telah melakukan pekerjaan terbaik untuk menerapkan AI yang aman.

### ***Menyebarkan Model***

Setelah model lulus pengujian dan disetujui oleh proses tata kelola, tim harus mempermudah penyebarannya di setiap lingkungan. Ini berarti menulis skrip untuk mengotomatiskan proses penyebaran. Lebih baik dapat mendorong model ke lingkungan dengan penyebaran sekali klik daripada harus memasukkan perintah manual setiap kali. Semua aktivitas ini harus diotomatisasi menggunakan jenis alat yang sama yang dimanfaatkan di seluruh praktik rekayasa perangkat lunak. Setelah seorang ilmuwan AI membantu memutuskan model atau model-model yang telah diuji, ia membutuhkan dukungan dari tim rekayasa AI DevOps atau pembelajaran mesin yang berpengalaman. Tim AI DevOps sama pentingnya dengan ilmuwan AI itu sendiri: tim ini bertanggung jawab untuk menyiapkan alur integrasi berkelanjutan/penyebaran berkelanjutan (CI/CD) untuk model AI.

Setelah ilmuwan AI memeriksa kode untuk model AI, insinyur AI mengemas model AI dan layanan mikro rekayasa fitur dan membuat unit penyebaran model seperti yang ditunjukkan pada Gambar 8.10. Unit penyebaran model inilah yang disebarkan ke lingkungan

pengujian dan produksi sesuai kebutuhan, termasuk perubahan konfigurasi infrastruktur yang diperlukan. Selanjutnya, tim memasang unit penyebaran model dan secara otomatis menjalankan pengujian relevan yang telah disebutkan sebelumnya. Unit ini mengirimkan notifikasi untuk perbaikan jika terjadi kegagalan, atau jika tidak diperlukan, unit ini akan membongkar lingkungan dan melanjutkan ke tahap berikutnya.

### **Mengaktifkan Model**

Setelah model divalidasi dan disebar, pengembang menulis kode dalam aplikasi perusahaan yang akan menggunakan keluaran model AI. Kode ini digunakan untuk memanggil, atau mengaktifkan, model setiap kali prediksi perlu dibuat. Ada banyak cara untuk melakukan ini. Misalnya, jika algoritma pembelajaran mesin mencoba memutuskan apakah sebuah email adalah spam atau bukan, sistem email akan meneruskan email tersebut ke model melalui antarmuka program aplikasi (API) setiap kali email baru diterima. API adalah perangkat lunak yang dapat menerima masukan (dalam hal ini, email) dan mengirimkan keluaran (prediksi model tentang apakah email tersebut adalah spam atau bukan) ke perangkat lunak lain. Cara umum lain untuk mengaktifkan model adalah dengan menampilkan hasilnya dalam laporan intelijen bisnis (BI), tempat pengguna dapat melihat output dan mengambil keputusan atau tindakan. Cara ketiga untuk menerapkan model adalah dalam mode batch, di mana inferensi model terjadi saat data dibaca, diproses, dan ditulis menggunakan alur data. Dalam hal ini, output inferensi model disimpan dalam basis data yang dapat diakses melalui aplikasi antarmuka pengguna (front-end).

### **Pemantauan Produksi**

Pemantauan model AI yang diterapkan, terkadang disebut operasi kecerdasan buatan (AIOps), sangat penting untuk memastikan model berkinerja baik. Hal ini karena akurasi prediktif suatu model cenderung menjadi kurang akurat seiring waktu, karena data pelatihan yang menjadi dasar model tersebut kemungkinan akan berubah. Jika kinerjanya tidak memadai, tim perlu memodifikasi model, baik melalui pelatihan ulang model maupun mengatasi masalah lain yang muncul. Pemantauan dan perbaikan ini harus bersifat proaktif, bukan hanya reaktif.

Pemantauan proaktif melacak kinerja saat ini berdasarkan serangkaian perjanjian tingkat layanan (SLA). Sebagian besar sistem TI memiliki SLA, seperti utilisasi CPU dan memori, throughput puncak, dan waktu respons. Selain itu, tim dukungan dan pemeliharaan perlu mengetahui bahwa model itu sendiri berkinerja baik artinya, model tersebut mempertahankan tingkat akurasi yang sama. Untuk melakukannya, ada dua pendekatan yang dapat diambil. Salah satunya adalah menangkap keluaran model AI misalnya, apa yang diprediksinya dan kemudian, ketika hasil aktualnya diketahui, menangkap apakah prediksi itu akurat atau tidak. Namun, dalam beberapa kasus, terdapat jeda antara prediksi dan hasil aktual, terkadang berminggu-minggu. Hal ini dapat terjadi ketika sebuah model memprediksi bahwa mengirimkan rekomendasi produk melalui email dan penawaran diskon akan mengonversi pelanggan, tetapi ia baru membeli produk tersebut dan menggunakan kupon diskon tersebut tiga minggu kemudian. Dalam kasus lain, hasil aktualnya mungkin tidak pernah diketahui karena beberapa tindakan telah diambil untuk memitigasi prediksi: misalnya,

pelanggan yang diprediksi akan churn tidak pernah pergi karena ia diberi diskon berdasarkan prediksi churn. Dalam kasus ini, inferensi kausal perlu dimodelkan untuk menghubungkan titik-titiknya. Pendekatan lainnya adalah memantau pergeseran fitur, suatu kondisi di mana statistik data input mulai berubah. Contoh pergeseran fitur adalah model AI yang memprediksi kemungkinan gagal bayar pinjaman untuk pelanggan yang salah satu fitur inputnya adalah distribusi pendapatan pelanggan. Jika distribusi tersebut berubah secara material, dengan lebih banyak pelanggan kini berada dalam rentang pendapatan yang lebih tinggi atau lebih rendah daripada data pelatihan, hal ini dianggap sebagai pergeseran fitur. Hal ini dapat menjadi indikasi bahwa pembelajaran dari set data pelatihan mungkin sudah tidak relevan lagi. Pergeseran fitur merupakan indikator utama bahwa model perlu diperbarui, melampaui ambang batas tertentu. Melacak prediksi terhadap aktual merupakan indikator tertinggal.

## **8.7 KESIMPULAN**

Bab ini telah menjelaskan proses yang dilalui tim AI untuk mengimplementasikan model AI sebagai bagian dari kasus penggunaan. Namun, proses ini tidak berdiri sendiri proses ini membutuhkan orang-orang yang terorganisir termasuk peran yang dibutuhkan dalam sebagian besar proyek TI, dan membutuhkan platform untuk mempermudah dan mempercepat proses. Proses yang dijelaskan dalam bab ini dapat dipandang sebagai persyaratan untuk detail yang dibahas di Bagian III buku ini. Di Bagian IV, kami menjelaskan contoh kerja dan membuat model menggunakan proses ini bagi pembaca yang tertarik untuk mengetahui lebih detail. Semua ini proses, organisasi, dan platform bekerja sama untuk memungkinkan keberhasilan usaha AI perusahaan.

## **BAB 9**

### **MEMBANGUN MESIN AI YANG SEMPURNA**

Ada kepercayaan yang telah lama dipegang bahwa bagian terpenting dari membangun inisiatif AI adalah merancang aplikasi AI algoritma atau model perangkat lunak yang dibuat oleh ilmuwan AI untuk membuat prediksi. Namun seperti yang telah kita lihat, sebagian besar upaya yang dilakukan untuk membangun aplikasi AI perusahaan dihabiskan untuk pengumpulan, pembersihan, dan pelabelan data; pembuatan jalur data; DevOps; penerapan; membangun aplikasi bisnis untuk pengguna akhir; serta memantau dan menyempurnakannya dari waktu ke waktu. Ketika perusahaan mengabaikan tahapan-tahapan ini, mereka seringkali berakhir dengan proses pemodelan yang sangat tidak efisien dan orang-orang bekerja secara terpisah. Hal ini membuang-buang uang dan waktu serta menunda perolehan manfaat bisnis dari model tersebut. Cara untuk menghindari hal ini adalah dengan membangun platform AI yang holistik.

#### **9.1 PLATFORM AI VERSUS APLIKASI AI**

Platform AI adalah solusi perangkat lunak yang kohesif dan terintegrasi dengan baik yang berjalan pada perangkat keras yang skalabel yang mempercepat siklus hidup penuh proyek dan aplikasi AI di seluruh perusahaan. Platform AI meringankan beban kerja, mendorong kerja sama, dan mempercepat adopsi. Platform yang tangguh mendukung semua tugas di seluruh siklus hidup AI. Platform semacam itu merupakan fondasi penting dari aplikasi AI perusahaan. Aplikasi AI dibangun dan dijalankan pada platform AI; aplikasi tersebut terdiri dari perangkat lunak yang memecahkan masalah spesifik atau menjawab pertanyaan tertentu. Dalam 10 tahun terakhir, perusahaan yang telah menggunakan AI berfokus terutama pada pengembangan aplikasi. Misalnya, ada perusahaan ritel yang membangun satu aplikasi AI untuk menargetkan pesan pemasaran, satu lagi untuk rekomendasi daring, dan satu lagi untuk peramalan permintaan dalam rantai pasokan. Masing-masing aplikasi ini membutuhkan versi manajemen data, infrastruktur komputasi, proses penerapan, dan model AI-nya sendiri.

Saat ini, perusahaan-perusahaan terkemuka membangun platform AI untuk mengembangkan dan mengimplementasikan aplikasi-aplikasi ini. Pola pikir platform ini sebagian didorong oleh fakta bahwa membangun platform sekali saja lalu membangun aplikasi di atasnya lebih murah, dan sebagian lagi karena memungkinkan lebih banyak data untuk dibagikan antar departemen seperti pemasaran, e-commerce, dan rantai pasokan.

#### **9.2 APA YANG SEHARUSNYA DILAKUKAN ARSITEKTUR PLATFORM AI**

Secara umum, hal-hal penting yang seharusnya dilakukan platform AI adalah memungkinkan perusahaan AI untuk tumbuh dan berkembang seiring waktu seiring perkembangan teknologi dan kebutuhan, serta mudah digunakan dan oleh karena itu tersedia bagi lebih banyak staf perusahaan dan melakukannya dalam skala perusahaan. Lebih spesifik lagi, platform AI harus meningkatkan probabilitas keberhasilan dengan memungkinkan tim

menjalankan lebih banyak eksperimen tanpa meningkatkan biaya secara linear. Platform harus menyediakan akses mandiri ke teknologi AI dasar untuk mengurangi hambatan dari tim AI, sehingga memungkinkan produktivitas, kecepatan, dan skalabilitas mereka.

Meningkatkan jumlah eksperimen adalah wawasan yang dimiliki Wright bersaudara yang membantu mereka terbang sebelum orang lain. Para pesaing Wright membutuhkan waktu sembilan bulan untuk membangun sebuah pesawat terbang, lalu mencoba menerbangkannya dengan melemparkannya dari tebing (kecil). Setelah pesawat jatuh, mereka akan mengumpulkan serpihan-serpihannya, kembali ke bengkel, menganalisisnya, dan mencoba mencari tahu apa yang perlu ditingkatkan. Ketika Wright bersaudara menyadari hal ini, mereka mengajukan pertanyaan penting: Bagaimana kita dapat meningkatkan jumlah eksperimen yang kita lakukan? Untuk melakukannya, mereka membangun terowongan angin pertama di dunia. Kini, alih-alih menguji desain pesawat setiap sembilan bulan, mereka dapat membangun model baru setiap hari, memasukkannya ke dalam terowongan angin, dan belajar darinya, melakukan 270 eksperimen dengan risiko dan biaya yang lebih rendah daripada pesaing mereka dalam hal waktu yang dibutuhkan pesaing mereka untuk melakukan satu eksperimen.

Seperti Wright bersaudara, ilmuwan AI perlu menghabiskan waktu mereka dalam eksperimen cepat untuk belajar lebih cepat, tanpa harus menunggu atau membuang waktu saat mereka menyediakan lingkungan, menangani penyerapan dan pembersihan data, atau mengelola daya komputasi. Setiap ilmuwan AI harus dapat dengan cepat memanfaatkan alat yang dibutuhkannya, kapan pun dibutuhkan. Memastikan bahwa setiap pengalaman pengguna ini sesederhana dan seefisien mungkin memerlukan pertimbangan yang matang mengenai komponen apa yang akan disertakan dalam sebuah platform.

Platform AI yang fleksibel memungkinkan pembangunan beberapa aplikasi dalam kerangka kerja teknis yang sama. Fleksibilitas ini memungkinkan produk dan aplikasi masa depan untuk dikembangkan dan diterapkan lebih cepat dan hemat biaya dibandingkan jika dibangun sebagai aplikasi mandiri. Misalnya, di bank, platform AI dapat mendukung semua fungsi di berbagai divisi dan memungkinkan penggunaan komponen bersama. Platform ini dapat digunakan untuk mengembangkan dan menerapkan model untuk anti pencucian uang (AML), personalisasi dalam layanan pelanggan, dan persetujuan kredit. Model AML terintegrasi ke dalam alur kerja pencegahan kejahatan keuangan dalam aplikasi AML; model personalisasi terintegrasi ke dalam situs web dan aplikasi seluler bank; dan model kredit untuk memprediksi risiko gagal bayar pinjaman terintegrasi ke dalam alur kerja kredit dalam aplikasi kredit. Seiring waktu, platform yang dibangun dengan baik bahkan dapat memungkinkan perusahaan untuk memperluas jaringan AI-nya hingga mencakup pemasok, vendor, dan mitra, sehingga mereka juga dapat membangun platform tersebut. Perusahaan dapat memperoleh banyak manfaat dari perluasan jaringan AI ini.

Platform AI harus terintegrasi dengan baik ke dalam perusahaan. Platform ini harus mendukung analis, pengembang produk, dan pemangku kepentingan lainnya, serta model AI yang dibangun sendiri dan model yang dikembangkan secara komersial yang telah dibeli. Platform ini juga harus memungkinkan tim penerapan untuk mengakses dan mengonfigurasi

data, model AI, dan sumber daya komputasi dengan mudah, efisien, dan andal. Untuk melakukan hal ini, platform perlu menyertakan alat otomatis yang mudah digunakan dan meminimalkan tugas berulang.

Dengan pesatnya perkembangan di bidang ini, kapabilitas dan layanan baru terus tersedia, termasuk algoritma baru, kode sumber terbuka, komponen dan solusi dari perusahaan rintisan, serta kerangka kerja AI baru. Di saat yang sama, sulit, bahkan mustahil, untuk memprediksi kebutuhan AI perusahaan mana pun di masa depan. Oleh karena itu, platform AI yang ideal mampu berevolusi. Platform ini, yang tidak terbatas pada satu jenis atau solusi AI saja, memungkinkan penyediaan lingkungan komputasi berkinerja tinggi yang cepat untuk mendukung berbagai jenis AI. Singkatnya, platform ini mengubah AI dari serangkaian solusi titik terbatas menjadi kapabilitas perusahaan yang dapat ditingkatkan seiring waktu untuk memenuhi kebutuhan bisnis perusahaan yang terus berkembang. Ketika sebuah platform dapat melakukan hal ini, platform tersebut dikatakan memiliki arsitektur evolusioner.

Untuk membantu evolusi ini, platform AI harus memiliki portal tempat pengguna dapat terhubung dan berbagi informasi tentang apa yang mereka lakukan, serta alat dan fitur apa yang mereka gunakan atau ketahui yang mungkin bermanfaat bagi rekan kerja. Portal harus menjadi tempat bagi ilmuwan AI untuk meminta agar perangkat lunak sumber terbuka atau komersial baru disertakan dalam platform, atau tempat pengguna dapat memperoleh ide untuk kasus penggunaan baru. Portal juga harus menjadi tempat alur kerja tata kelola atau pengadaan dimulai. Portal memfasilitasi kolaborasi yang lebih cepat, lebih efisien, dan lebih efektif antara ilmuwan dan insinyur AI, memungkinkan distribusi pekerjaan yang lebih merata dan memungkinkan penyelesaian penelitian dan eksperimen secara efisien dan tepat waktu.

Ada yang mungkin merasa bahwa mengatasi kompleksitas membangun platform AI sangat mirip dengan membangun pesawat terbang saat sedang terbang. Mereka khawatir bahwa teknologi baru muncul lebih cepat daripada siklus anggaran mereka. Ketika seorang eksekutif tidak yakin apa yang akan ia butuhkan di masa depan atau solusi apa yang memenuhi kebutuhan tersebut, bagaimana ia dapat memutuskan di mana harus berinvestasi? Jawabannya adalah berfokus bukan pada teknologi tetapi pada keseluruhan arsitektur. Memanfaatkan arsitektur evolusioner membuatnya lebih cepat dan lebih murah untuk mengganti teknologi baik saat menjadi usang atau saat opsi yang lebih baik muncul. Organisasi yang paling sukses di bidang AI adalah mereka yang meluangkan waktu untuk membangun platform AI perusahaan yang tepat bagi bisnis mereka, yang memungkinkan mereka memberikan nilai lebih dengan lebih cepat tidak hanya hari ini, tetapi juga seiring munculnya peluang baru di masa depan. Investasi yang dilakukan perusahaan dalam platform AI seharusnya memadai untuk mencapai semua ini.

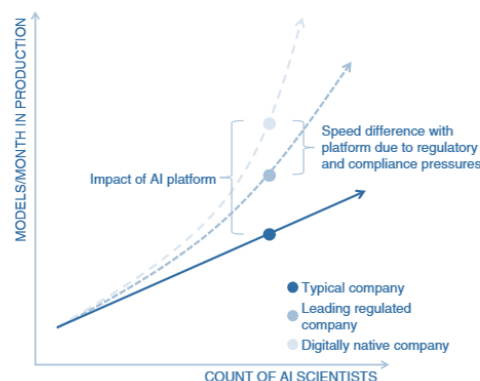
Platform AI yang baik memungkinkan bisnis meningkatkan produktivitas dengan cepat. Dengan menggunakan indikator kinerja utama (KPI) dan membandingkan hasil sebelum dan sesudah penerapan platform AI yang tangguh, perusahaan dapat dengan mudah melihat manfaat pendekatan platform. Beberapa perusahaan menyaksikan tim AI yang sama mengerjakan proyek hingga lima kali lebih banyak per tahun pada platform yang dirancang dengan baik dibandingkan sebelum mereka memilikinya. Perusahaan lain meningkatkan

kecepatan membangun aplikasi baru pada platform AI mereka dari rata-rata delapan bulan per proyek menjadi delapan minggu dengan ukuran tim yang sama. Beberapa perusahaan meningkatkan skala eksperimen hingga 30 kali lebih banyak dalam jangka waktu yang sama menggunakan platform, dibandingkan dengan yang mereka lakukan sebelumnya. Platform AI juga memungkinkan pembentukan tim AI yang besar tanpa terhambat. Dalam melakukan semua ini, platform membantu mengatasi kesenjangan keterampilan, berfungsi sebagai titik fokus untuk orientasi talenta baru, dan mempercepat peningkatan anggota tim baru serta mendistribusikan pekerjaan secara lebih efektif di antara mereka. Platform ini juga dapat membantu mendukung praktik terbaik di seluruh tim ilmuwan AI.

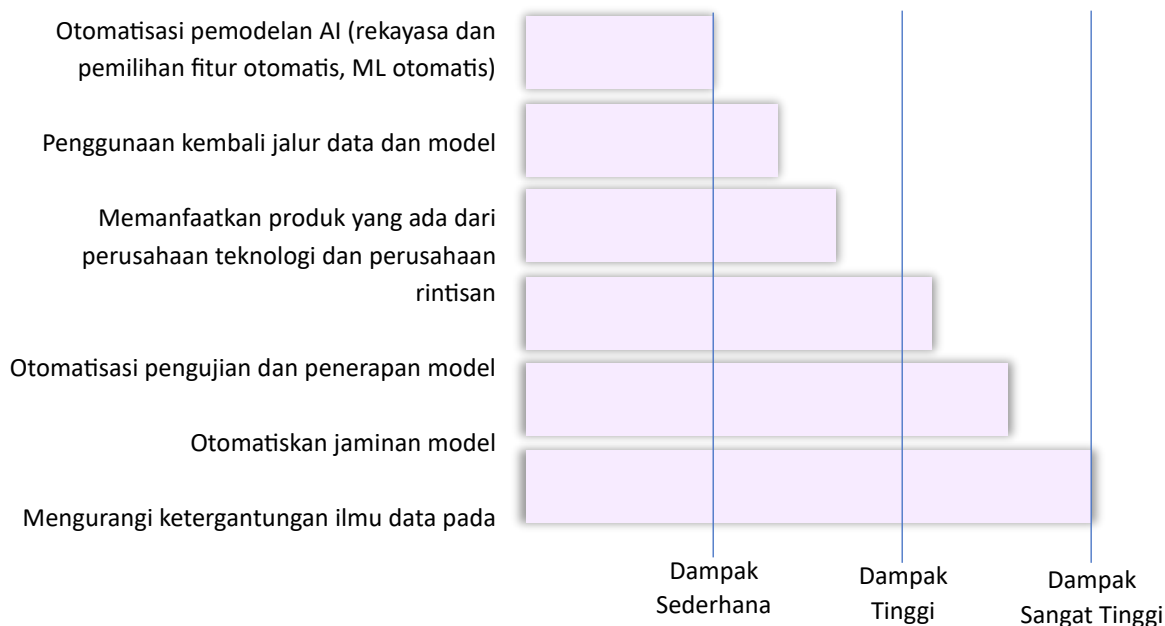
Banyak proyek AI yang terbengkalai karena penyelesaiannya terlalu lama. Seringkali, bagian paling lambat dari proses ini adalah penyediaan lingkungan data dan lingkungan komputasi berkinerja tinggi oleh tim infrastruktur TI. Penundaan lain yang dapat menyebabkan terbengkalainya proyek dapat berasal dari proses tata kelola dan jaminan risiko model yang bersifat ad hoc atau manual yang membutuhkan waktu terlalu lama untuk dieksekusi dan disetujui. Penundaan juga dapat disebabkan oleh kurangnya ketelitian, struktur, dan otomatisasi dalam penerapan dan hosting model produksi.

Gambar 9.1 menceritakan sebagian kisah tentang apa yang dapat dilakukan oleh platform AI yang baik. Perusahaan yang berada pada tahap kematangan platform AI yang rendah berada pada garis lurus; perusahaan-perusahaan ini harus terus menambah ilmuwan data untuk memproduksi semakin banyak model AI. Perusahaan yang berada pada garis putus-putus telah menerapkan platform AI yang skalabel. Bahkan bisnis yang harus menggunakan tata kelola risiko model yang ketat karena tunduk pada tekanan regulasi yang tinggi pun berada pada garis putus-putus pendek, bukan garis lurus, jika mereka memiliki arsitektur evolusioner yang tepat.

Gambar 9.2 menunjukkan asal mula kecepatan dan skala platform AI yang baik, serta dampak relatifnya. Manfaat diperoleh dari mengintegrasikan model atau komponen yang ada ke dalam platform dan membangun komponen yang dapat digunakan kembali melalui setiap proyek; komponen-komponen ini dapat ditambahkan ke platform dan digunakan kembali pada proyek-proyek mendatang untuk mempersingkat jangka waktunya.



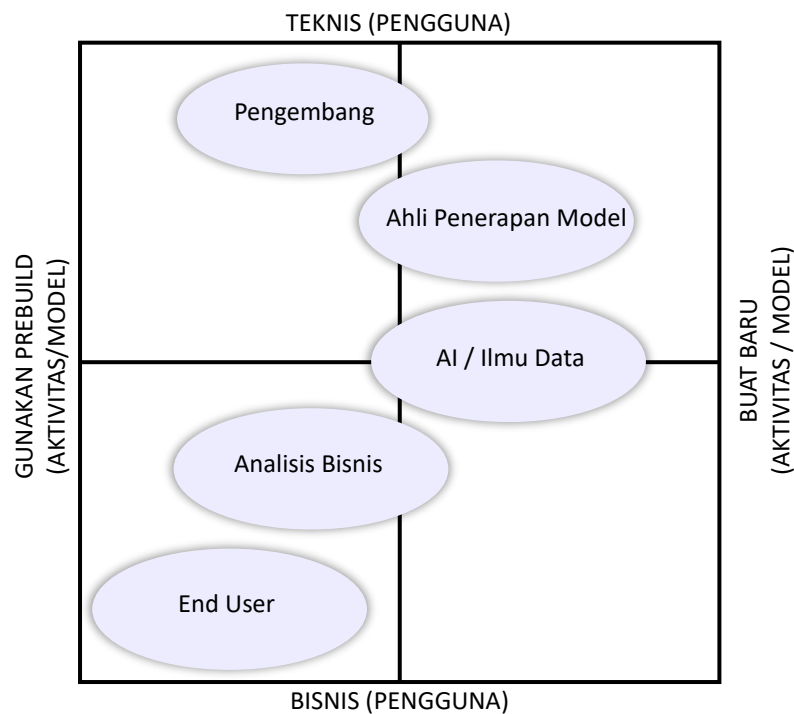
**Gambar 9.1** Dampak penggunaan platform AI.



**Gambar 9.2** Ringkasan manfaat penggunaan platform AI.

Ada juga manfaat mengotomatiskan pekerjaan ilmuwan AI dengan mengintegrasikan alat-alat inkremental seperti alat untuk menghasilkan dan menilai fitur secara otomatis, menghasilkan model secara otomatis menggunakan alat seperti AutoML, serta alur kerja sintesis data dan pelabelan data. Ilmuwan AI harus dapat menghabiskan waktu mereka dalam eksperimen cepat tanpa harus berkebutuhan dengan penyediaan lingkungan, penyerapan dan pembersihan data, serta daya komputasi yang terbatas, sembari berupaya untuk dengan mudah dan cepat mengimplementasikan model mereka yang paling sukses. Karyawan harus dapat dengan cepat memanfaatkan alat atau kombinasi alat yang mereka butuhkan.

Pada akhirnya, kebutuhan "pengguna" harus menentukan jenis platform AI yang akan dibangun, yang membuka jalan bagi organisasi untuk menghadirkan layanan dan produk cerdas, baik saat ini maupun di masa mendatang. Dan yang dimaksud dengan pengguna bukan hanya ilmuwan AI atau bahkan karyawan perusahaan. Yang kami maksud adalah siapa pun yang berinteraksi dengan platform (lihat Gambar 9.3), mulai dari ilmuwan AI hingga departemen TI, insinyur perangkat lunak, ahli strategi, pelanggan, dan lainnya. Pengguna teknis meliputi ilmuwan AI yang menciptakan model baru, insinyur DevOps yang menerapkan model, dan pengembang yang mengintegrasikan model ke dalam aplikasi lain. Pengguna bisnis mencakup analis bisnis yang menggunakan model yang ada untuk analisis, dan pengguna akhir yang mengoperasikan bisnis dan menggunakan keluaran model. Masing-masing akan memanfaatkan platform dengan cara yang berbeda, dan platform harus dapat berinteraksi dengan mereka semua dalam konteksnya, bertukar informasi secara optimal.



**Gambar 9.3** Jenis pengguna platform AI (sumbu vertikal) dan bagaimana mereka berinteraksi dengan model AI (sumbu horizontal).

Pengguna tidak perlu membangun alur data baru secara manual, menyediakan infrastruktur baru, atau mengadaptasi model secara manual untuk sistem produksi; platform pembelajaran mesin menyederhanakan lingkungan pengembangan dan produksi, menyederhanakan pengelolaan infrastruktur untuk swalayan, dan meningkatkan atau menurunkan skala sumber daya bila diperlukan. Interaksi harus intuitif, dengan pelatihan pengguna minimal. Beberapa pengguna mungkin lebih menyukai respons visual yang imersif, atau berinteraksi dengan sistem melalui perangkat seluler atau suara mereka; fitur-fitur ini harus tersedia bila diperlukan. Alat penting lainnya bagi pengguna mencakup mekanisme untuk mengelola penerapan berbagai model AI dalam kontainer di seluruh platform dan dashboard yang memantau kinerja platform.

### 9.3 BEBERAPA PERTIMBANGAN PENTING

Sebelum membangun platform AI, seorang arsitek perusahaan harus berfokus pada beberapa pertimbangan penting.

#### ***Haruskah Sistem Berbasis Cloud, Berbasis di Tempat (On-site) di Organisasi, atau Gabungan Keduanya?***

Ada pro dan kontra untuk masing-masing pilihan. Cloud unggul dalam banyak hal. Penggunaannya membuat bisnis memasarkan lebih cepat: biaya startup dan pemeliharaan lebih sedikit dengan menggunakan cloud, dan pemulihan bencana dapat lebih cepat dan lebih murah, mengingat jumlah pusat data cloud dan keragaman geografisnya. Mengenai produktivitas, cloud memungkinkan perusahaan untuk mengalihdayakan beberapa fungsi TI kepada pakar khusus yang mungkin sulit untuk dipekerjakan, dilatih, dan dipertahankan oleh

perusahaan. Penggunaan cloud berarti menghindari tugas mengoordinasikan berbagai tim di tempat serta proses manajemen yang berpotensi rumit dan birokratis. Penggunaan cloud juga berarti tidak perlu membangun dan mengelola infrastruktur sebelum mulai mengembangkan model yang menciptakan nilai. Di cloud, interaksi perangkat lunak atau operator yang sederhana dapat memberikan perusahaan apa yang dibutuhkannya; di lokasi (on-premise), tim mungkin harus berurusan dengan proses yang melibatkan serangkaian persetujuan dan penundaan. Cloud juga memungkinkan skalabilitas, elastisitas, dan fleksibilitas yang luar biasa. Karena fitur-fitur ini, bahkan orang yang membangun infrastruktur internal pun merencanakan migrasi cloud ketika sudah berskala besar.

Namun, ada beberapa area di mana cloud mungkin kurang memadai. Di cloud, perusahaan hanya membayar untuk apa yang mereka gunakan, dan biaya lisensi perangkat lunak sudah termasuk. Namun, perusahaan-perusahaan tersebut mungkin membayar lebih untuk mendapatkan manfaat seperti biaya yang lebih rendah dan pemulihan bencana yang cepat. Jika mereka sepenuhnya memanfaatkan pusat data di tempat dan jumlah perangkat lunak, pemrosesan, dan kemampuan jaringan mereka secara optimal, mereka mungkin akan membayar lebih rendah daripada jika mereka melakukan outsourcing ke cloud. Keamanan dan privasi data adalah masalah lain yang perlu dipertimbangkan. Banyak organisasi percaya bahwa data mereka lebih aman di tempat mereka sendiri daripada di cloud. Seorang eksekutif mungkin mendapati dirinya terlibat dalam perdebatan panjang dengan arsitek perusahaan dan kepala petugas keamanan informasi mengenai hal ini, dan untuk bisnis yang diatur, kepatuhan data harus direncanakan, diuji, dan disetujui sebelum cloud dapat digunakan.

Latensi juga perlu diperhatikan. Meskipun bekerja di cloud mudah dan cepat, informasi tidak dapat bergerak lebih cepat dari kecepatan cahaya. Jika sebuah perusahaan menjalankan fungsi seperti perdagangan saham frekuensi tinggi, waktu yang dibutuhkan data untuk berpindah dari model ke pusat data cloud dan kembali dapat menimbulkan masalah besar. Para pedagang menempatkan platform mereka di dekat hub jaringan agar tidak membuang waktu karena alasan ini. Fleksibilitas, kinerja, dan skalabilitas yang dibutuhkan dalam platform AI modern dapat berarti solusi berbasis cloud adalah yang paling sesuai untuk bisnis tertentu. Ada juga opsi hibrida, yang menggunakan penyimpanan lokal untuk area perusahaan yang memiliki keuntungan berada di lokasi misalnya, dalam situasi yang melibatkan keamanan data atau latensi rendah. Dengan perangkat lunak terkelola seperti OpenShift dari RedHat dan Anthos dari Google, pengelolaan lingkungan multicloud dan cloud serta lokal menjadi lebih mudah.

### ***Haruskah Bisnis Menyimpan Datanya di Gudang Data, Danau Data, atau Pasar Data?***

Baik gudang data maupun danau data terdiri dari prosesor dan perangkat lunak yang menjalankannya, serta perangkat penyimpanan dan jaringan yang menghubungkannya. Perbedaan di antara keduanya terletak pada struktur data yang dapat disimpan masing-masing. Gudang data menyimpan data yang sangat terstruktur dalam format yang telah ditentukan sebelumnya untuk mengoptimalkannya bagi pelaporan standar. Data terstruktur memiliki bidang-bidang yang terdefinisi dengan baik, seperti nama, alamat, nomor telepon, dan sebagainya, serta baris data untuk bidang-bidang ini. Data tidak terstruktur adalah teks

yang mungkin diperoleh dari sumber seperti email atau tweet, atau gambar dan video seperti sinar-X atau pemindaian MRI.

Gudang data bekerja paling baik ketika kebutuhan informasi dan data perusahaan berkembang secara perlahan. Namun, saat ini hanya sedikit bisnis yang tidak berada dalam kondisi perubahan yang konstan. Mengingat iklim bisnis yang terus berubah, lanskap teknis yang berkembang pesat, dan akumulasi data yang tak henti-hentinya, gudang data saja mungkin tidak dapat memenuhi kebutuhan perusahaan. Gudang data tradisional mengikat data yang dikumpulkan ke dalam kategori yang sangat terstruktur dan kaku sebelumnya; danau data menyimpan data mentah dalam format aslinya: misalnya, berkas teks. Dengan kata lain, danau data memerlukan lebih sedikit transformasi data sebelum disimpan (meskipun akan memerlukan beberapa transformasi selama proses pengambilan). Danau data dapat mengakomodasi jenis data baru seiring dengan perkembangan kebutuhan organisasi, yang pada dasarnya menciptakan kumpulan informasi yang dapat diakses dan dikueri kapan saja. Danau data juga dapat mengagregasi dan menstandarisasi berbagai macam data serta memungkinkan tampilan data longitudinal tunggal.

Pasar data adalah danau data dengan lapisan yang terorganisir dan mudah diakses untuk membuat penemuan data dan penggunaan bisnis lebih alami bagi tim non-TI. Hal ini sangat penting karena memungkinkan pebisnis dan ilmuwan AI untuk mulai berbicara dalam bahasa data yang sama. Pasar data memudahkan pengguna bisnis untuk menemukan dan memahami data yang mereka butuhkan, memungkinkan mereka memperoleh data tanpa melibatkan pemrograman atau TI. Pasar data juga memiliki sistem dan alat yang membatasi akses data hanya kepada pengguna yang berwenang. Pasar data membantu mengubah data dari aset teknis menjadi aset bisnis, mendorong perubahan budaya perusahaan menuju literasi data yang lebih besar dan keputusan yang lebih berbasis data. Perubahan budaya ini juga membantu meningkatkan kualitas data berkelanjutan karena mendorong lebih banyak penggunaan data. Pengguna bisnis yang menggunakan pasar data menjadi lebih nyaman meminta dan menggunakan model yang lebih canggih dari waktu ke waktu, sehingga menciptakan adopsi AI yang lebih besar.

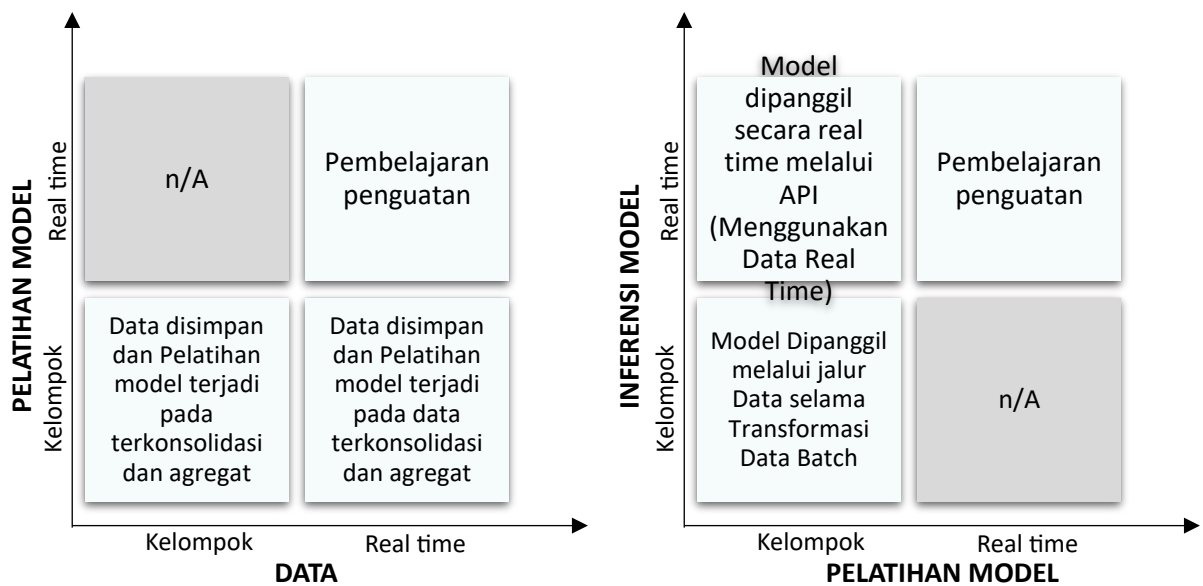
Karena perbedaan ini, arsitek perusahaan beralih ke penyimpanan danau data dan penggunaan pasar data, meskipun sebagian besar perusahaan saat ini juga menggunakan gudang data jika kebutuhan mengharuskannya. Keputusan untuk menggunakan salah satu atau yang lain didasarkan pada kasus penggunaan perusahaan. Bagi perusahaan yang hanya berfokus pada standarisasi, kontrol, dan pelaporan, gudang data mungkin sudah memadai. Bagi perusahaan yang menginginkan fleksibilitas, berbagi, dan beragam jenis penggunaan, danau data dan pasar data menjadi sangat penting.

### ***Haruskah Bisnis Menggunakan Pemrosesan Batch atau Real-Time?***

Ada tiga cara untuk melakukan pemrosesan bagi bisnis: batch, real-time (atau streaming), atau gabungan keduanya. Pengumpulan data bisa batch atau streaming. Menyebut suatu model sebagai batch atau real-time dapat menyebabkan kebingungan jika kita tidak menentukan jenis pemrosesannya apakah yang kita maksud adalah pelatihan model atau inferensi model. Untuk model AI, masing-masing aktivitas ini bisa batch atau real-time (lihat

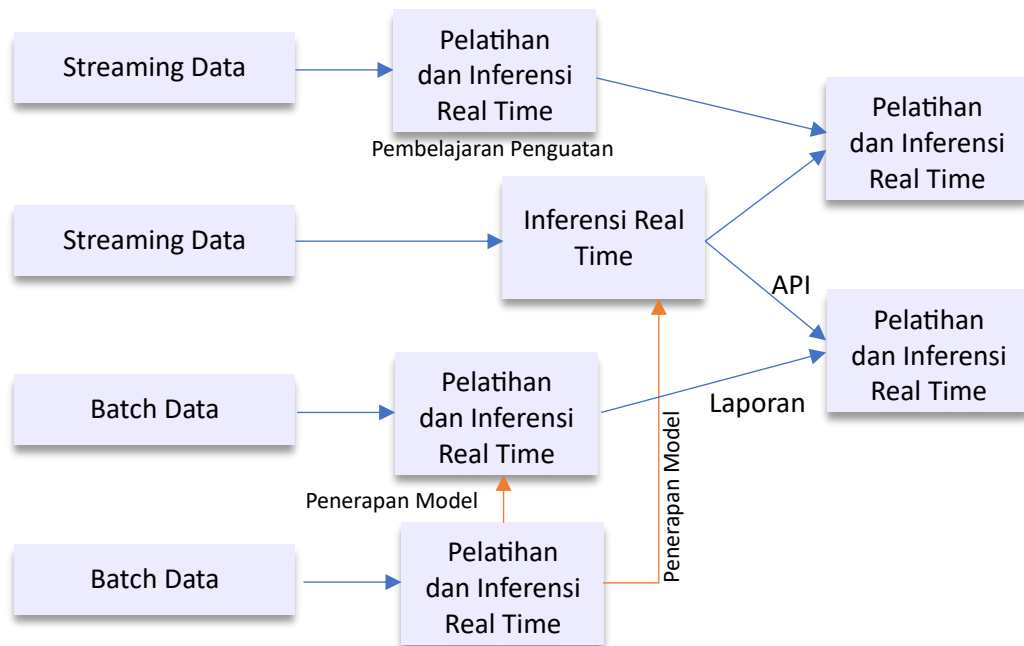
Gambar 9.4). Pada sebagian besar aplikasi, pelatihan terjadi dalam mode batch, dan inferensi terjadi dalam mode batch atau real-time. Dengan semakin umum digunakannya pembelajaran penguatan dan metode serupa, pelatihan real-time semakin sering digunakan. Seiring dengan peningkatan model AI, banyak keputusan diambil secara otonom tanpa campur tangan manusia. Beberapa orang menyebut pengambilan keputusan otonom ini sebagai real-time. Keperluan pemrosesan real-time, dan jika ya, untuk jenis apa, akan bergantung pada kasus penggunaan yang ingin dipecahkan oleh bisnis.

Pemrosesan batch, juga disebut inferensi batch, mengacu pada input yang diakumulasikan selama jangka waktu tertentu dan kemudian diproses secara berkala (misalnya, setiap jam atau setiap hari). Prediksi yang dihasilkan disimpan dalam basis data yang dapat diakses oleh pengembang atau pengguna akhir. Inferensi batch sederhana, efisien, dan hemat biaya, sehingga umum digunakan. Inferensi batch bekerja dengan baik dengan volume data yang besar dan dalam situasi di mana pembaruan hasil model AI hanya dapat diterima secara berkala. Misalnya, bisnis yang melakukan valuasi rumah berdasarkan ukuran, jumlah kamar tidur, dan lokasi tidak perlu memperbarui nilai setiap rumah sepanjang hari.



**Gambar 9.4** Batch versus waktu nyata untuk data, pelatihan model, dan inferensi model.

Namun, inferensi batch mungkin tidak memuaskan perusahaan jika membutuhkan pengambilan keputusan kritis waktu atau aplikasi interaktif. Gambar 9.5 menunjukkan berbagai skenario untuk pemrosesan batch dan waktu nyata. Inferensi waktu nyata diperlukan saat data baru tersedia dan pengguna perlu melihatnya saat itu juga (kendaraan otonom, misalnya, memerlukan inferensi waktu nyata). Namun, lebih umum bagi bisnis untuk menggunakan inferensi waktu nyata. Misalnya, mesin rekomendasi yang merekomendasikan item berdasarkan keranjang belanja online seseorang harus mengambil data waktu nyata tentang keranjang belanja dan menjalankan model AI untuk mendapatkan rekomendasi.



**Gambar 9.5** menunjukkan berbagai pola data batch atau streaming, pelatihan model, inferensi model, dan penggunaan.

(Jika rekomendasi hanya didasarkan pada item yang paling populer, model AI dapat dijalankan dalam mode batch dan hasilnya disimpan.) Inferensi dalam waktu nyata mendekati memberikan hasil sesuai permintaan menggunakan data terbaru, biasanya dipanggil melalui panggilan API. Penggunaan API memungkinkan pengiriman output dari model AI secara hampir real-time atau sesuai permintaan ke program lain, yang kemudian diintegrasikan ke dalam alur kerja, layar, atau laporan aplikasi kedua.

Seiring dengan peningkatan model AI, banyak keputusan ditindaklanjuti secara otonom tanpa campur tangan manusia. Beberapa orang menyebutnya sebagai pengambilan keputusan real-time, atau streaming. Diperlukan atau tidaknya pengambilan keputusan real-time bergantung pada kasus penggunaan; sebagian besar sistem yang keputusannya telah diotomatisasi menggunakan inferensi real-time semacam ini.

Dalam bentuk pemrosesan hibrida, bisnis dapat melakukan inferensi batch hasil spesifik dan menggunakan inferensi near-real-time untuk hasil lainnya. Misalnya, perhatikan contoh nasabah bank yang mengajukan pinjaman. Proses batch dapat mencakup penggunaan OCR dan NLP untuk membaca dokumen dan menangkap informasi terstruktur. Dan proses real-time mungkin terjadi ketika semua informasi tersedia, yang memicu proses pengambilan keputusan pinjaman. Jika nasabah menelepon dengan informasi terkini, petugas kredit di bank yang memproses permintaan pinjaman dapat memperbarui informasi tersebut dan menjalankan evaluasi ulang segera atas penilaian risiko kredit dari model AI melalui API.

**Haruskah Bisnis Menggunakan Arsitektur Monolitik atau Mikroservis?**

Arsitektur monolitik terdiri dari satu struktur tunggal: sebuah aplikasi perangkat lunak di mana berbagai komponen digabungkan menjadi satu program yang berjalan pada satu platform. Arsitektur monolitik secara historis merupakan cara sistem perangkat lunak

perusahaan direkayasa; arsitektur ini cenderung berasal dari satu vendor dan memecahkan masalah tertentu. Arsitektur mikroservis adalah sistem yang terdiri dari banyak mikroservis yang terhubung secara longgar. Layanan-layanan ini berbeda, terisolasi, dan dapat digunakan kembali, dan mereka bekerja sama untuk menyediakan satu kapabilitas bisnis. Saat ini, semakin umum untuk menggabungkan pedoman khusus yang diambil dari dunia mikroservis dan arsitektur berbasis peristiwa ke dalam suatu sistem, karena arsitektur ini dapat dimodifikasi tergantung pada kebutuhan perusahaan, sehingga memudahkan pembuatan atau penghentian satu mikroservis tanpa memengaruhi mikroservis lainnya. Fitur ini, yang dikenal sebagai isolasi, memberikan kemampuan untuk mengembangkan platform secara bertahap seiring layanan baru dikembangkan atau tersedia, tanpa perlu merancang ulang arsitektur dari awal. Ini mengurangi waktu pemasaran dan waktu untuk mendapatkan nilai serta memungkinkan struktur evolusioner platform AI.

Perangkat lunak yang berorientasi pada layanan mikro lebih mudah untuk diskalakan, digunakan kembali, dan diintegrasikan baik di dalam maupun di seluruh perusahaan. Dalam sistem AI, di mana fleksibilitas dan adaptasi cepat sangat penting, ini memberikan keunggulan tersendiri dibandingkan arsitektur monolitik tradisional. Misalnya, perusahaan yang menggunakan layanan mikro dari Google untuk mengubah ucapan menjadi teks dalam asisten virtual akan mudah beralih dari satu layanan mikro ke layanan mikro lainnya jika Microsoft menemukan solusi yang lebih baik. Arsitektur AI harus mengadopsi praktik terbaik ini, yang memungkinkan solusi menjadi lebih sederhana, lebih fleksibel, dan lebih mudah diintegrasikan dengan bagian lain dari bisnis.

Banyak dari hampir 2.000 perusahaan teknologi terkait AI saat ini menyediakan layanan mikro. Namun, jika perusahaan masih menggunakan arsitektur monolitik dan ingin beralih, tersedia paket perangkat lunak yang membuat lapisan layanan mikro di sekitar aplikasi lama. Lapisan layanan mikro ini dapat terintegrasi langsung ke dalam aplikasi lama atau dapat terintegrasi menggunakan perangkat lunak RPA sebagai perantara. Pendekatan ini memungkinkan arsitektur perusahaan untuk sepenuhnya berorientasi pada layanan mikro guna mendukung inisiatif AI di dalam perusahaan tanpa harus mengganti setiap aplikasi lama. Namun, pada akhirnya, aplikasi lama perlu dimodernisasi.

#### **Platform yang Mendorong Uber: Michelangelo**

Michelangelo, platform pembelajaran mesin internal Uber, “memungkinkan tim internal untuk membangun, menerapkan, dan mengoperasikan solusi pembelajaran mesin dengan lancar dalam skala Uber. Platform ini dirancang untuk mencakup alur kerja ML menyeluruh: mengelola data, melatih, mengevaluasi, dan menerapkan model, membuat prediksi, dan memantau prediksi. Sistem ini juga mendukung model ML tradisional, peramalan deret waktu, dan pembelajaran mendalam.”<sup>1</sup> Uber mengembangkan Michelangelo pada tahun 2015, dan telah melayani perusahaan tersebut sejak tahun 2017. Michelangelo dibangun karena upaya AI sebelumnya tidak memadai untuk membantu bisnis sebesar dan sebesar Uber. Tim-tim teknik yang terpisah telah menciptakan sistem sekali pakai yang

terbatas berdasarkan sebagian besar perangkat lunak sumber terbuka. Tidak ada tempat standar untuk menyimpan data, tidak ada tempat standar untuk menyimpan hasil eksperimen pelatihan, tidak ada cara untuk membandingkannya satu sama lain, dan tidak ada kemampuan untuk melatih model yang lebih besar dari yang muat di mesin desktop. Yang terpenting, tidak ada jalur standar untuk menerapkan model ke tahap produksi. Michelangelo mengubah semua itu, dan saat ini model tersebut digunakan di seluruh perusahaan dalam berbagai kasus penggunaan.

#### 9.4 ARSITEKTUR PLATFORM AI

Komponen fungsional utama arsitektur platform AI harus selaras dengan siklus hidup AI. Komponen-komponen ini dapat dipecah menjadi empat area tingkat tinggi: pengelola data, pembuat model, penggerak inferensi, dan pengelola kinerja.

##### ***Pengendali Data***

Selain staf bisnis, data merupakan aset paling berharga. Komponen pengendali data memproses data, mengkurasinya, dan membuatnya dapat diakses melalui layanan mandiri dari seluruh perusahaan, sehingga tersedia untuk konsumsi pemodelan AI. Ilmuwan AI dapat merevisi kode saat membuat model AI, tetapi mereka hanya membuat perubahan pada salinan data yang telah dikurasi, sehingga data mentah tidak tersentuh. Ilmuwan AI tidak perlu menyediakan data setiap kali mereka perlu membuat model; jauh lebih cepat dan efisien untuk menggabungkan dan membuat kumpulan data yang dapat diakses serta menggunakannya berulang kali, sehingga siapa pun yang memiliki akses dapat membagikannya. Ini berarti tidak perlu menyiapkan kumpulan data lebih dari sekali, dan memungkinkan konsistensi dalam penggunaan data. Selain itu, komponen pengendali data memungkinkan tim untuk bekerja sama secara lebih efisien serta menghilangkan pengulangan beberapa aktivitas yang mahal, seperti menyalin dan mengekstrak data serta mengelola kualitas data.

Sebuah platform harus memungkinkan penggunaan bahasa khusus bisnis untuk pemilihan dan transformasi fitur. Bahasa ini memungkinkan penggunaan nomenklatur yang familiar bagi pengguna data, alih-alih menyajikan terminologi teknis yang membingungkan dan tidak mereka kenali. Kosakata umumnya diimplementasikan melalui katalog data dan alat silsilah data menggunakan terminologi bisnis. Sebuah platform harus memungkinkan penambahan tipe data terstruktur dan tak terstruktur baru tanpa jeda waktu dan biaya standarisasi data atau penyediaan perangkat keras dan perangkat lunak yang mahal. Platform tersebut harus secara akurat mengenali siapa yang memiliki hak akses. Selain itu, ketika sebuah tim telah mengerjakan rekayasa fitur suatu dataset untuk tujuan pemodelan, dataset tersebut harus ditambahkan ke pasar fitur; dataset tersebut merupakan kumpulan fitur rekayasa yang dikurasi dan dapat digunakan kembali di berbagai model.

##### ***Pembuat Model***

Komponen pembuat model dari arsitektur platform memungkinkan pembuatan model AI baru dan eksperimen yang lebih cepat sekaligus menghindari duplikasi upaya, mengotomatiskan tugas-tugas bernilai rendah, dan meningkatkan reproduktifitas serta

penggunaan ulang semua pekerjaan. Lingkungan ini memungkinkan tim untuk secara kolaboratif memelihara, mengkloning, menggunakan kembali, dan memperluas model AI serta mendukung kontrol versi. Komponen ini berisi banyak komponen teknologi dan mendukung berbagai jenis algoritma yang digunakan ilmuwan AI untuk mengembangkan model AI. Komponen ini harus mampu tumbuh dan berkembang layaknya bisnis, memberikan organisasi cara yang terstruktur namun fleksibel untuk menciptakan solusi berbasis AI saat ini dan dalam jangka panjang.

Komponen ini melacak setiap model dengan mudah dan efisien, menjawab pertanyaan seperti: Dataset mana yang digunakan? Algoritma apa? Hiperparameter apa? Hasil apa yang diperoleh? Komponen ini memiliki kemampuan untuk mereproduksi hasil yang identik ketika model yang sama dijalankan ulang dengan data yang sama. Hal ini juga memungkinkan ilmuwan AI untuk kembali dan mengulangi eksperimen di beberapa titik di masa lalu dengan data baru, serta memahami mengapa model tertentu bekerja paling baik.

Pembuat model memungkinkan setiap fungsi atau bisnis dalam suatu organisasi untuk bekerja secara paralel, memanfaatkan alat dan plug-in yang mereka butuhkan. Hal ini memungkinkan standarisasi alur kerja dan alat di seluruh tim melalui sistem ujung ke ujung yang memungkinkan pengguna untuk membangun dan mengoperasikan aplikasi pembelajaran mesin dengan cepat. Standarisasi memungkinkan mereka untuk berbagi data, yang sangat penting karena mereka kemungkinan akan menjalankan model yang berbeda pada data tersebut, masing-masing berorientasi pada kasus penggunaan yang berbeda.

Untuk memungkinkan bisnis mengevaluasi model AI-nya, platform ini memiliki kemampuan untuk melacak setiap eksperimen. Idealnya, platform ini secara otomatis menghasilkan fitur optimal, secara otomatis memilih model optimal, dan secara otomatis mengoptimalkan hiperparameter. Platform ini menghasilkan keluaran, terkadang dalam bentuk visual, untuk berbagai macam pengguna, mulai dari eksekutif hingga pengguna berpengalaman hingga ilmuwan AI.

### **Aktivator Inferensi**

Komponen aktivator inferensi adalah tempat platform mendukung propagasi versi model tertentu ke tahap produksi. Tujuannya adalah untuk berpindah dari pengembangan model ke penerapan model secepat dan seefisien mungkin. Penerapan model mencakup kemampuan untuk menerapkan model secara independen atau dengan kode dan data dependen. Komponen ini memungkinkan pengujian paralel atau pengujian A/B untuk berbagai model. Komponen ini melakukan ini secara otomatis di semua tahap siklus hidup model: sandbox, validasi model, uji integrasi sistem (SIT), uji penerimaan pengguna (UAT), staging, dan produksi. Setiap kali model lolos inspeksi di salah satu lingkungan ini, platform akan mendorongnya ke lingkungan berikutnya dengan satu klik.

Banyak hal yang dapat mencegah atau menunda aktivasi model, yaitu, menempatkan model untuk digunakan dalam produksi. Salah satunya adalah kurangnya otomatisasi proses penerapan menggunakan AI DevOps yang tepat. Faktor lainnya termasuk mengelola risiko model, mendapatkan dukungan pengguna, dan mempertimbangkan antara kinerja dan risiko. Sekalipun perusahaan telah meningkatkan jumlah eksperimen dan melakukan banyak

pembuktian konsep, tidak ada nilai yang terealisasi jika tidak menerapkannya ke dalam produksi. Nol dikalikan apa pun tetaplah nol. Sangat penting bagi pengguna akhir untuk mengakses keluaran model, karena wawasan ini hanya berguna ketika pengguna akhir memilikinya, dan terkadang hanya dalam jangka waktu tertentu.

Aktivator inferensi memiliki desain modular, sehingga produk cerdas terbaik yang tersedia untuk kebutuhan perusahaan dapat dipasang atau diganti. Beberapa di antaranya adalah produk niche yang memiliki satu kegunaan spesifik, seperti mesin personalisasi yang menyesuaikan situs web perusahaan untuk setiap pengguna. Lainnya bersifat plug-and-play, seperti model ucapan-ke-teks yang dapat disesuaikan atau model analisis sentimen yang dapat digunakan kembali oleh tim dalam berbagai aplikasi. Arsitektur evolusioner yang fleksibel memungkinkan ilmuwan AI untuk menggabungkan berbagai fungsinya menjadi sistem yang sukses yang dapat menangani keadaan perusahaan yang terus berkembang. Hal ini memungkinkan fleksibilitas untuk mengonfigurasi kapabilitas kognitif yang baru tersedia dan mengganti atau meningkatkan yang sudah ada, baik itu model yang dibangun oleh tim perusahaan atau komponen yang tersedia secara komersial yang dapat dimanfaatkan dalam platform. Komponen-komponen ini mungkin awalnya menjadi bagian penting dari model yang tersedia.

Sebuah platform tidak boleh terlalu terikat dengan sistem TI internal yang ada; dengan cara ini, perubahan yang mahal dan integrasi dengan sistem lama dapat diminimalkan dan biaya implementasi platform kognitif baru terkandung. Penggandengan longgar dapat dikelola melalui lapisan API. Mempertahankan fleksibilitas ini dalam arsitektur memungkinkan peningkatan platform kognitif dari waktu ke waktu. Ini juga akan membantu menghindari API yang diterbitkan berlebihan: antarmuka program aplikasi yang memberi tahu bisnis dan dunia cara berbicara dengan sistem AI bisnis itu.

Beberapa platform menawarkan fungsionalitas dasbor terintegrasi atau menyediakan koneksi ke alat visualisasi, yang lebih dari sekadar menyediakan API untuk hasil model. Hal ini membuat keluaran model jauh lebih mudah diakses untuk dikonsumsi atau diintegrasikan. Namun, pengguna yang masuk ke sistem baru dapat mengganggu alur kerja mereka. Itulah sebabnya beberapa perusahaan memilih menggunakan API untuk mengintegrasikan model ke dalam sistem BI atau aplikasi bisnis mereka.

### **Manajer Kinerja**

Komponen manajer kinerja platform adalah tempat kualitas model dipantau dari waktu ke waktu. Keluaran model AI, seperti prediksi peristiwa, dicatat, begitu pula peristiwa aktual. Kemudian keduanya dibandingkan. Perbedaan antara keluaran model dan aktual harus berada dalam ambang batas kinerja yang sama dengan yang digunakan saat model disetujui untuk diproduksi.

Manajer kinerja memantau penyimpangan fitur, sehingga statistik masukan model tidak lebih buruk daripada saat model dilatih, terutama ketika data aktual membutuhkan waktu untuk dikumpulkan. Penyimpangan fitur dapat diterima dalam beberapa ambang batas, tetapi jika ambang batas penyimpangan atau ambang batas kinerja terlampaui, model perlu dilatih ulang. Platform mungkin juga perlu memiliki alur kerja terpisah untuk berbagai tingkat

risiko bisnis. Manajer kinerja juga mendukung audit prediksi atau rekomendasi yang dibuat oleh model AI. Persyaratan ini jauh lebih ketat dalam lingkungan yang sangat diatur, seperti keuangan atau asuransi. Jika regulator ingin mengaudit model dan melihat semua keputusan yang dibuat dalam produksi selama tiga bulan terakhir, platform perlu mendukung permintaan tersebut.

Penting juga untuk memiliki sistem yang memungkinkan bisnis mengenali, menyoroti, dan membatasi keluaran yang tidak diinginkan dalam suatu model. Keluaran yang tidak diinginkan adalah keluaran yang tidak boleh ditindaklanjuti. Komponen platform ini harus memiliki pembatas, terkadang juga disebut pemutus sirkuit atau mekanisme pengaman agar sistem tahu apa yang harus dilakukan dalam keadaan seperti itu. Pembatas memperingatkan pengguna dan menjalankan mekanisme pengaman, yang secara otomatis mencegah sistem mengambil tindakan tertentu ketika suatu model melampaui batas yang telah ditentukan (yaitu, ketika hasilnya terlihat aneh atau salah). Kemampuan untuk menonaktifkan model dan mempertahankan pemahaman tentang dampak hilir dari tindakan tersebut sangatlah penting. Mekanisme pengamanan juga harus mencakup penggantian manual, model failover, atau kombinasi keduanya. Mekanisme ini harus mencakup prosedur dan eskalasi untuk menangani kesalahan, serta komunikasi apa pun yang diperlukan, baik secara internal maupun eksternal.

#### **Memilih Kerangka Kerja AI**

Kerangka kerja AI adalah kumpulan perangkat lunak dan algoritma pra-paket yang bekerja sebagai bagian dari platform yang dapat Anda gunakan untuk membangun model AI terlatih. Memilih kerangka kerja AI yang tepat merupakan bagian penting dalam menciptakan platform AI yang sesuai, dan lebih jauh lagi, perusahaan AI yang sukses. Setiap kerangka kerja memiliki kekuatan dan kelemahan dan dapat berubah seiring waktu. Misalnya, mungkin saja perangkat yang awalnya dipilih kini sudah usang karena munculnya kerangka kerja yang lebih baru. Dalam hal ini, perusahaan ingin dapat mengubah kerangka kerja tersebut secara keseluruhan atau menambahkan satu atau lebih untuk melengkapi kerangka kerja yang sudah berjalan. Atau, kerangka kerja yang telah digunakan mungkin tidak lagi didukung dan sedang dihapuskan. Karena alasan ini, kerangka kerja harus cukup fleksibel untuk mengakomodasi berbagai perangkat AI saat ini dan yang akan datang.

Sebuah kerangka kerja juga harus mendukung bahasa pemrograman yang disukai bisnis minimal, Python dan R, yang saat ini paling umum di lingkungan AI serta cukup mudah digunakan. Sebaiknya pertimbangkan untuk memanfaatkan toolkit umum daripada produk khusus. Manfaatkan API yang sudah ada jika memungkinkan juga bermanfaat, alih-alih membuat model baru yang menduplikasi API yang sudah ada. API yang sudah ada telah teruji di lapangan, dan itu berarti API tersebut cenderung lebih andal. Kerangka kerja populer saat ini antara lain PyTorch dari Facebook, TensorFlow dari Google, CNTK dari Microsoft, MXNET dari Amazon, dan SciKit-learn.

## **BAB 10**

### **MENGELOLA RISIKO MODEL**

Orang mungkin tidak menyadari AI dalam kehidupan sehari-hari mereka, tetapi AI memang ada. Seperti yang kita lihat di Bagian II buku ini, program berbasis pembelajaran mesin kini meninjau banyak aplikasi hipotek. Algoritma AI memilah resume untuk menemukan sekelompok kecil kandidat yang sesuai sebelum wawancara kerja dijadwalkan. Sistem AI mengkurasi konten untuk setiap individu di Facebook. Dan panggilan telepon ke departemen layanan pelanggan penyedia kabel, perusahaan utilitas, dan bank, di antara lembaga lainnya, dijawab oleh sistem pengenalan suara berbasis AI.

Namun, AI yang "tak terlihat" ini dapat membuat dirinya terlihat dengan beberapa cara yang tidak disengaja dan terkadang menjengkelkan. Raksasa ritel Target menggunakan AI untuk memahami apa yang dibeli pembeli dan apa yang harus direkomendasikan kepada mereka, tetapi taktik itu menjadi bumerang ketika Target mengirimkan kupon kepada putri remaja seorang pria yang menampilkan furnitur kamar bayi dan pakaian hamil. Beberapa saat setelah menyerbu ke toko Target di luar Minneapolis dan mengecam manajer karena mendorongnya untuk hamil, ayah gadis itu yang marah menemukan bahwa Target mengetahui sesuatu yang tidak diketahuinya. "Ternyata ada beberapa kegiatan di rumah saya yang belum sepenuhnya saya sadari," ungkapnya dalam sebuah wawancara. "Dia akan melahirkan pada bulan Agustus." Bukan karena prediksi Target tentang pelanggan itu tidak akurat; masalah muncul ketika toko secara otomatis menindaklanjutinya.

Lalu ada Tay, chatbot yang dirilis oleh Microsoft pada Maret 2016. Tay seharusnya memiliki kepribadian seorang remaja berusia 19 tahun yang normal, tetapi hanya beberapa jam setelah dirilis, ia berubah menjadi monster rasis yang gila seks. Tampaknya pengguna Tay telah mengeksploitasi model pembelajaran Tay dan mengajarkannya cara mengatakan hal-hal yang sangat menyinggung. Tay dikatakan telah "menjadi nakal" setelah dirilis ke publik. Ini tentu saja memberikan momen yang dapat diajarkan, tetapi itu adalah momen yang mengkhawatirkan.

Masalah serius lainnya muncul ketika PredPol, sebuah perusahaan yang dibentuk untuk membuat dan memasarkan perangkat lunak kepolisian prediktif yang dirancang untuk memprediksi kegiatan kriminal di lingkungan sekitar, menemukan bahwa algoritmanya memprediksi kejahatan di komunitas yang memiliki proporsi ras minoritas yang tinggi, terlepas dari apakah mereka, pada kenyataannya, merupakan daerah dengan kejahatan tinggi. Tidak seperti contoh Target sebelumnya, prediksi ini tidak tepat. Dalam kasus lain tentang keadilan yang salah, organisasi berita ProPublica menemukan masalah dengan alat penilaian risiko yang banyak digunakan dalam sistem peradilan pidana. Mesin tersebut dirancang untuk memprediksi residivisme (kambuhnya perilaku kriminal) di antara populasi penjara. Estimasi risiko secara keliru menetapkan terdakwa Afrika-Amerika lebih mungkin melakukan kejahatan di masa mendatang daripada terdakwa Kaukasia.

Ada kekhawatiran yang berkembang bahwa algoritma AI menimbulkan bias dan diskriminasi terhadap kelompok yang dilindungi serta "potensi pengkodean diskriminasi dalam keputusan otomatis" dari bias dalam data historis, seperti terhadap minoritas dalam kepolisian, atau terhadap perempuan dalam pekerjaan dan keputusan hipotek. Pelajaran di sini adalah bahwa kecerdasan buatan membutuhkan tata kelola yang cermat untuk memitigasi risiko model, sesuatu yang menjadi semakin penting karena semakin banyak perusahaan yang menggunakan AI.

### **10.1 KETIKA ALGORITMA BERJALAN SALAH**

Pada tahun 2018, Amazon menghentikan beberapa perangkat lunak rekrutmen AI-nya karena menunjukkan bias terhadap perempuan. Sebagaimana dilaporkan oleh Reuters, spesialis pembelajaran mesin Amazon sendiri menyadari bahwa data pelatihan algoritma mereka telah diambil dari pola dalam resume yang dikirimkan selama 10 tahun. Sayangnya, selama periode ini, laki-laki mendominasi industri perangkat lunak, sehingga sebagian besar resume yang digunakan untuk pelatihan berasal dari laki-laki. Contoh ini menunjukkan dua poin penting: sangat penting untuk memilih data pelatihan yang tepat, dan manusia dapat dengan mudah membuat kesalahan dalam melakukannya. Manusia juga dapat menimbulkan masalah dalam kasus algoritma AI itu sendiri, terlepas dari upaya ilmuwan komputer untuk melakukan pekerjaan sebaik mungkin dengan alat yang tersedia bagi mereka. Hal ini dapat mengakibatkan algoritma AI mengalami bias atau masalah lainnya. Ada baiknya mengkaji situasi ini dan bagaimana cara mengatasinya.

Perangkat lunak itu sendiri tidaklah baik atau buruk. Perangkat lunak dapat menjadi buruk karena cara seseorang menggunakannya, misalnya, ketika malware dibuat dengan niat jahat, atau ketika desainnya menciptakan konsekuensi negatif yang tidak diinginkan. Konsekuensi yang tidak diinginkan tidak terlalu menjadi masalah di masa lalu, mengingat persyaratan ditulis dan kemudian dikodekan secara manual, ditinjau, dan diuji. Bahkan saat itu, ada beberapa beberapa dalam proses pengkodean dan beberapa karena celah dalam desain. AI belajar dari data yang ada tanpa tingkat intervensi manusia tersebut, dan tergantung pada berbagai keadaan, hal ini dapat menyebabkan masalah lebih lanjut. Misalnya, data yang digunakan untuk pelatihan mungkin mencerminkan bias manusia dalam suatu organisasi. Demikian pula, penggunaan hanya sebagian data untuk pelatihan dapat menciptakan bias yang dapat terus-menerus mengkodifikasi masalah ke dalam praktik bisnis yang digerakkan oleh AI. Ini akan menjadi tambahan pada tantangan sebelumnya dengan perangkat lunak.

Komputer menafsirkan algoritma secara harfiah. Mereka melakukan apa yang diminta untuk mereka lakukan atau, dalam kasus pembelajaran mesin, apa yang telah dilatih untuk mereka lakukan, tanpa penerapan akal sehat apa pun. Komputer tidak berperilaku seperti manusia. Mereka tidak memiliki penilaian di luar baris kode mereka atau pembelajaran dari data. Ada banyak contoh komputer yang mengorbankan manfaat jangka panjang untuk keuntungan jangka pendek. Situs web yang menggunakan pembelajaran mesin untuk meningkatkan rasio klik iklan, misalnya, tanpa mempertimbangkan kepuasan pelanggan dapat

membuat penggunanya tenggelam dalam artikel umpan klik yang tidak berguna dan sering kali menyinggung karena model AI mereka dilatih pada tujuan yang salah.

Algoritma juga bersifat rakus dalam hal ini, istilah "rakus" merupakan istilah teknis. Algoritma ini diprogram untuk membuat pilihan terbaik di setiap langkah proses tanpa melihat ke masa depan untuk memprediksi ke mana arah semua jalur ini. Terkadang hal ini memberikan jawaban terbaik; terkadang tidak. Analoginya adalah sistem navigasi GPS yang mengoptimalkan waktu di atas pertimbangan lain. Algoritma semacam itu, yang dirancang untuk mengoptimalkan keuntungan sekecil apa pun, akan mengarahkan pengemudi ke rute memutar yang menggunakan lebih banyak bahan bakar atau bahkan lebih berbahaya, selama ia memberikan keunggulan, bahkan satu nada per detik, atas arah yang lebih mudah. Lebih buruk lagi, beberapa belokan ini mungkin membutuhkan waktu lebih lama dari yang diperkirakan oleh algoritma, sehingga menghilangkan keuntungan apa pun.

Masalah-masalah ini diperparah oleh fakta bahwa banyak model AI yang dibangun di atas pembelajaran mesin tidak mudah diinterpretasikan. Kotak hitam ini menerima masukan dan memberikan keluaran atau prediksi dengan sedikit atau tanpa indikasi tentang bagaimana atau mengapa prediksi tersebut dibuat. Karena komputer "menciptakan" logika berdasarkan data yang dipelajarinya, terkadang sulit untuk memprediksi bagaimana komputer akan berperilaku dalam situasi tertentu, atau mengapa ia berperilaku dengan cara tertentu. Sekalipun penjelasannya dapat dipahami secara matematis, hanya sedikit pengguna bisnis yang memiliki latar belakang untuk memahaminya secara intuitif. Selain itu, hanya ada sedikit cara untuk menanyakan algoritma AI tentang mengapa ia membuat keputusan tersebut.

Semakin orang memandang AI sebagai kotak hitam, semakin mereka khawatir dengan sifatnya. Kita sudah melihat publik, serta regulator dan media, khawatir bahwa AI adalah jin yang tidak dapat dikendalikan. Kekhawatiran ini hanya akan semakin parah seiring waktu seiring AI menjadi semakin umum. Seiring AI semakin merambah pasar keuangan, kedokteran, hukum, manufaktur, dan di mana pun, orang perlu tahu bagaimana, apa, dan mengapa algoritma AI melakukan apa yang mereka lakukan dan bagaimana mengurangi risiko penggunaan model. Jika tidak, kekhawatiran akan menghambat penerapan AI dalam bisnis.

## 10.2 MITIGASI RISIKO MODEL

Mengelola risiko model AI bukan hanya ide yang bagus, tetapi juga keharusan bisnis. Risiko ini digunakan tidak hanya untuk mematuhi hukum, tetapi juga untuk melakukan hal yang benar. Sebagai pemimpin perusahaan, karyawan, warga negara, dan pemangku kepentingan dalam hasil bisnis dan sosial, kita harus memikirkan dan bertindak berdasarkan tata kelola risiko model untuk model AI di setiap langkah pengembangan dan penggunaannya. Untuk memitigasi risiko model AI, perusahaan harus menerapkan tata kelola model di seluruh siklus hidup pemodelan: sebelum model dikembangkan, saat model sedang dikembangkan, dan setelah model dibangun.

### ***Sebelum Pemodelan***

Sebelum pemodelan, fokus harus diberikan pada data. Tim harus memastikan mereka memiliki data yang tepat untuk membangun model mereka. Data yang baik memerlukan

penilaian kualitas data, seperti mencari nilai yang hilang, kesalahan sistematis, dan keterlambatan ketersediaan data. Tata kelola data untuk pemodelan harus mencakup pemahaman tentang kualitas data; silsilah data (mengetahui data apa yang digunakan dan dari mana asalnya); privasi data (memastikan bahwa model yang dipertimbangkan sejalan dengan privasi data berdasarkan kebijakan); proses propagasi data (untuk mencegah, misalnya, kebocoran data uji ke dalam lingkungan produksi); dan pelacakan penggunaan atribut yang dibatasi, seperti jenis kelamin atau ras. Jika tim menggunakan jenis kelamin atau ras dalam suatu model, model tersebut perlu melalui langkah validasi atau tata kelola tambahan, seperti pengujian kewajaran, yang akan kita bahas nanti di bab ini. Tim juga harus memetakan regulasi ke entitas dan atribut serta menandai batasan yang sesuai.

Salah satu aspek penting dari data yang baik adalah data tersebut harus memiliki keluasan dan kedalaman. Dalam data pelanggan, misalnya, keluasan data melibatkan pengamatan terhadap sejumlah besar pelanggan, sedangkan kedalaman data melibatkan fokus pada lebih banyak data tentang setiap pelanggan. Kumpulan data yang lebih luas dan mendalam memungkinkan model untuk menangani situasi dengan lebih baik dan lebih terprediksi serta membantu mengurangi bias; faktanya, kurangnya keluasan data inilah yang harus diatasi Amazon dalam perangkat lunak perekrutannya. Salah satu cara untuk memahami hal ini adalah bahwa kedalaman dan keluasan data menambahkan sedikit "akal sehat" pada algoritma pembelajaran.

Keadilan juga perlu menjadi tujuan dalam pengambilan keputusan algoritmik. Meskipun penggunaan data dan algoritma aktual dapat membantu menghilangkan bias yang ada, hal itu juga dapat menimbulkan bias baru atau mengkodifikasi bias yang sudah ada dari data historis. Untuk mengatasi hal ini, penelitian AI baru-baru ini telah mencoba mendefinisikan apa yang dimaksud dengan "adil" dalam pemodelan prediktif dan bagaimana menangani keadilan. Dua pendekatan utama yang mulai digunakan perusahaan dalam bisnis mereka adalah keadilan kelompok, yang mengharuskan hasil prediksi serupa untuk semua kelompok, dan keadilan individu, yang mengharuskan individu yang serupa diperlakukan serupa terlepas dari keanggotaan kelompok.

Salah satu cara untuk mengatasi kewajaran sebelum pemodelan adalah dengan menilai data untuk mencari bias, dan melakukan praproses data untuk menghilangkan bias sebelum membuat model darinya. Ada beberapa pendekatan untuk melakukan hal ini. Salah satunya, yang disebut penekanan, menghilangkan atribut yang dilindungi, seperti jenis kelamin, ras, agama, atau disabilitas, dari data pelatihan. Namun, hanya menghapus variabel-variabel ini ternyata tidak efektif, tidak melakukan apa pun untuk mengatasi bias dan bahkan mungkin menyembunyikannya. Hal itu karena "kewajaran melalui ketidaksadaran" ini, sebagaimana disebut, mengabaikan penyandian yang redundan cara untuk menyimpulkan atribut yang dilindungi dari fitur yang tidak dilindungi. Misalnya, jika kita membangun model untuk menyetujui aplikasi pinjaman, menghapus jenis kelamin pemohon sebagai input seringkali tidak cukup, karena, berdasarkan data historis, jika kita mengetahui input lain dan keputusan akhir, kita seringkali dapat memprediksi jenis kelamin dengan cukup akurat. Prediktibilitas ini

berarti bahwa bias telah dikodekan dalam variabel lain. Misalnya, kode pos atau nama dapat digunakan untuk menyimpulkan ras, jenis kelamin, atau kebangsaan.

Untuk mengatasi hal ini, pertama-tama kami mencari semua fitur dan atribut lain yang sangat berkorelasi dengan atribut yang dilindungi (dan yang dapat digunakan untuk memprediksi karakteristik yang dilindungi). Kemudian, kami melakukan praproses data untuk menghilangkan bias menggunakan pendekatan yang mencakup transformasi data, di mana data masukan (historis) ditransformasikan menggunakan algoritma optimasi untuk menghilangkan diskriminasi, dan pembobotan ulang data, di mana bobot diberikan pada set data masukan untuk membuat data pelatihan bebas diskriminasi dan memastikan kewajaran selama pelatihan model. Dalam setiap opsi praproses ini, penurunan diskriminasi mengorbankan akurasi dan harus dioptimalkan sebagai kompensasi. Meskipun saya secara khusus menggunakan contoh kewajaran di sini, masing-masing pendekatan ini berlaku untuk semua jenis bias yang tertanam dalam data baik data tersebut tentang orang, atau tentang mesin dari sensor, atau data perdagangan tentang perusahaan. Bias harus diidentifikasi dan diatasi dalam data sebelum pemodelan.

### ***Selama Pemodelan***

Saat pemodelan, validasi kinerja prediktif model diperlukan. Ambang batas kinerja seperti akurasi, positif palsu, dan negatif palsu harus ditetapkan sejak dini dan dibandingkan dengan tolok ukur. Seringkali terdapat trade-off di antara keduanya, sehingga pemilihan ambang batas yang akan dioptimalkan sangatlah penting. Ingatlah bahwa dalam contoh deteksi kanker dari Bab 8, negatif palsu dapat berpotensi mengancam jiwa, sedangkan positif palsu tidak akan berakibat fatal, meskipun hampir pasti akan memberikan pengalaman buruk bagi pasien.

Penting juga untuk menggunakan serangkaian tujuan yang tepat atau beragam untuk melatih model AI guna menghindari jebakan tujuan yang sempit dan literal. Pertimbangkan contoh clickbait. Akan lebih baik untuk membuat model yang dirancang untuk kepuasan pengguna atau pembelian aktual, alih-alih hanya klik iklan. Namun, untuk melakukan hal tersebut, model harus dilatih berdasarkan tujuan kepuasan pengguna, yang diukur, misalnya, dengan jumlah waktu yang dihabiskan di halaman yang diklik, atau pembelian aktual, alih-alih hanya klik iklan. Dengan kata lain, tujuan (yaitu, target model) harus berubah, setidaknya sebagian.

Ada banyak situasi di mana sebuah tim mungkin perlu mengintegrasikan berbagai tujuan tersebut ke dalam sistem AI. Misalnya, sebuah bisnis mungkin ingin meningkatkan penjualan daring dan memiliki tingkat pengembalian yang rendah. Itu berarti kemungkinan akan mempertimbangkan target pembelajaran seperti jumlah penjualan dikurangi jumlah pengembalian atau nilai dolar yang setara. Jika dalam contoh clickbait kami ingin juga meningkatkan klik karena kami ingin pengguna mengunjungi beberapa halaman, kami dapat menyertakan waktu yang dihabiskan dan jumlah klik sebagai tujuan dengan mendefinisikan target tujuan baru, yang merupakan perkalian jumlah klik dan waktu yang dihabiskan di halaman yang dihasilkan. Ini akan membantu mengabaikan skenario di mana pengguna selalu pergi ke satu halaman dan membacanya untuk waktu yang lama dan tidak pergi ke tempat

lain, misalnya. Model yang dibangun berdasarkan beberapa tujuan, menggunakan dimensi yang berbeda dari variasi data, mengurangi kemungkinan bahwa model ini akan dengan satu pikiran mencoba memecahkan hanya untuk satu hal. Namun, dalam prosesnya, data yang tidak sesuai harus dihindari. Apa yang tidak sesuai tergantung pada kasus penggunaan. Misalnya, untuk seseorang yang melakukan diagnosis medis, jenis kelamin pasien dan usia pasien mungkin sangat relevan untuk meningkatkan kekuatan diagnostik prediktif model. Di sisi lain, memasukkan data terkait gender dan usia yang sama dalam memutuskan aplikasi hipotek atau pinjaman mungkin tidak etis atau bahkan ilegal.

Untuk mengatasi bias selama proses pemodelan, setelah bias dari data dihilangkan sebelum pemodelan, salah satu pendekatan umum adalah melakukan pascaproses keluaran – yaitu, mengambil keluaran model pembelajaran mesin dan memprosesnya lebih lanjut menggunakan algoritma lain untuk mengurangi bias, sehingga meningkatkan keadilan bagi anggota kelompok yang dilindungi. Pendekatan lain adalah menambahkan kendala, yang disebut istilah regularisasi bias, ke tujuan optimasi pelatihan model. Biasanya, tujuan tersebut hanya mencakup kesalahan, dan dalam hal ini, kesalahan tersebut merupakan jumlah kesalahan dan bias yang harus diminimalkan bersama. Variasi lain dari hal ini yang baru-baru ini dikembangkan, disebut peluang yang disamakan, membatasi algoritma sehingga baik positif palsu maupun negatif palsu tidak memengaruhi subkelompok yang dilindungi secara tidak proporsional. Ada alternatif dan variasi lain yang harus dieksplorasi oleh ilmuwan AI, tergantung pada kebutuhan bisnis untuk bias rendah.

Sangat penting untuk memeriksa proses pembuatan model dan penggunaan data guna memastikan reproduktifitas. Pembelajaran mesin berupaya menemukan fungsi dengan kesalahan terkecil. Jika proses pemodelan tidak menemukan minimum global melainkan hanya minimum lokal, maka setiap iterasi pelatihan dapat menemukan minimum yang berbeda dan, karenanya, menghasilkan prediksi yang berbeda. Bayangkan seorang pendaki yang sedang mencari titik terendah di daerahnya. Ia mencapai apa yang tampaknya merupakan titik terendah ini, yang dikenal sebagai minimum. Jika ia bergerak sedikit ke arah mana pun, ia akan mendaki lagi. Namun, titik terendah ini mungkin hanya minimum lokal, bukan minimum global: mungkin ada lembah lain (minimum lokal) di lokasi lain yang lebih rendah daripada tempat ia berada. Dalam ekuivalen algoritmik dari situasi ini, proses pelatihan mungkin telah menemukan minimum lokal dari fungsi kesalahan tetapi bukan minimum globalnya. Menjalankan kembali proses pelatihan dapat mengubah algoritma yang dihasilkan secara signifikan jika menemukan satu minimum lokal dan kemudian, setelah pelatihan ulang, menemukan minimum yang berbeda, sehingga memberikan hasil yang berbeda. Proses pemodelan juga harus mencakup tinjauan kode untuk memastikan praktik terbaik dan mengonfirmasi bahwa model atau bagian model dapat digunakan kembali di seluruh tim.

### **Setelah Pemodelan**

Setelah model dibangun dan divalidasi, penting untuk melakukan berbagai uji risiko model, termasuk uji kepatuhan, uji kewajaran, uji sensitivitas, dan uji kondisi batas. Model juga harus diperiksa untuk interpretabilitas hasilnya.

Uji kewajaran melihat apakah kelompok kurang mampu atau terlindungi berdasarkan hal-hal seperti gender atau ras diperlakukan secara adil. Ini sama dengan uji bias untuk jenis data lainnya. Ini memastikan bahwa model akhir tidak bias. Uji ini menguji bagaimana model berperilaku dengan input yang berbeda misalnya, jika data yang sama digunakan saat menentukan hasil aplikasi hipotek tetapi gendernya diubah, apakah output atau keputusannya berubah? Uji ini juga menguji hasil agregat misalnya, apakah jumlah pria dan wanita yang disetujui atau ditolak pada dataset baru sama banyaknya?

Sebagian besar teknik ini dapat dievaluasi berdasarkan berbagai metrik kewajaran algoritmik seperti peluang yang disamakan, paritas statistik, dampak yang berbeda, paritas prediktif, kesempatan yang sama, peluang rata-rata, atau kesetaraan prediktif. Namun, baik metrik kewajaran maupun metrik akurasi perlu didefinisikan selama pemodelan, karena beberapa mungkin bekerja lebih baik bersama-sama daripada yang lain.

Analisis sensitivitas memvalidasi bahwa berbagai kombinasi bidang masukan berperilaku sesuai. Analisis sensitivitas dan pengujian kondisi *batas* relatifnya, melibatkan penyediaan berbagai masukan yang terganggu (di mana masukan sedikit berubah) dan memeriksa keluaran untuk menemukan faktor masukan mana yang paling penting dalam menentukan hasil tertentu dan apakah, dalam beberapa kondisi masukan, model dapat "menjadi liar" dengan memberikan keluaran yang berpotensi tidak valid. Analisis sensitivitas juga dapat menggambarkan bagaimana keluaran model bergantung pada masukannya di berbagai wilayah ruang masukan data.

Ada berbagai cara untuk melakukan ini. Pertama, tim perlu memahami distribusi nilai fitur. Katakanlah dua fitur yang dihadapi bisnis adalah usia dan tinggi badan. Tim mulai dengan meringkas data, menentukan nilai minimum (misalnya usia dan tinggi badan terendah), dan nilai maksimum (misalnya usia dan tinggi badan tertinggi). Kemudian, ditentukan nilai rata-rata, atau nilai persentil ke-50, untuk masing-masing fitur. Dari sana, tim menghitung nilai-nilai lain yang mungkin berguna, biasanya persentil ke-25 untuk setiap fitur (usia atau tinggi badan saat seperempat titik data lebih rendah) dan persentil ke-75.

Tim kemudian menetapkan titik referensi dengan menjalankan model menggunakan nilai rata-rata fitur-fiturnya, lalu mengubah nilai satu fitur pada satu waktu (misalnya, usia), sambil mempertahankan nilai rata-rata fitur lainnya (tinggi badan). Tim kemudian dapat terus mengubah usia ke nilai yang berbeda, seperti nilai minimum, persentil ke-25, persentil ke-75, dan maksimum, serta mengamati bagaimana prediksi bervariasi. Tim juga dapat mengubah nilai dua fitur sekaligus, alih-alih satu, atau dalam kasus variabel yang lebih banyak, mencoba semua kombinasi pada interval yang sama.

Misalnya, seorang ilmuwan AI menggunakan fitur usia dan tinggi badan untuk memprediksi berat badan. Untuk melihat apakah model memberikan respons yang wajar dengan input yang berbeda, ia menjalankan model pada rata-rata tinggi badan dan usia, misalnya, 175 cm dan 130 cm. Setelah menyesuaikan model dengan data, modelnya memprediksi bahwa individu dengan tinggi badan 175 cm dan 130 cm memiliki berat badan 68 kg. Untuk menentukan sensitivitas model, ia kemudian dapat menjalankannya menggunakan nilai yang sama untuk tinggi badan tetapi mengubah nilai usia menjadi 44, lalu

46. Ini memberi tahu seberapa sensitif berat badan yang diprediksi terhadap perubahan kecil pada usia. Sekarang ia dapat menggunakan persentil ke-25, ke-75, atau persentil lainnya. Jika keluaran tidak banyak berubah, bahkan dengan modifikasi substansial pada masukan, ia telah mengetahui bahwa usia tidak penting dalam memprediksi berat badan. Jika berubah, maka usia merupakan fitur penting dalam model. Selain itu, ini akan menunjukkan titik tertentu di mana model mulai menjadi lebih sensitif.

Pengujian kondisi batas melibatkan penggunaan nilai ekstrem masukan sebagai kasus uji. Ini dapat berarti mengubah nilai fitur usia secara lebih dramatis, misalnya, ke persentil ke-95 atau bahkan persentil ke-99,9. Keluaran yang tidak terduga, seperti berat badan 3.000 pon dalam skenario ini, akan menunjukkan bahwa model tersebut bermasalah. Tim perlu menemukan model yang lebih baik yang dapat diandalkan untuk berkinerja baik. Mereka harus memastikan bahwa tidak ada kombinasi masukan yang wajar tetapi ekstrem yang menghasilkan keluaran yang tidak wajar.

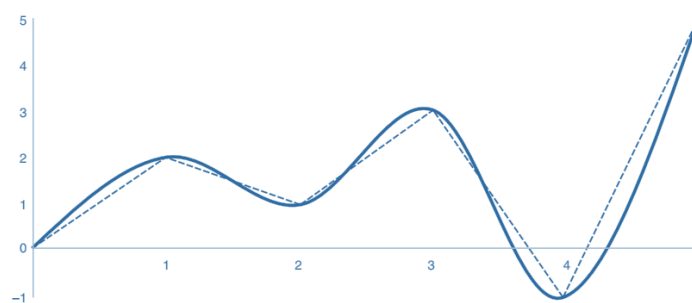
Interpretabilitas model berguna untuk mengembangkan pemahaman intuitif tentang suatu model. Interpretasi model adalah praktik mengembangkan penjelasan post hoc atas hasil yang dihasilkan oleh model AI. Interpretabilitas adalah istilah yang agak samar yang berarti kemampuan untuk melihat apa yang dipikirkan komputer dengan cara yang dapat dipahami oleh manusia. Karena tidak ada karakterisasi yang tepat atau pemahaman umum tentang interpretabilitas, orang-orang telah mengusulkan definisi seperti: pemahaman mekanistik tingkat rendah tentang model, yang diterapkan pada fitur, parameter, model, atau algoritma pelatihan; atau, model yang menggunakan sebab-akibat untuk membenarkan keputusannya. Cara memandang interpretabilitas model ini terkadang disebut AI yang dapat dibenarkan.

Lebih bermanfaat untuk menganggap interpretabilitas sebagai sesuatu yang memungkinkan pengguna menggabungkan informasi model dengan pengetahuan mereka sendiri. Hal ini memungkinkan mereka untuk mulai memahami keluaran model, mengembangkan intuisi yang lebih baik, dan dari sana membuat perubahan pada model atau menyempurnakan penjelasan yang mungkin diperlukan untuk mengimplementasikan keluaran yang direkomendasikan (misalnya, mengapa strategi perdagangan tertentu harus diikuti). Hal ini bisa sangat berguna meskipun tidak menjelaskan cara kerja internal model secara pasti. Bahkan, dalam praktiknya, hal ini biasanya lebih bermanfaat daripada memahami mekanisme algoritma secara tepat. Misalnya, jika sebuah model memprediksi churn untuk pelanggan tertentu, model diagnostik mungkin dapat memberikan intuisi kepada pengguna dengan menunjukkan pelanggan serupa yang telah berhenti berlangganan. Dalam hal ini, "serupa" berarti bahwa informasi tentang pelanggan ini dalam hal penggunaan, panggilan ke pusat panggilan, dan sebagainya, menyerupai informasi pelanggan lain yang telah berhenti berlangganan. Atau, model diagnostik mungkin dapat mengatakan bahwa churn diprediksi karena panggilan pelanggan ke layanan pelanggan telah meningkat secara signifikan. Membangun intuisi model seperti membangun intuisi mengemudi seorang pengemudi tidak perlu mengetahui mekanisme detail cara kerja mobil, tetapi perlu memahami bagaimana

mobil tersebut berperilaku dalam menanggapi kondisi jalan dan dalam menanggapi masukannya.

Meskipun interpretabilitas terus menjadi bidang penelitian akademis yang aktif, beberapa pendekatan sedang diimplementasikan secara produktif. Bisnis saat ini memanfaatkan teknik-teknik ini untuk mengatasi berbagai masalah. Hal ini memungkinkan tim untuk berhasil mengelola model sehingga mereka dapat terus menerapkannya ke dalam produksi. Salah satu metode tersebut melibatkan penggunaan model lokal pengganti, yang merupakan model yang sangat mencerminkan model asli tetapi lebih dapat diinterpretasikan, meskipun mungkin kurang akurat. Sebuah analogi untuk memahami model pengganti adalah dengan memikirkan cara mengaproksimasi fungsi polinomial kompleks menggunakan fungsi pengganti linier di berbagai bagian ruang data masukan (yaitu sumbu  $x$ ). Gambar 10.1 menunjukkan polinomial yang didekati oleh lima model pengganti. Yang pertama hanya berlaku untuk data masukan (sumbu  $x$ ) dari 0 hingga 1, dan yang kedua hanya berlaku untuk data masukan dari 1 hingga 2, dan seterusnya. Anda tidak dapat berasumsi bahwa fungsi pengganti pertama (garis lurus antara 0 dan 1) akan berfungsi dengan baik untuk semua masukan data (misalnya untuk  $x = 3$ ). Selain itu, setiap model pengganti kurang akurat dibandingkan polinomial. Namun, fungsi linear lebih intuitif dan mudah diinterpretasikan.

Demikian pula, model pembelajaran mesin yang berfungsi untuk semua data adalah fungsi multidimensi yang rumit dan belum tentu mudah diinterpretasikan. Salah satu cara untuk membangun intuisi tentang model tersebut adalah dengan membuat beberapa model yang lebih sederhana, menggunakan subset data yang berbeda yang lebih mudah diinterpretasikan.

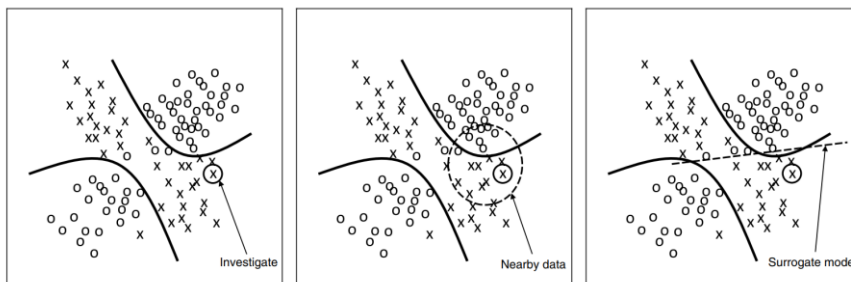


**Gambar 10.1** Aproksimasi fungsi polinomial menggunakan fungsi linear yang lebih sederhana di berbagai bagian sumbu  $x$ .

Gambar 10.2 memberikan skema bagaimana pendekatan model pengganti bekerja dalam praktik. Contoh ini menunjukkan potensi masalah klasifikasi, misalnya dengan  $X$  mewakili pelanggan yang mungkin menghadapi kesulitan keuangan dalam 12 bulan ke depan dan  $O$  mewakili pelanggan yang tidak akan mengalaminya. Model AI dilatih menggunakan algoritma kotak hitam yang canggih, dan klasifikasi  $X$ -dan- $O$  adalah outputnya. Model klasifikasi ditunjukkan dengan garis tebal pada kotak 1 yang memisahkan  $X$  dari  $O$ .

Untuk mencoba memahami bagaimana keputusan dibuat untuk pelanggan tertentu, nilai input dan output dari kasusnya dipilih untuk interpretasi (kotak 1). Kemudian, model

pengganti dibuat di sekitar titik ini dengan mengidentifikasi titik-titik terdekat (kotak 2). Titik-titik ini kemudian diganggu, atau diubah, dengan jumlah kecil, dan data yang terganggu digunakan sebagai input ke dalam model kotak hitam asli untuk mendapatkan output baru. Masukan-masukan ini dan keluarannya yang sesuai digunakan untuk melatih model yang lebih sederhana, biasanya linier (kotak 3), di mana masukan yang terganggu diberi bobot berdasarkan kedekatannya dengan instans yang dijelaskan (rekaman asli yang dipilih). Model linier yang lebih sederhana ini kemudian digunakan untuk mengidentifikasi fitur-fitur signifikan dan kontribusi peringkatnya terhadap klasifikasi. Parameter-parameter ini dapat berfungsi sebagai penjelasan untuk hasil yang dimaksud.



**Gambar 10.2** Contoh bagaimana model pengganti dapat membantu interpretabilitas.

Perhatikan bahwa model baru ini berfungsi sebagai batas lokal untuk dua jenis klasifikasi X dan O di dekat titik data teridentifikasi yang sedang diselidiki. Penting untuk diingat bahwa model ini tidak berfungsi secara global. Misalnya, jika Anda bergerak ke arah tengah atau kiri pada kotak 3, garis pemisah model pengganti dan garis pemisah model asli berbeda secara signifikan. Dalam model pengganti (garis putus-putus pada kotak 3), di sekitar titik data asli, apa pun di atas garis adalah O dan di bawah garis adalah X. Ketika Anda bergerak ke kiri, model pengganti ini sekarang akan memberikan prediksi yang salah, karena masih akan memprediksi apa pun di atas garis sebagai O. Namun, kita dapat melihat bahwa data sebenarnya menunjukkan X di area ini. Model pengganti hanya akan berfungsi di sekitar titik data yang awalnya diidentifikasi. Jika Anda memerlukan interpretasi untuk wilayah grafik yang berbeda, proses ini harus diulang untuk titik data awal yang berbeda.

Untuk setiap opsi terkait interpretabilitas ini, tim AI juga harus mempertimbangkan kemudahan penggunaan output melalui visualisasi interaktif. Ini akan sangat membantu dalam membuat pengguna akhir memahami model dalam arti heuristik. Pada saat yang sama, tim AI harus ingat bahwa mereka harus menghindari menempatkan batasan yang tidak perlu pada model AI mereka. Mereka ingin algoritme membuat prediksi yang akurat, seringkali melampaui kapasitas manusia mana pun; namun, mereka tidak ingin menerima bias atau risiko yang tidak beralasan atau opasitas yang mendiskualifikasi. Mereka menginjak garis tipis di sini, dan opasitas tertentu mungkin diperlukan. Mengorbankan terlalu banyak kekuatan prediktif untuk meningkatkan interpretabilitas dapat merugikan kasus penggunaan dengan membatasi potensi penuh aplikasi AI.

### **Interpretabilitas Model dalam Pembelajaran Mendalam**

Jaringan pembelajaran mendalam sangat baik dalam menetapkan klasifikasi pada sampel, seperti kucing, bukan-kucing. Namun, pembelajaran mendalam hanya memberikan sedikit wawasan tentang cara kerja algoritma. Meskipun interpretabilitas pembelajaran mendalam masih menjadi bidang penelitian yang aktif, studi terbaru menunjukkan hasil yang menjanjikan. Sebagaimana dibahas dalam Bab 2, beberapa lapisan jaringan ini menerima masukan dan meneruskannya dari satu lapisan ke lapisan lainnya, memberikan deskripsi tingkat tinggi semakin jauh dari lapisan masukan. Terkadang, ketika Anda semakin dekat dengan lapisan keluaran, deskripsi tingkat tinggi ini dapat melacak hal-hal yang menyerupai cara seseorang mendeskripsikannya. Deskripsi tersebut mungkin, misalnya, memiliki simpul atau kelompok simpul yang tampaknya menunjukkan fitur seperti bulu, ekor, atau telinga. Pengujian telah menunjukkan bahwa untuk tugas-tugas tertentu, seperti mengklasifikasikan gambar, melihat keluaran dari setiap lapisan membantu memberikan pemahaman yang lebih baik kepada pengguna tentang cara kerja model.

Setelah model AI memberikan output atau membuat prediksinya, tim harus memikirkan cara terbaik untuk menggunakan output tersebut. Pertimbangkan kembali contoh Target yang dibahas sebelumnya. Tim Target mengamati pembelian pelanggan sebelumnya untuk memprediksi produk lain yang harus mereka rekomendasikan. Namun, mereka tidak memiliki filter yang mencegah mereka mendapatkan pers negatif. Mungkin tidak selalu jelas jenis filter apa yang harus digunakan, dan itu membutuhkan pemikiran melalui berbagai skenario untuk menentukan kapan harus memicu jenis tindakan apa (atau tidak) dari output model.

Dalam upaya untuk memberikan transparansi kepada individu yang terdampak oleh keputusan algoritmik, Peraturan Perlindungan Data Umum (GDPR) memberi orang hak untuk mendapatkan penjelasan ketika dihadapkan dengan keputusan algoritmik seperti mengapa mereka ditolak pinjaman bank. Seperti yang dapat dibayangkan, hal ini dapat menjadi masalah karena sifat input yang kompleks dan ketidakjelasan algoritma AI. Seringkali, mungkin tidak ada penjelasan yang siap untuk keputusan algoritma. Pasal 22 GDPR menyatakan "Pengambilan keputusan individu otomatis, termasuk pembuatan profil, dibuat dapat diperdebatkan." Ini berarti bahwa perusahaan harus dapat menafsirkan keluaran model dan membuktikan bahwa algoritmanya tidak membuat klaim palsu terhadap pelanggan.

Hak hukum untuk penjelasan ini membutuhkan transparansi, dan beberapa perusahaan di Eropa dan di tempat lain telah mengungkapkan beberapa bagian dari kode sumber mereka sebagai hasilnya. Namun, ini hanya memberikan sedikit gambaran tentang apa yang mampu dilakukan kode dalam hal bias. Selain itu, seperti yang disebutkan sebelumnya untuk pembelajaran mesin, bias mungkin ada dalam data pelatihan dan tidak dalam kode sama sekali. Oleh karena itu, interpretabilitas model merupakan persyaratan yang harus dimiliki oleh bisnis yang perlu mematuhi GDPR. Peraturan ini berlaku untuk semua perusahaan

multinasional yang melakukan bisnis di mana pun di Eropa. Selain itu, karena mereka memerlukan perangkat standar privasi, perusahaan harus mempertimbangkan untuk menerapkannya ke seluruh pelanggan mereka di seluruh dunia.

### 10.3 KANTOR RISIKO MODEL

Saat mendefinisikan dan menerapkan pendekatan tata kelola model, pertimbangkan siklus hidup model AI secara keseluruhan, seperti yang dibahas di Bab 8. Untuk proses tata kelola model yang sehat, independen, dan dapat diulang, sebuah bisnis harus membentuk kantor atau tim tata kelola model. Konsekuensi dari tidak melakukannya bisa sangat signifikan. Perhatikan contoh di mana sebuah bank regional menciptakan model AI yang berharga. Para eksekutif segera ingin memproduksinya, yang akan memungkinkan mereka melakukan lebih banyak penjualan silang dan penjualan atas. Namun, CEO tersebut khawatir: "Saya tidak tahu apa yang akan terjadi. Bagaimana jika menjadi liar? Terlalu besar risiko bagi reputasi kami."

Masalahnya adalah bank tersebut lalai menciptakan proses tata kelola untuk algoritma AI. Butuh dua setengah bulan untuk mengembangkan model AI mereka; butuh sembilan bulan untuk menyiapkan sistem tata kelola dan menggunakannya untuk menyetujui model tersebut. Seandainya mereka mulai menyiapkan sistem tata kelola ini sebelum mengerjakan algoritma pertama mereka, mungkin hanya butuh dua hari untuk menjalankan model baru melalui langkah-langkah tata kelola. Akan lebih efisien dan hemat biaya untuk menyiapkan proses tata kelola sejak dini, memastikannya berlaku untuk setiap algoritma AI yang dikembangkan.

Tidak semua kasus penggunaan perlu melalui tingkat ketelitian yang sama dalam hal tata kelola. Bagi bank, satu kasus penggunaan mungkin untuk mempersonalisasi konten di situs web, dan yang lainnya mungkin untuk memperkirakan arus kas dan suku bunga. Kesalahan pada kasus pertama mungkin memiliki konsekuensi yang lebih rendah, sedangkan kesalahan pada kasus kedua dapat menyebabkan masalah likuiditas bagi bank. Tim harus memutuskan langkah tata kelola mana yang harus dilalui model dengan mempertimbangkan kemungkinan masalah dan tingkat dampak atau konsekuensi yang dihasilkan dari masalah tersebut. Jika setiap model melewati tingkat ketelitian tertinggi, hal itu akan memperlambat model berisiko rendah untuk diproduksi dan berpotensi mengalihkan fokus dari model berisiko tinggi. Berbagai faktor yang spesifik untuk bisnis tertentu mungkin berlaku, dan perusahaan harus mengembangkan kerangka kerja risiko model untuk menentukan tingkat risiko untuk berbagai jenis model dalam konteks bisnis mereka. Kerangka kerja ini harus mempertimbangkan tingkat konsekuensi untuk jenis dan penggunaan model, dan menentukan tata kelola apa yang perlu diterapkan untuk berbagai kategori model selama setiap langkah dalam siklus hidup untuk setiap tingkat risiko (tinggi, sedang, atau rendah). Kerangka kerja tersebut harus menentukan langkah mana yang berlaku untuk setiap kategori model, dan langkah mana yang harus dihilangkan untuk mempercepat dan mengurangi biaya untuk model berisiko sedang dan rendah. Bisnis kemudian perlu memutuskan, untuk setiap model yang mereka kembangkan, kategori risiko apa yang termasuk dalam kerangka kerja tersebut, dan dengan demikian langkah tata kelola apa yang perlu dilalui model tersebut.

Untuk pengujian kewajaran, harus ada serangkaian validasi standar. Tim harus memeriksa untuk melihat apakah model tersebut mendiskriminasi kelompok individu yang dilindungi (seperti mereka yang merupakan anggota kategori gender atau ras yang disebutkan sebelumnya). Penting juga bahwa model tersebut mematuhi persyaratan peraturan.

Sebelum menghubungkan model ke proses produksi dan memicu penggunaannya, tim harus memastikan bahwa model tersebut bekerja dengan baik dengan sistem hulu (biasanya data) dan hilir (biasanya penggunaan). Tim juga harus memvalidasi bahwa ada mekanisme pengaman yang tersedia sehingga jika terjadi kesalahan dengan model dalam produksi, ada sesuatu untuk diandalkan, baik model lain atau beberapa tindakan cadangan. Jika suatu model ditemukan tidak bekerja dengan baik, tim mungkin kembali ke model sebelumnya yang diketahui berfungsi, atau alternatif berbasis aturan yang lebih sederhana yang akan berfungsi sementara. Bergantung pada bagaimana model digunakan, alternatif lain mungkin untuk meningkatkan frekuensi intervensi manual dalam suatu keputusan misalnya, jika keluaran model disimpan sebelum digunakan (model batch), maka keluaran ini dapat ditinjau untuk hasil yang buruk dan ditimpa secara manual, hingga model diperbaiki. Jika alternatif pengaman adalah model AI lain, model itu juga harus sesuai dengan langkah-langkah tata kelola. Sebaiknya pertimbangkan mekanisme pengaman ini dan mekanisme pengaman lainnya sejak awal.

Selain itu, situasi lain mungkin perlu ditangani, seperti model yang mulai mengalami penurunan performa dalam produksi atau tidak lagi berfungsi seperti yang diharapkan atau diinginkan, misalnya, karena statistik data masukan telah berubah dari set data pelatihan. Pemantauan berkelanjutan terhadap kinerja model dalam produksi harus dilakukan untuk menemukan situasi ini. Ini termasuk menangkap prediksi model dan data aktual saat tersedia. Dengan demikian, seseorang di perusahaan dapat mendeteksi dan menangani penurunan kinerja model dengan tepat dan cepat. Terkadang, hal ini dapat memicu pengaman performa yang disebutkan sebelumnya, dan terkadang model perlu dilatih ulang.

Semua aspek mitigasi risiko ini penting untuk dipertimbangkan sebelum mengimplementasikan model AI dalam skala besar. Selanjutnya, kita akan melihat bagaimana sebuah organisasi mempertahankan dan mengembangkan kapabilitas AI dengan mempertimbangkan struktur organisasi terbaik untuk implementasinya.

## **BAB 11**

### **MENGAKTIFKAN KAPABILITAS ORGANISASI**

Sejauh ini, kita telah membahas proses dan alat inti yang digunakan perusahaan untuk menerapkan strategi AI perusahaan. Namun, dibutuhkan lebih dari sekadar proses dan alat untuk mengembangkan dan mempertahankan kapabilitas AI di seluruh perusahaan; hal ini membutuhkan orang-orang yang terorganisasi. Ali Ghodsi dari UC Berkeley mengatakan, “Hanya sekitar 1% perusahaan yang berhasil dalam AI. Sisanya berjuang dengan tiga masalah, yang pertama adalah membuat orang bekerja sama dan berkolaborasi satu sama lain.” Terdapat hambatan signifikan dalam mengintegrasikan AI di sebagian besar organisasi; beberapa aspek model operasional perusahaan harus dikonfigurasi ulang agar berhasil menerapkan AI dalam skala besar. Untuk mempersiapkan organisasi dalam penggunaan AI secara luas, perusahaan perlu membantu tim manajemen, karyawan bisnis, dan pihak-pihak lain untuk berpikir dengan cara yang baru dan berbeda. Namun, bahkan sebelum itu, hal pertama yang harus dilakukan adalah mencapai keselarasan dengan pemangku kepentingan terkait AI.

#### **11.1 MENYELARASKAN PEMANGKU KEPENTINGAN**

Membangun kapabilitas AI secara sukses di dalam suatu organisasi membutuhkan dukungan yang kuat dari dewan direksi dan tim manajemen, sesuatu yang juga berlaku untuk membangun kapabilitas organisasi apa pun. Masalahnya, manajemen mungkin merasa seolah-olah bertaruh (terkadang perusahaan) pada teknologi yang belum terbukti di dalam organisasi. Perasaan ini bisa meresahkan, dan untuk mengatasinya, para eksekutif harus meluangkan waktu untuk memahami cara kerja AI dan pembelajaran mesin; para eksekutif tersebut perlu memungkinkan pengambilan keputusan yang lebih baik dengan membangun intuisi mereka sendiri tentang apa yang nyata dan apa yang hanya sensasi dalam AI, dan memikirkan langkah-langkah yang diperlukan untuk mengimplementasikan solusi AI di seluruh organisasi mereka. Tim manajemen kemudian dapat mendefinisikan visi untuk AI. Buku ini dapat berfungsi sebagai panduan saat mengembangkan visi tersebut.

Namun, yang krusial bagi keberhasilan setiap program AI adalah dukungan tidak hanya dari para eksekutif tetapi juga sebagian besar, jika tidak semua, pemangku kepentingan lainnya. Para pemimpin harus menyatukan organisasi mereka di belakang tujuan visioner mereka. Karyawan harus memahami mengapa AI penting bagi perusahaan mereka dan memahami bagaimana AI akan memengaruhi pekerjaan mereka serta bagaimana mereka dapat berkontribusi pada organisasi yang berpusat pada AI. Para eksekutif perlu menjelaskan bahwa pekerjaan kemungkinan akan berubah, tetapi menjadi lebih baik, ditingkatkan dengan penambahan alat baru berbasis AI. Para pemimpin bisnis perlu menekankan bahwa AI tidak menggantikan sebagian besar pekerjaan; AI menggantikan tugas, yang seringkali berat, sehingga memungkinkan karyawan untuk fokus pada strategi dan kreativitas, alih-alih pekerjaan yang rutin atau membosankan. Stereotip negatif AI harus dilawan untuk

menghilangkan rasa takut kehilangan otoritas dan mendorong lingkungan kerja yang sehat di mana manusia dan mesin bekerja sama dan berkembang.

Pimpinan harus siap menghadapi beberapa penolakan. Menerapkan strategi AI dapat melibatkan perubahan yang signifikan, dan hal itu dapat membuat orang merasa tidak nyaman. Dalam situasi ini, sebaiknya dimulai dengan perubahan bertahap atau proyek percontohan di setiap fungsi atau departemen agar kelompok merasa nyaman dengan AI. Ini berarti menjangkau personel yang mungkin tidak berinteraksi secara teratur, seperti mereka yang bekerja di pemasaran, rantai pasokan, SDM, dan TI. Selain itu, karyawan di perusahaan harus memahami cara berpartisipasi bagaimana mereka dapat mengidentifikasi potensi kasus penggunaan, langkah apa yang perlu mereka ambil untuk memulai pengembangan solusi, dan bagaimana mereka dapat memanfaatkan AI dalam peran mereka. Hasil dari hal ini adalah tercapainya keselarasan tujuan AI dan meningkatnya minat untuk berkontribusi pada solusi AI di berbagai lini bisnis, dari lini depan hingga lini belakang, lintas departemen dan merek.

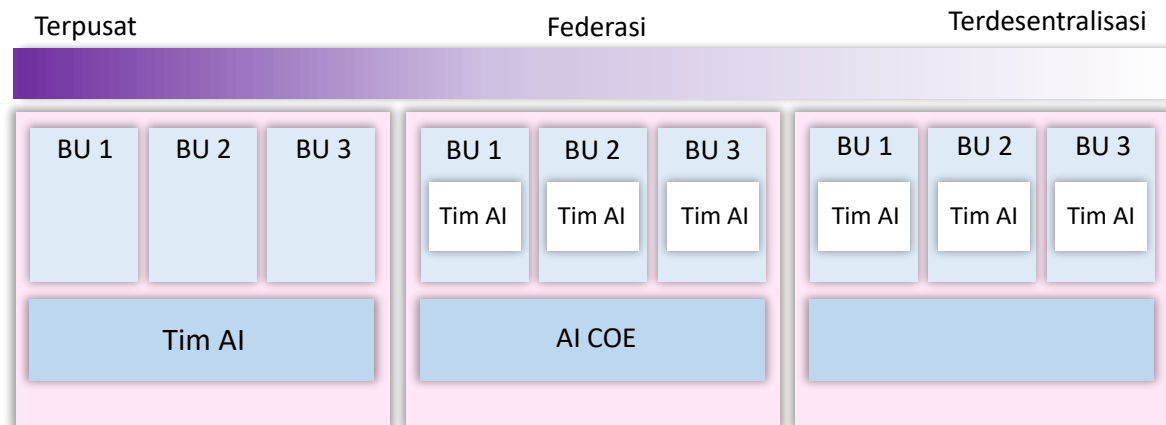
## **11.2 MENGORGANISIR UNTUK SKALA**

Setelah ada keselarasan visi dan tujuan AI, tim kepemimpinan perlu menentukan di mana tim AI berada dalam organisasi dan bagaimana mereka berinteraksi dengan kelompok lain. Harus ada struktur organisasi yang dipahami dengan baik yang mencakup tim bisnis, TI, dan AI, serta yang menetapkan peran dan tanggung jawab terkait untuk keberhasilan pelaksanaan dan pemeliharaan proyek AI. Berbagai fungsi perlu berkolaborasi di seluruh struktur organisasi dengan nyaman. Tanpa kolaborasi ini, bisnis tidak dapat maju. Memilih struktur terbaik dapat menjadi tindakan penyeimbangan yang rumit, karena mungkin tidak ada konsensus tentang di mana kapabilitas AI dan analitik harus ditempatkan dalam suatu bisnis. Selain itu, struktur terbaik bersifat spesifik untuk setiap organisasi sesuatu yang berhasil untuk satu perusahaan mungkin belum tentu berhasil untuk perusahaan lain (lihat Gambar 11.1).

Salah satu pendekatan adalah memiliki struktur terpusat untuk tim AI dan analitik, di bawah satu kepemimpinan. Dalam struktur ini, ilmuwan AI melapor kepada kepala AI atau kepala analitik dan ditugaskan ke proyek sesuai kebutuhan. Sentralisasi mendorong kolaborasi dan berbagi pengetahuan di antara rekan-rekan AI. Hal ini mengarah pada kepatuhan yang lebih baik terhadap kebijakan AI perusahaan dan standarisasi aplikasi yang lebih tinggi di seluruh unit bisnis dengan memastikan bahwa setiap orang dalam grup bekerja berdasarkan seperangkat aturan AI yang sama. Hal ini juga memungkinkan percepatan prioritas strategis AI karena perencanaan dan pelaksanaan terpusat.

Namun, struktur terpusat berarti ilmuwan AI tidak "berada di ruangan yang sama" dengan pengguna bisnis mereka agak jauh dari para pengambil keputusan di unit bisnis dan dari pengguna yang mengoperasikan bisnis. Hal ini dapat menghambat kemampuan tim AI untuk berhasil mengeksekusi kasus penggunaan karena lebih sulit untuk mendapatkan masukan dari unit bisnis dan mendapatkan persetujuan untuk menggunakan model setelah dikembangkan. Struktur terpusat juga memiliki kelemahan lain. Inisiatif dapat melambat jika harus melalui tim terpusat setiap kali unit bisnis ingin meluncurkan proyek AI. Selain itu, jika

unit bisnis ditugaskan ilmuwan AI yang berbeda setiap kali memulai proyek, kurangnya kontinuitas yang dihasilkan dapat menyebabkan unit bisnis merasa bahwa struktur terpusat tidak sepenuhnya mendukung kebutuhan mereka.



**Gambar 11.1** Model operasi terpusat, terdesentralisasi, dan terfederasi untuk AI.

Pendekatan alternatifnya adalah struktur terdesentralisasi, di mana masing-masing unit bisnis membangun kapabilitas sains dan rekayasa AI mereka sendiri dan merekrut tim mereka sendiri. Setiap unit bisnis menetapkan agenda dan prioritasnya secara independen dari unit bisnis lain dan bertanggung jawab untuk memecahkan tantangan AI dengan caranya sendiri. Dalam model ini, ilmuwan AI mendapatkan instruksi langsung dari unit bisnis, memungkinkan tim AI memperoleh pemahaman yang lebih mendalam tentang persyaratan dan nuansa bisnis tersebut. Hal ini memungkinkan tim AI dan anggota tim bisnis untuk berkomunikasi lebih sering dan terbuka. Hal ini membuat tim AI merasa dilibatkan dalam proses tersebut. Interaksi yang lebih dekat dan berkelanjutan sering kali meningkatkan penerimaan bisnis terhadap hasil dari tim AI.

Model terdesentralisasi dapat memiliki kekurangan. Misalnya, model ini mengharuskan setiap unit bisnis memiliki manajer yang memiliki pengalaman dan fasilitas yang cukup untuk merekrut ilmuwan AI yang tepat dan membimbing mereka dalam pekerjaan mereka. Selain itu, desentralisasi dapat menyebabkan duplikasi upaya, sehingga organisasi tidak dapat memanfaatkan apa yang telah dibangun dan mengakibatkan berbagai proyek serupa dikelola di seluruh unit bisnis. Kerugian lainnya adalah bahwa masing-masing unit bisnis mungkin hanya mampu memecahkan masalah yang lebih kecil, karena tidak memiliki anggaran atau kapasitas individual untuk menghadapi tantangan yang lebih substansial di seluruh bisnis.

Pertimbangkan contoh sebuah perusahaan di mana CTO menemukan enam chatbot berbeda sedang dikembangkan di antara berbagai unit bisnis dengan fungsionalitas yang tumpang tindih secara signifikan, semuanya berada pada tahap penyelesaian yang berbeda-beda. Setiap unit ini hanya berfokus pada pemecahan masalah yang lebih mudah, yaitu pengenalan maksud dalam pertanyaan pengguna, menghindari masalah yang lebih kompleks, yaitu penggunaan AI untuk menemukan jawaban atas pertanyaan secara otomatis dari teks

yang tidak terstruktur. Upaya yang berlebihan ini dapat dihindari jika perencanaan dan koordinasi kelompok telah ditangani sejak tahap awal.

Pilihan ketiga adalah mengambil pendekatan hibrida struktur terfederasi, atau model hub-and-spoke, yang memanfaatkan kekuatan model terpusat dan terdesentralisasi sekaligus meminimalkan kelemahannya. Model terfederasi ini memiliki hub pusat kita sebut saja pusat keunggulan AI (CoE) dan spoke yang berada di dalam unit bisnis yang berbeda. Dalam struktur seperti ini, sebagian besar ilmuwan AI bekerja di unit bisnis tetapi berkoordinasi melalui CoE AI, sehingga mereka mendapatkan visibilitas yang diperlukan untuk memanfaatkan pengetahuan dan solusi secara lintas unit.

### 11.3 PUSAT KEUNGGULAN AI

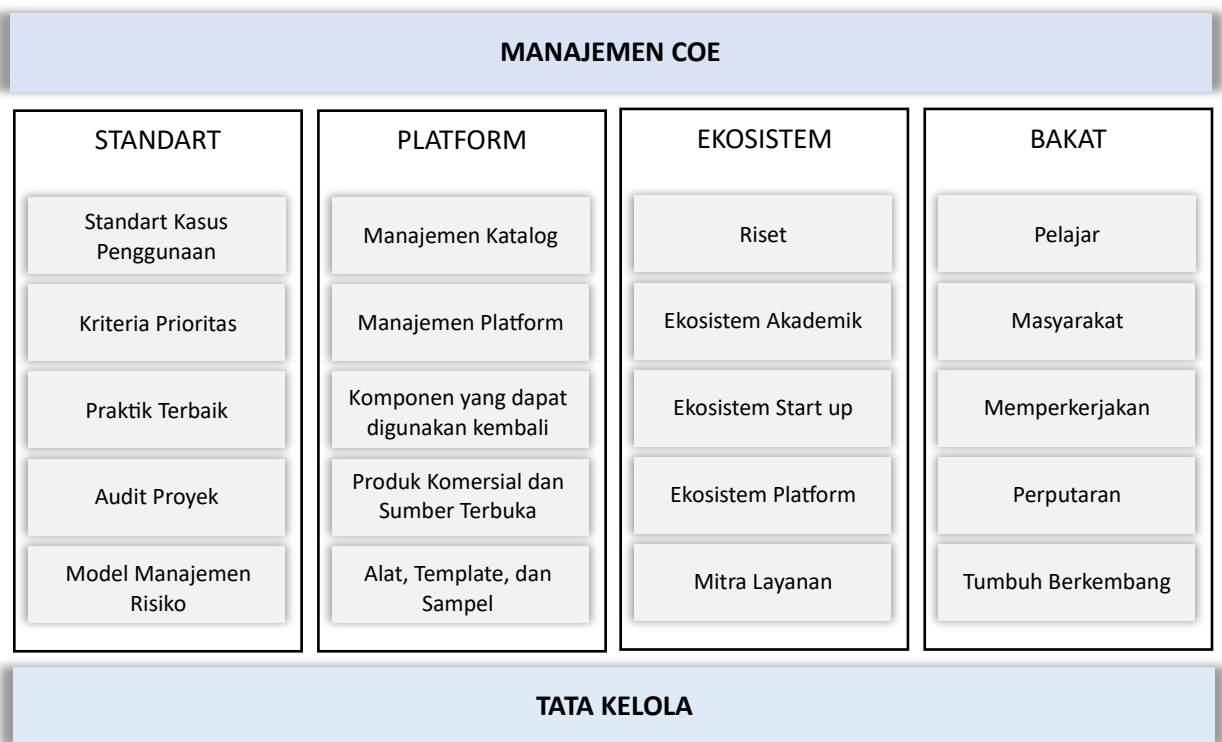
Pusat Keunggulan AI bertanggung jawab atas bagian-bagian pekerjaan AI yang diuntungkan dengan sentralisasi (lihat Gambar 11.2). CoE AI bertanggung jawab kepada manajemen puncak dan unit bisnis, yang untuknya CoE AI memberikan panduan dan dukungan. Beberapa contoh fungsi yang dicakup CoE AI di banyak organisasi meliputi:

- Edukasi karyawan dan mendapatkan dukungan mereka untuk menggunakan AI di dalam organisasi.
- Menetapkan standar untuk proyek AI di seluruh siklus hidup AI dan memastikan bahwa proyek mematuhi standar tersebut.
- Berbagi pengetahuan dan membangun komunitas AI.
- Menentukan dan mengoordinasikan pelatihan serta membantu merekrut talenta AI di seluruh perusahaan.
- Memiliki dan mengembangkan platform AI.
- Mempertahankan kesadaran dan menjadi titik sentral bagi partisipasi perusahaan dalam ekosistem AI eksternal, termasuk perusahaan teknologi besar, ruang startup AI, dan tren penelitian akademis AI.

Hal-hal ini dibahas lebih detail di subbagian berikut.

#### ***Standar dan Tata Kelola Proyek***

CoE AI bertanggung jawab untuk menetapkan standar tentang bagaimana proyek AI dijalankan dan untuk memastikan bahwa standar yang relevan dipatuhi untuk setiap proyek. Sering kali dibentuk oleh kepala AI, CoE dapat berperan penting dalam membangun momentum untuk inisiatif AI. CoE terdiri dari tim tata kelola yang terdiri dari para pemimpin termasuk dari unit bisnis organisasi, dan dirancang untuk memastikan bahwa tim proyek, yang didistribusikan antara CoE dan unit bisnis, berkolaborasi dan berbagi akuntabilitas, terlepas dari bagaimana peran dan tanggung jawab dibagi. Agar CoE AI berfungsi dengan baik, struktur tata kelola harus ada yang memastikan akuntabilitas ini dan menyediakan alat yang dibutuhkan setiap tim untuk membuat keputusan yang baik dan membangun solusi yang kuat. Struktur tata kelola yang terencana dengan baik mendorong kolaborasi antar anggota, memungkinkan mereka untuk berbagi pengetahuan dengan rekan-rekan mereka.



**Gambar 11.2** Fungsi utama dalam pusat keunggulan AI.

CoE AI harus diberi wewenang untuk mewajibkan proyek berpartisipasi dalam tinjauan proyek dan memberikan bukti bahwa standarnya dipenuhi. Dalam model terfederasi, di mana unit bisnis (BU) menetapkan prioritas untuk berbagai proyek AI mereka, AI CoE berfungsi sebagai fasilitator, memberikan bantuan untuk hal-hal seperti mengelola penugasan vendor untuk sumber daya terkait AI. Salah satu aspek kunci dari perannya adalah melihat komponen yang dapat digunakan kembali di seluruh BU dan memberikan saran tentang urutan proyek yang optimal untuk mendapatkan manfaat maksimal dari komponen bersama yang dibangun di atas platform.

Di awal perjalanan perusahaan, sebagian besar proyek AI ditinjau oleh AI CoE untuk membantu memastikan bahwa tim memiliki staf yang memadai dengan keterampilan yang tepat serta untuk memastikan bahwa semua jebakan yang dapat menggagalkan proyek AI diperhitungkan. Proyek-proyek awal ini menetapkan standar untuk memantau dan mengelola kinerja berbagai aplikasi AI yang dibangun. Setiap BU atau grup memantau aplikasi yang dimilikinya. Dalam bidang tata kelola model, CoE menetapkan kriteria dan menerbitkan serangkaian standar yang harus diikuti oleh tim, yang mencakup cara untuk menentukan apakah suatu tim proyek telah mematuhi standar. Tim kemudian harus menunjukkan kepada pusat keunggulan (atau kelompok lain dalam BU) bahwa mereka melakukan semua pekerjaan yang diamanatkan, yang mungkin mencakup pengumpulan dan penentuan prioritas kasus penggunaan, penilaian kualitas data, proses siklus hidup AI, pengujian kewajaran, memanfaatkan platform untuk pemodelan dan pemantauan, dan melakukan penilaian risiko model, mungkin dengan menunjukkan hasil pengujian itu sendiri.

### ***Komunitas, Pengetahuan, dan Pelatihan***

Untuk berhasil menerapkan strategi AI, tidak hanya dibutuhkan ilmuwan AI yang unggul dan memahami bisnis yang mereka geluti; tetapi juga dibutuhkan karyawan, mulai dari CEO hingga ke bawah, untuk memahami AI secara memadai dan apa yang ditawarkannya. Pemahaman ganda ini sangat penting untuk menyukseskan usaha AI. AI dapat menganalisis suatu proses secara mendalam, baik itu bagaimana pelanggan bereaksi terhadap kampanye iklan baru atau bagaimana kendaraan pengantar organ dapat menghemat waktu berharga dengan mengambil rute yang berbeda. Kapasitas ini membutuhkan lingkungan di mana setiap orang, mulai dari manajemen puncak hingga karyawan di lapangan, berpartisipasi dalam mengidentifikasi kasus penggunaan AI yang relevan. Banyak inovasi yang memungkinkan AI bermanfaat terjadi di dalam suatu proses, di mana karyawan memiliki pemahaman terbaik. Semakin baik pengetahuan mereka tentang AI dan semakin nyaman mereka dengan penggunaannya, semakin besar manfaat yang akan diperoleh perusahaan mereka.

Pusat Kepegawaian AI harus mendorong siklus umpan balik di mana pembelajaran bersama di seluruh organisasi dikumpulkan dan disebar untuk memperkuat kapabilitas AI unit bisnis dari waktu ke waktu. Siklus ini memastikan bahwa setiap orang, termasuk mereka yang berada dalam tim di dalam Unit Bisnis (BU), memiliki pengetahuan terbaru tentang apa yang berhasil, cara mengatasi tantangan spesifik, alat dan produk baru apa yang akan tersedia, dan apa saja praktik terbaik yang inovatif.

CoE AI harus berkoordinasi lintas BU untuk memahami dan menentukan kebutuhan pelatihan bagi organisasi. Pelatihan harus spesifik untuk setiap audiens. Pelatihan harus mencakup kesadaran umum dan cara kerja proyek bagi semua karyawan bisnis, cara menerapkan dan memanfaatkan model AI dalam aplikasi bagi tim TI, serta topik AI tingkat lanjut dan pengetahuan bisnis bagi tim AI. Pelatihan karyawan secara umum harus mencakup cara memahami seperti apa proyek AI; cara mengidentifikasi kasus penggunaan potensial; proses untuk memulai implementasi bagi kasus-kasus tersebut; dan peran apa yang harus dimainkan karyawan saat proyek bergerak dari konsepsi hingga penerapan dan seterusnya. Pendidikan AI terbaik adalah kombinasi pelatihan tatap muka dan daring dengan kecepatan mandiri. Namun, rencana ini tidak bisa hanya berupa daftar kursus pelatihan poin dan program eksekutif yang bersifat ad hoc. Pelatihan ini perlu didekati sebagai latihan pengembangan kapabilitas yang luas dan berskala penuh. Salah satu contoh terbaru adalah dari Amazon, yang mengatakan bahwa mereka berencana untuk menginvestasikan Rp 7 triliun untuk melatih karyawan. Bisnis harus mengumpulkan tim eksekutif untuk seminar. Para duta besar harus "terjun ke lapangan" untuk mendorong semua orang agar ikut serta. Pelatihan dan persiapan ini akan memobilisasi karyawan untuk memanfaatkan AI sebaik-baiknya.

Selain semua ini, CoE dapat menjadikan perusahaan tempat yang lebih menarik untuk bekerja, yang dapat membantu bisnis merekrut talenta AI terbaik yang tersedia. Seringkali, anggota tim AI dapat berotasi masuk dan keluar dari CoE AI saat dibutuhkan, yang memungkinkan mereka untuk berpindah antar unit bisnis dan dengan demikian memberi mereka berbagai masalah baru untuk dipecahkan. CoE juga membantu mempertahankan

komunitas praktik yang lebih kuat, dengan ilmuwan dan insinyur AI yang tersebar di seluruh unit bisnis dan di lokasi yang berbeda.

### ***Platform dan Ekosistem AI***

CoE AI, bersama dengan departemen TI, harus memiliki, memelihara, dan meningkatkan platform AI berdasarkan kebutuhan bisnis di seluruh Unit Bisnis (BU). CoE membantu mengoordinasikan komponen mana dari berbagai proyek yang harus dibangun sebagai komponen yang dapat digunakan kembali sehingga proyek-proyek mendatang dapat memanfaatkannya untuk mempercepat. Dalam beberapa kasus, CoE AI dapat mengambil alih proyek sebelum dibutuhkan dalam BU, terutama jika proyek tersebut melibatkan pembangunan komponen kompleks yang dapat dibagikan di seluruh BU. Lebih sering, CoE mengembangkan akselerator pada platform yang dapat digunakan setiap BU untuk mempercepat proyek mereka. Contohnya adalah kerangka kerja untuk manajemen risiko model atau alat untuk penemuan sinyal data otomatis.

Tidak ada perusahaan yang menggunakan AI yang berdiri sendiri, melainkan berada dalam ekosistem aktivitas AI. CoE AI harus menjaga hubungan dengan ekosistem ini untuk memastikan selalu membuat keputusan terbaik tentang alat, kerangka kerja, metodologi, algoritma, atau aplikasi. Manajemen ekosistem ini harus mencakup pemeliharaan hubungan dengan perusahaan teknologi besar seperti Microsoft yang menghabiskan miliaran dolar untuk meningkatkan platform AI mereka; dengan perusahaan rintisan yang memiliki aplikasi yang relevan dengan perusahaan; dengan perusahaan konsultan dan jasa yang dapat menyediakan anggota tim dan pengetahuan; dan dengan universitas serta lembaga penelitian lain yang sedang mengembangkan algoritma dan model baru untuk memecahkan berbagai jenis masalah baru. Keuntungan tambahan dari mempertahankan kemitraan universitas adalah bahwa hal itu merupakan cara lain untuk merekrut talenta terbaik, yang memungkinkan perusahaan untuk mengidentifikasi dan mengembangkan hubungan dengan ilmuwan AI yang menjanjikan. Memantau situs-situs seperti Kaggle, yang mempromosikan kompetisi antar tim AI amatir dan profesional, juga merupakan ide yang bagus. Kaggle menyediakan data yang identik kepada tim dan menantang mereka untuk membuat algoritma yang memecahkan masalah tertentu.

### **11.4 MENYUSUN TIM UNTUK EKSEKUSI PROYEK**

Dibutuhkan data dalam jumlah besar, rekayasa terbaik, pengalaman bisnis yang luas, dan banyak lagi untuk berhasil menciptakan aplikasi berbasis AI. Dibutuhkan juga tim yang berbakat. Tim AI biasanya dibentuk di awal proyek; dalam sistem terfederasi, tim ini merekrut talenta dari CoE, TI, dan Unit Bisnis.

Ada beberapa hal yang perlu diingat saat membentuk tim proyek. Penting bagi ilmuwan AI dan insinyur data untuk bekerja sama dengan insinyur dan analis bisnis lain agar proyek dapat berjalan optimal. Tim terbaik membina hubungan baik di antara spesialisasi ini. Tanpa hubungan ini, inisiatif AI dapat mengalami penundaan yang merugikan dan menghasilkan penurunan kinerja model. Ketika insinyur data memikul tanggung jawab untuk tugas-tugas

seperti persiapan data, ilmuwan AI bebas untuk fokus pada apa yang seharusnya mereka lakukan membangun dan memvalidasi model.

Tim yang terorganisir dengan baik juga lebih mampu menangani potensi hambatan. Banyak di antaranya terjadi di tahap akhir proyek masalah seperti ketidakcocokan alat atau tata kelola data, penerapan, atau masalah keamanan. Anggota tim juga mungkin kesulitan membuat pengguna akhir memahami dan memanfaatkan prediksi yang dihasilkan model mereka. Cara terbaik untuk menangani situasi seperti ini, tentu saja, adalah mengantisipasi dan menanganinya sebelum muncul. Tim yang baik mengintegrasikan solusi untuk masalah-masalah ini di awal proyek, berkoordinasi dengan para pemangku kepentingan untuk memastikan bahwa perencanaan proyek mempertimbangkan kebutuhan mereka; bahwa ekspektasi realistis; dan bahwa sumber daya yang memadai tersedia. Hal ini membutuhkan tim yang memiliki beragam keahlian dan memiliki keterampilan sejak awal.

Untuk membentuk tim yang efektif dalam pelaksanaan proyek, manfaatkan orang-orang yang paling andal dalam peran kepemimpinan. Mereka adalah individu yang dapat menggabungkan ketajaman bisnis, keahlian AI dan pembelajaran mesin, serta kapabilitas TI, sehingga memungkinkan mereka memahami kebutuhan dan mengusulkan solusi. Pemimpin biasanya adalah ilmuwan AI senior yang memahami kebutuhan bisnis dan memiliki ketajaman TI, atau pebisnis berpengalaman dengan pengalaman analitik kuantitatif dan data. Beberapa peran kunci yang dibutuhkan dalam tim proyek AI dijelaskan selanjutnya.

- *Seorang manajer tim AI* bertanggung jawab untuk memanfaatkan pengalaman dan keahlian timnya agar proyek dapat seproduktif mungkin. Ia harus memiliki keterampilan komunikasi untuk terhubung dengan personel nonteknis dan keahlian teknis yang memadai untuk berkomunikasi dengan anggota timnya, memahami apa yang mereka lakukan, dan mendukung mereka dengan cara apa pun yang diperlukan.
- *Ilmuwan AI* bertanggung jawab untuk mengeksplorasi data dan mengembangkan model serta algoritma untuk membangun dan menerapkan solusi intelijen yang optimal, menilai data untuk deteksi sinyal, dan mengidentifikasi pendekatan AI yang paling sesuai dengan kasus penggunaan. Mereka adalah individu yang dapat memahami tujuan bisnis untuk suatu kasus penggunaan dan mengubahnya menjadi tujuan tugas AI.
- *Insinyur AI (atau ML)* berspesialisasi dalam membangun dan menerapkan platform skala perusahaan dan produk cerdas, memproduksi model bukti konsep (PoC) dengan menggabungkan keterampilan rekayasa perangkat lunak dan pemodelan. Mereka menerapkan model pada pembelajaran mesin terdistribusi dan kerangka kerja data (misalnya Tensorflow, Spark, dll.), dan bertanggung jawab atas pelatihan ulang, pemantauan, dan pembuatan versi model yang diterapkan. Insinyur pembelajaran mesin adalah spesialis dalam infrastruktur dan operasi lingkungan pembelajaran mesin, dan mereka bekerja sama erat dengan insinyur data dalam mendukung proses ini. Beberapa insinyur pembelajaran mesin mampu membuat prototipe antarmuka pengguna akhir.

- *Insinyur data* bertanggung jawab untuk menangani tantangan penyerapan, penyimpanan, pemrosesan, dan kueri data berskala besar memungkinkan analitik dengan memanfaatkan teknologi big data. Insinyur data memahami dan merancang arsitektur big data. Mereka merancang dan membangun gudang data, danau data, dan aliran data pada sistem terdistribusi. Mereka bertanggung jawab untuk menjaga integritas data, sehingga memungkinkan ilmuwan AI untuk berfokus pada algoritma yang akan memanfaatkan data ini.
- *Analisis data* mengukur, menginstrumentasikan, dan mengeksplorasi data secara iteratif serta mengembangkan alat visualisasi dan pelaporan data untuk memungkinkan pengambilan keputusan berbasis data. Mereka memastikan bahwa data dikumpulkan dengan tepat, dan terkadang diminta untuk menginterpretasikan hasil analitik guna mendapatkan wawasan. Mereka juga membantu pengguna memahami cara menggunakan model yang ada.
- Staf *AI DevOps* biasanya merupakan bagian dari organisasi TI perusahaan. Sebagian besar organisasi TI sudah memiliki tim DevOps; mereka dapat dilatih dalam ML atau model ops, atau anggota tim lain dengan pengalaman MLOps yang sesuai dapat direkrut. Sangat penting untuk memastikan bahwa mereka semua mendapatkan pelatihan yang dibutuhkan untuk mendukung teknisi AI atau teknisi data. Mereka bertanggung jawab atas pengembangan, pengujian, dan rilis perangkat lunak yang cepat, mengawasi alur integrasi berkelanjutan dan penerapan berkelanjutan (CI/CD) otomatis, mendukung perbaikan bug, dan pemecahan masalah.
- *Manajer proyek teknis* menjembatani kesenjangan di seluruh tim. Mereka tidak perlu ahli di bidang TI, tetapi mereka harus memiliki keahlian teknis serta ketajaman politik untuk menjaga hubungan kerja yang baik dengan setiap fungsi yang terlibat, serta keterampilan manajemen untuk memastikan bahwa proyek mematuhi rencana.

Meskipun ini adalah beberapa peran yang lebih penting dalam proyek AI, mereka bukanlah satu-satunya fungsi yang diperlukan untuk membentuk tim yang lengkap. Serupa dengan proyek rekayasa lainnya, proyek AI merupakan kolaborasi lintas tim, termasuk analisis bisnis, pengembang dan insinyur full-stack, manajer produk, master SCRUM, arsitek solusi, analisis intelijen bisnis, arsitek data, dan insinyur visualisasi data. Sebagian besar organisasi memiliki peran serupa dalam proyek TI sehari-hari mereka, sehingga kami tidak membahasnya di sini.

## 11.5 MENGELOLA BAKAT DAN PEREKRUTAN

Ketika diminta untuk mendefinisikan apa itu ilmuwan AI, banyak perusahaan menggambarkan seseorang yang telah menguasai keterampilan pembelajaran mesin, memiliki wawasan bisnis yang cukup untuk memahami dengan tepat kasus penggunaan mana yang terbaik untuk dikejar, memiliki keterampilan komunikasi yang diperlukan untuk memotivasi tim, memiliki pengetahuan rekayasa data untuk dapat membangun jalur data yang kuat, memiliki pengetahuan DevOps yang cukup untuk menyiapkan dan memelihara infrastruktur, memiliki bakat seorang pengembang yang dapat membuat prototipe aplikasi pengguna akhir,

dan banyak lagi. Namun, kecil kemungkinan Anda akan menemukan semua keterampilan ini dalam satu orang.

Sebaliknya, perusahaan seharusnya berfokus pada membangun tim yang keahliannya saling tumpang tindih dan mencakup berbagai bidang. Tim dengan keahlian campuran menghadirkan beragam perspektif yang bermanfaat. Mempekerjakan ilmuwan AI yang merupakan bagian dari insinyur data, pengembang yang familier dengan rekayasa data, atau analis data dengan pengalaman bisnis lebih mudah dilakukan, dan individu yang memiliki beragam keahlian seringkali bekerja sama secara lebih efektif, karena keahlian mereka di lebih dari satu bidang menjadikan mereka kolaborator yang baik. Akan bermanfaat, misalnya, jika seorang ilmuwan AI tidak hanya menguasai pemrograman teknis dan pemodelan matematika, tetapi juga memiliki pemahaman tentang bisnis yang bersangkutan, sehingga ia dapat mengartikulasikan keputusan pemodelan yang ia buat sebagai potensi trade-off bisnis.

Jika organisasi tidak dapat menemukan ilmuwan AI yang serba bisa, ada solusi lain. Sebuah bisnis mungkin ingin mempekerjakan dua orang: satu yang dapat menangani algoritma dan satu yang dapat menangani strategi bisnis, karena tidak banyak orang yang memiliki keahlian di keduanya. Namun, masing-masing membutuhkan pengalaman di bidang yang lain. Alternatif lain adalah berkomitmen untuk melatih ilmuwan AI baru dalam strategi bisnis, keahlian yang sudah dimiliki organisasi. Mengingat meningkatnya permintaan akan ilmuwan AI, sebuah bisnis mungkin harus mempekerjakan seseorang yang belum memiliki banyak pengalaman di dunia nyata dan bahkan merasa tidak nyaman beroperasi di lingkungan bisnis. Lagipula, mereka adalah pakar AI, bukan bisnis, TI, atau penerapan. Sekolah pascasarjana mengajarkan teori dan sains AI, bukan cara mempraktikkannya di bisnis. Bantuan apa pun yang dapat diberikan oleh AI CoE atau rekan mereka di unit bisnis patut dihargai. Perusahaan mungkin juga ingin melatih setidaknya beberapa insinyur data mereka yang ada dalam bidang AI. Banyak kursus daring dan perguruan tinggi yang mengarah pada sertifikasi sebagai profesional AI, termasuk dari Microsoft, Coursera, dan banyak universitas.

Meskipun demikian, kita perlu mengakui bahwa, sebagaimana disebutkan dalam Bab 7, praktisi AI yang cangguh sangat langka. Jumlah lowongan pekerjaan jauh lebih besar daripada kumpulan bakat yang tersedia. Kelangkaan ini tidak hanya berarti gaji yang tinggi, tetapi juga membuat ilmuwan AI memiliki banyak pilihan. Kabar baiknya adalah mereka mencari hal yang sama dengan kebanyakan orang berbakat tingkat tinggi. Mereka menginginkan pekerjaan yang bermakna dengan beragam kasus penggunaan yang dapat mereka gunakan bakat mereka dan beragam masalah untuk dipecahkan. Mereka menginginkan kesempatan untuk melakukan penelitian mereka sendiri dan kemampuan untuk menyarankan dan mengimplementasikan proyek-proyek baru. Mereka menginginkan lingkungan kerja yang menyediakan hal-hal yang mereka butuhkan untuk sukses, termasuk akses ke banyak data, dan mereka tidak ingin menghabiskan sebagian besar waktu mereka untuk membersihkan data sebelum mereka dapat memulai pekerjaan mereka. Mereka membutuhkan akses ke platform AI swalayan sehingga mereka dapat beroperasi secara mandiri dan tidak harus bergantung pada departemen TI untuk menyiapkan lingkungan atau staf pengadaan untuk mendapatkan perangkat lunak.

Untuk merekrut seseorang yang memenuhi kriteria yang diperlukan, organisasi harus memberi tahu kandidat bahwa mereka akan bekerja di lingkungan yang kreatif dan kolaboratif dengan banyak orang cerdas lainnya, bahwa perusahaan serius dengan misinya untuk menciptakan program AI, dan bahwa akan ada anggaran yang memadai untuk proyek-proyek serta dukungan penuh dari manajemen puncak. Perusahaan juga harus menjelaskan dengan jelas bagaimana organisasi mereka berbeda dari pesaing. Misalnya, jika peluang di dalam perusahaan tidak mendukung apa yang ingin dilakukan oleh orang AI tersebut selanjutnya seperti beralih ke penelitian terapan atau mengerjakan produk baru orang tersebut kemungkinan akan mencari pekerjaan lain. Seorang ilmuwan AI yang baik menerima banyak tawaran; perusahaan perlu mengartikulasikan apa yang membedakan mereka agar dapat menutup kesepakatan.

Sayangnya, kekurangan ilmuwan AI bukan satu-satunya masalah yang dihadapi berbagai organisasi saat ini. Merekrut insinyur data mereka yang bertanggung jawab menyediakan data bersih dan andal yang dibutuhkan ilmuwan AI dan pihak lain di organisasi juga menjadi semakin sulit. Pencarian di Glassdoor menemukan bahwa jumlah lowongan pekerjaan insinyur data empat kali lebih banyak daripada jumlah lowongan untuk ilmuwan AI, dan gaji awal seringkali dimulai dari Rp 1.250.000.000 atau lebih per tahun. Perusahaan harus siap menghadapi proses pencarian yang aktif.

#### **Karier di Bidang AI**

Semakin banyak perguruan tinggi dan universitas yang menawarkan jurusan kecerdasan buatan. Ada beberapa universitas ternama, seperti Carnegie Mellon, Stanford, MIT, Columbia, dan Harvard, tetapi universitas negeri seperti Colorado State University dan Eastern Michigan University juga memiliki program studi yang ternama. Profesional yang terlatih dalam AI dapat berkarier sebagai insinyur pembelajaran mesin, ilmuwan data, ilmuwan riset, insinyur visi komputer, dan pengembang intelijen bisnis. Banyak perusahaan telah merekrut akademisi terkemuka untuk mendukung program AI mereka. Pada tahun 2013, Google merekrut ilmuwan komputer Geoffrey Hinton untuk memimpin upaya penelitian AI-nya. Facebook merekrut Yann LeCun, yang bekerja di bidang pembelajaran mesin dan visi komputer, untuk menjadi kepala ilmuwan AI-nya. Pada tahun 2015, Uber berhasil membawa kabur hampir 50 peneliti dan ilmuwan dari Pusat Rekayasa Robotika Nasional Carnegie Mellon, termasuk spesialis robotika dan otomasi terkemuka, David Stager, yang kini menjadi insinyur sistem utama<sup>3</sup> yang memimpin inisiatif Uber untuk membangun mobil swakemudi. Yoshua Bengio adalah seorang profesor di Université de Montréal dan penasihat Microsoft; Institut Algoritma Pembelajaran Montréal (MILA) miliknya menjalin kemitraan dengan IBM. Masih banyak contoh lainnya. Meskipun perguruan tinggi diperkirakan akan meluluskan 400.000 lulusan ilmu komputer (bukan hanya AI) pada tahun 2020, diperkirakan akan ada 1,4 juta lapangan pekerjaan yang harus diisi pada saat itu.

## 11.6 LITERASI DATA, EKSPERIMEN, DAN KEPUTUSAN BERBASIS DATA

Karena kemungkinan AI akan tumbuh menjadi bagian yang lebih besar dalam bisnis, budaya data yang sehat menjadi semakin penting. Memiliki kapabilitas AI atau analitik di dalam organisasi tanpa budaya data yang sesuai untuk memanfaatkannya sepenuhnya akan mengurangi kekuatannya. Para eksekutif dan Pusat Manajemen AI perlu bekerja sama untuk merencanakan transisi menuju pengambilan keputusan yang lebih berbasis data di dalam organisasi. Dalam masa transisi saat ini, ada tiga aspek budaya data yang perlu diperhatikan.

**Literasi data:** Informasi terlengkap dan terkini di dunia tidak berharga bagi karyawan yang tidak dapat membaca dan memanfaatkannya. Perlu ada rencana dan perangkat untuk melatih karyawan agar dapat memanfaatkan data yang akan menjadi bagian yang lebih umum dalam pekerjaan mereka. Peningkatan literasi data juga mengubah kecenderungan budaya mereka yang memperolehnya. Selain pelatihan, pasar data yang konsisten (lihat Bab 9) perlu diaktifkan bagi pengguna bisnis. Ketika pengguna bisnis, ilmuwan AI, dan insinyur TI menggunakan pasar data dan katalog data yang sama, mereka mulai berbicara dalam bahasa data yang sama. Cara lain untuk mendorong hal ini adalah dengan mempromosikan pemanfaatan lebih banyak data dalam rapat rutin. Jelaskan kepada peserta area-area di mana keputusan yang lebih baik dapat dibuat karena mereka mengandalkan data.

**Eksperimen:** Seperti yang telah disebutkan sebelumnya, dalam AI, iterasi merupakan komponen penting kesuksesan. Pendekatan uji-dan-pelajari terhadap eksperimen membantu menemukan kasus penggunaan, set data, dan model prediktif terbaik. Eksperimen yang sering juga mendorong budaya perusahaan yang lebih berpusat pada data. Perusahaan berbasis data yang paling sukses tidak takut membuat "kesalahan"; mereka terus-menerus menguji ide. Mengelola tim AI lebih seperti mengawasi banyak eksperimen daripada sekadar memberikan tugas langsung di mana hasil yang diharapkan sudah diketahui. Namun, eksperimen perlu dilakukan dengan teliti. Bereksperimen bukan hanya berarti mencoba sesuatu untuk melihat apakah berhasil atau tidak. Eksperimen membutuhkan desain eksperimen yang dipikirkan dengan matang, serta kelompok kontrol dan kelompok uji yang siap untuk belajar dari hipotesis. Pendorong utama untuk bergerak menuju budaya eksperimen adalah menciptakan pengetahuan baru. Semakin banyak eksperimen yang dapat dilakukan perusahaan, semakin mereka menjadi budaya belajar. Dan dengan "belajar", yang kami maksud bukanlah pelatihan berdasarkan pengetahuan yang sudah ada; melainkan menciptakan dan memanfaatkan pengetahuan baru, dari eksperimen. "Kegagalan" bukan hanya sebuah pilihan, tetapi juga penting untuk pembelajaran. Sebuah studi terbaru oleh Robert Wilson, asisten profesor psikologi dan ilmu kognitif di University of Arizona, dan rekan-rekannya, menemukan bahwa komputer mempelajari tugas paling cepat ketika mereka benar 85% dari waktu dan gagal 15% dari waktu.

Agar berhasil dalam eksperimen membutuhkan perubahan pola pikir dalam cara memandang eksperimen. Mungkin tidak perlu 9.999 percobaan untuk mencapai keberhasilan, tetapi diragukan bahwa percobaan pertama akan menemukan solusi optimal; karyawan perlu terbiasa dengan hal ini. Eksperimen adalah hipotesis yang dapat dibuktikan benar atau salah.

Jika tidak ada kemungkinan salah, maka itu bukanlah sebuah eksperimen. Pertimbangkan contoh sebuah perusahaan di mana tim manajemen memutuskan untuk mengubah budayanya agar menerima eksperimen. Hal itu menantang. Manajemen sendiri harus belajar bagaimana berbicara dan berperilaku berbeda, serta mengubah cara mereka meminta pertanggungjawaban karyawan. Butuh waktu. Namun pada akhirnya berhasil. Para eksekutif berhenti berbicara tentang kegagalan dan mulai berbicara tentang hipotesis yang divalidasi atau dibatalkan, dan hipotesis yang dibatalkan dipandang bukan sebagai kegagalan tetapi sebagai bantuan untuk membentuk eksperimen di masa depan. Perusahaan ini memahami bahwa tidak membiarkan orang gagal berarti tidak membiarkan orang bereksperimen dan berhasil, dan bahwa tanpa eksperimen, tidak ada hal baru yang akan dikembangkan termasuk algoritma AI baru yang berhasil.

**Keputusan berbasis data:** Budaya berbasis data yang matang dapat ditemukan di perusahaan-perusahaan yang keputusannya didukung oleh data dan algoritma. Ini bukan berarti perusahaan harus menghasilkan setumpuk laporan setiap hari atau hanya menggunakan perangkat intelijen bisnis. Kunci untuk menciptakan budaya berbasis data adalah meningkatkan intuisi dengan data dan model. Para ilmuwan menyebut pengambilan keputusan dengan cara ini sebagai pengambilan keputusan berbasis bukti. Namun, mereka tidak hanya mengandalkan data. Data perlu dikombinasikan dengan intuisi, kreativitas, fleksibilitas, pengalaman pribadi, dan perspektif netral agar sepenuhnya berharga bagi perusahaan.

Literasi dan eksperimen data diperlukan untuk bergerak menuju keputusan berbasis data yang lebih bermakna di seluruh perusahaan. Selain itu, pengguna perlu dilatih dan merasa nyaman dalam mengambil keputusan dalam ketidakpastian, karena sebagian besar model berbasis data bersifat probabilistik. Untuk membantu hal ini, AI CoE dan tim manajemen dapat merencanakan serangkaian intervensi untuk mendorong lebih banyak keputusan berbasis data. Hal ini dapat dimulai dengan manajemen yang meminta tim bisnis untuk menjelaskan data yang muncul dalam laporan yang disajikan kepada mereka (pasca-keputusan), kemudian berlanjut dengan meminta tim untuk menjelaskan mengapa suatu keputusan harus mengarah pada arah tertentu berdasarkan berbagai model berbasis data alternatif (pra-keputusan). Terakhir, percakapan dapat beralih ke pertanyaan tentang optimasi pertanyaan seperti, dari semua model yang memungkinkan, mengapa suatu opsi tertentu merupakan yang terbaik? Kita telah menyaksikan ketiga tahap ini berhasil mengarahkan perusahaan yang melek data menuju keputusan berbasis data.

## 11.7 KESIMPULAN

Di bagian buku ini, kami telah membahas cara mengembangkan strategi AI di seluruh perusahaan dan cara mewujudkannya dengan orang, proses, dan teknologi yang tepat. Di Bagian IV buku ini, kami akan membahas detail lebih lanjut tentang proses dan keputusan bisnis yang terlibat di dalamnya dengan contoh langsung sebuah model dan membahas lebih detail subkomponen arsitekturnya serta cara kerjanya dalam berbagai pola solusi.

## BAB 12

### ARSITEKTUR DAN POLA TEKNIS

Bab ini membahas arsitektur teknis untuk platform AI, memperluas deskripsi tingkat tinggi dari Bab 9 dan membahas lebih dalam subkomponennya. Untuk memahami cara kerja platform, kita akan melihat lebih detail empat lapisan inti yang telah kita ulas di sana serta elemen-elemen di dalam setiap lapisan. Lapisan-lapisan ini adalah data minder untuk manajemen data, model maker untuk eksperimen dan validasi model, inference activator untuk deployment dan model serving, dan performance manager untuk pemantauan produksi berkelanjutan. Komponen-komponen ini mendukung siklus hidup AI yang dibahas dalam Bab 8. Kami juga akan membahas pola desain tentang cara menggunakan platform dalam berbagai skenario solusi, termasuk untuk chatbot dan asisten virtual cerdas, mesin personalisasi dan rekomendasi, deteksi anomali, perangkat IoT fisik, dan tenaga kerja digital.

#### 12.1 ARSITEKTUR PLATFORM AI

Alih-alih merakit platform AI dari awal, umumnya digunakan platform berbasis pembelajaran mesin berbasis cloud yang tersedia secara komersial dan dikembangkan oleh perusahaan perangkat lunak terkemuka. Microsoft Azure, Amazon AWS, dan Google Cloud menyediakan platform dasar ini, antara lain. Namun, sebagian besar perangkat lunak yang tersedia secara komersial ini tidak memiliki semua komponen yang dibutuhkan ilmuwan AI selama siklus hidup AI. Jika komponen-komponen ini kurang, perusahaan dapat mengisi kekurangan tersebut dengan membangun komponen yang hilang secara internal atau mengakuisisinya sebagai alat sumber terbuka atau aplikasi yang tersedia secara komersial yang dapat diintegrasikan ke dalam platform dasar AI.

##### ***Pengelola Data***

Pengelola data digunakan untuk mengelola integrasi dan data suatu platform. Komponen ini adalah tempat data dikumpulkan dan dibersihkan, serta akses ke data tersebut dikontrol sesuai kebutuhan. Lapisan ini mencakup danau data, basis data, berkas data, dan gudang data, dengan masukan dari sumber informasi di seluruh perusahaan maupun eksternal.

*Sumber data* adalah lokasi utama tempat data dikumpulkan secara langsung. Ini dapat mencakup sistem manajemen hubungan pelanggan (CRM) organisasi, sistem perencanaan sumber daya perusahaan (ERP), sistem manajemen pesanan (OMS), log dari situs web atau aplikasi seluler pelanggan, perangkat atau sensor *Internet of Things* (IoT), kamera diam atau video, perangkat yang dapat dikenakan, atau sumber lainnya. Beberapa lapisan data dan integrasi bahkan mungkin berfokus secara eksternal: yaitu, menyerap data dari sumber yang berada di luar perusahaan. Ini dapat masuk melalui jalur batch atau streaming.

*Penyimpanan data mentah* adalah tempat semua data yang masuk disimpan tanpa modifikasi, baik data terstruktur, yang mungkin berarti berasal dari tabel basis data, termasuk data deret waktu, maupun tidak terstruktur, seperti berkas gambar dari drone atau satelit,

video, atau dokumen Word. Penyimpanan data mentah dimaksudkan untuk menyimpan data dari sistem sumber "apa adanya", tetapi validasi diperlukan untuk memastikannya konsisten dengan data dari sistem sumber. Validasi ini dapat mencakup validasi struktur dan format serta validasi jumlah rekaman sumber dan tujuan atau distribusi dan pembuatan profil data. Data yang tidak valid ditandai dan disimpan untuk analisis dan koreksi lebih lanjut.

*Alur data* dibuat untuk mengalirkan data dari berbagai sumber ke berbagai tujuan. Data tersebut kemudian diproses lebih lanjut menjalani sanitasi dan penyaringan, standardisasi, normalisasi, pemeriksaan integritas referensial, dan jenis kurasi lainnya agar data dapat digunakan untuk analitik hilir dan AI. Data ini kemudian disimpan di penyimpanan *data yang dikurasi*. Beberapa pemrosesan data atau transformasi berbasis logika bisnis dilakukan pada tahap ini, sekaligus membuat proses hilir lebih efisien.

*Lab data* adalah tempat ilmuwan AI mengumpulkan data mereka untuk digunakan di seluruh proses eksperimen dan pemodelan. Mereka kemudian membuat perubahan yang diperlukan, termasuk transformasi, pengurangan bias, penskalaan, atau langkah-langkah persiapan data lainnya yang diperlukan sebelum atau selama pemodelan. Lab data dapat dianggap sebagai lingkungan pengembangan bagi para pemodel, dan setidaknya terdapat satu orang, atau tim kecil yang bekerja di setiap lab.

Komponen *tata kelola data* mengelola alur kerja tata kelola dan pengelolaan data, serta memanfaatkan komponen *kualitas data* untuk memahami kualitas data dan menentukan tindakan yang diperlukan untuk memastikan bahwa penggunaan data tersebut konsisten dengan kebijakan perusahaan dan peraturan perundang-undangan yang berlaku.

Komponen *penyintesis data* digunakan untuk melakukan dua hal. Pertama, mensintesis data baru untuk eksperimen model ketika informasi aktual belum tersedia. Data sintesis adalah data yang dihasilkan secara terprogram, berbeda dengan data dunia nyata yang dikumpulkan. Kedua, mensintesis data jika terdapat kekhawatiran privasi terkait data yang ada. Data sintesis dihasilkan dengan distribusi statistik yang dikontrol secara presisi yang mencerminkan skenario dunia nyata. Seringkali, model data buatan disiapkan untuk meniru model data nyata. Model AI yang dilatih pada data sintesis pada akhirnya dapat dilatih ulang pada data dunia nyata setelah dikumpulkan.

Komponen *pelabel data* juga digunakan untuk memberi label pada set data yang ada guna mempersiapkannya untuk pembelajaran terawasi. Label adalah atribut target yang coba diprediksi oleh algoritma pembelajaran mesin. Komponen ini mengelola alur kerja untuk pelabelan manusia, tetapi mungkin juga memiliki komponen pelabelan otomatis. Komponen ini menyajikan fitur kepada pengguna yang kemudian memasukkan label (bentuk bebas atau dari set yang telah ditentukan sebelumnya). Pelabelan ini dapat dilakukan dengan data terstruktur, audio, gambar, dan tipe data lainnya. Misalnya, pelabel manusia mungkin diminta untuk menggambar batas di sekitar berbagai objek dalam sebuah gambar.

Ada dua danau data untuk menyimpan data dari pemrosesan. Danau data fitur menyimpan setiap fitur terkomputasi yang dikembangkan dari data yang dikurasi, dan yang dapat digunakan kembali. Ini termasuk embedding yang dibuat untuk data kategorikal seperti pelanggan atau produk. Embedding adalah representasi matematis dari item kategorikal,

seperti produk ritel, yang mengubah setiap item dan propertinya menjadi vektor numerik. Danau data wawasan menyimpan hasil komputasi model apa pun dari menjalankan model secara batch atau waktu nyata. Pengguna biasanya melihat wawasan batch melalui laporan atau visualisasi. Hasil dari keluaran model waktu nyata dikembalikan melalui lapisan API ketika model dipanggil melalui API, tetapi hasil ini juga disimpan di danau data wawasan untuk pemantauan kinerja model.

*Grafik pengetahuan* adalah grafik entitas, hubungan, dan informasi lain yang dikumpulkan dari data teks tidak terstruktur seperti dokumen Word dan berkas PDF. Grafik ini digunakan untuk menemukan informasi yang ada di dalam dokumen secara lebih efisien dan tepat. Seringkali, terdapat lebih dari satu grafik pengetahuan, tergantung pada tujuan pembuatannya misalnya, untuk menemukan jawaban yang ditanyakan kepada asisten virtual tentang keahlian orang-orang di perusahaan atau tentang kebijakan perusahaan.

*Pasar data* adalah tampilan komprehensif dari semua data dan fitur yang tersedia bagi ilmuwan AI (dan pengguna lainnya). Informasi ini mudah ditelusuri sehingga pengguna dapat memahami data apa yang tersedia dan mencakup deskripsi data, metadata terkait, informasi tentang hubungan antar elemen data, silsilah data, dan sebagainya. Dalam banyak kasus, informasi yang tersedia juga mencakup profil data seperti jumlah dan distribusi baris data. Ilmuwan AI serta analis bisnis, pengguna bisnis, dan pengembang kecerdasan bisnis (BI) menggunakan pasar data untuk memahami semua data yang tersedia di perusahaan. Pasar data ini merupakan kunci untuk mewujudkan literasi data dan budaya pengambilan keputusan berbasis data di dalam perusahaan.

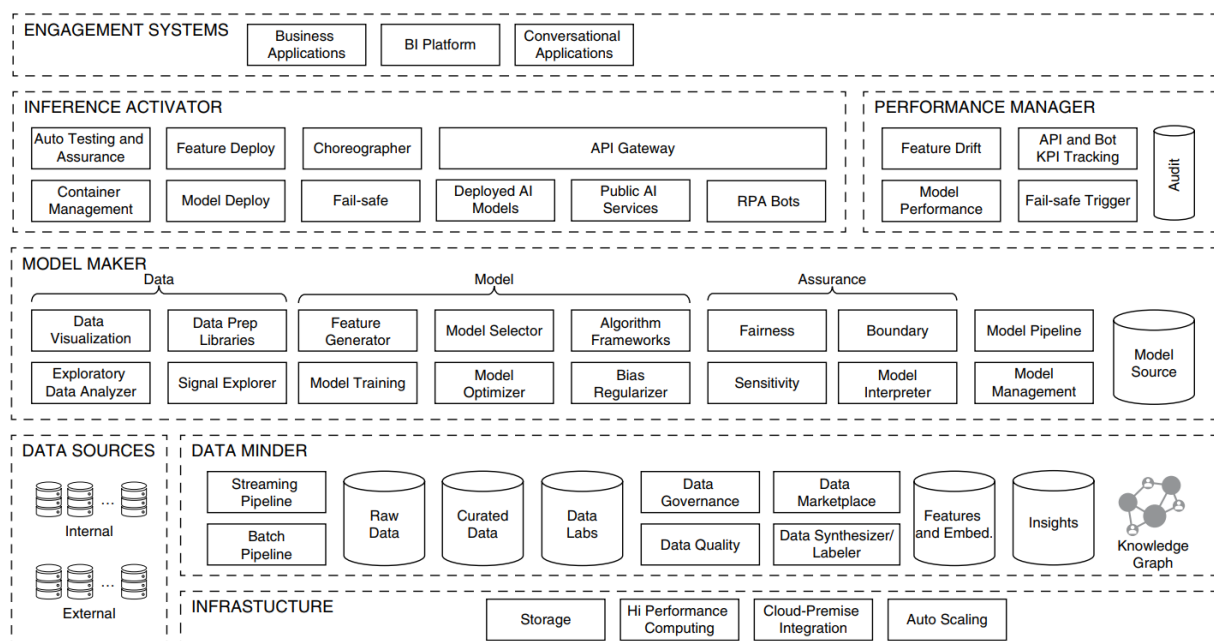
### ***Pembuat Model***

Pembuat model, juga dikenal sebagai lapisan eksperimen, adalah tempat ilmuwan AI mengembangkan, memvalidasi, dan mengulangi hipotesis mereka. Lapisan ini perlu mendukung mereka melalui pelatihan ratusan model sebelum mereka mencapai model yang ideal untuk kasus penggunaan tertentu. Lapisan ini juga akan membantu mereka melalui penilaian risiko model untuk setiap model dan dalam mengembangkan koreksi yang diperlukan. Pembuat model memiliki tiga kelompok komponen untuk digunakan ilmuwan AI (ditunjukkan sebagai tanda kurung kurawal pada komponen pembuat model pada Gambar 12.1) untuk bekerja dengan data, untuk membuat model, dan untuk jaminan model.

Komponen data dalam lapisan ini diperuntukkan bagi ilmuwan AI agar mereka dapat memahami kumpulan data mereka sebelum mereka dapat melanjutkan ke langkah-langkah pemodelan. Komponen visualisasi data adalah serangkaian pustaka yang memungkinkan pengguna untuk dengan mudah membuat berbagai visualisasi. Komponen persiapan data adalah serangkaian pustaka yang membantu imputasi data untuk data yang hilang dan mengonversi data ke format yang tepat untuk pemodelan. Penganalisis data eksploratori membantu menemukan pola, anomali, mencari bias dalam data, dan memahami berbagai ringkasan statistik data. Penjelajah sinyal menemukan korelasi dalam suatu set data ketika diberikan danau data dan kolom data target untuk diprediksi. Penjelajah sinyal mencari fitur lain yang berkorelasi dan kemungkinan merupakan prediktor untuk kolom target ini. Selain

waktu, keuntungan pendekatan otomatis dibandingkan pendekatan manual adalah kemampuan untuk menemukan pola yang tidak terduga dengan lebih mudah.

Ilmuwan AI menggunakan komponen pemodelan untuk mengembangkan model mereka. Kerangka kerja algoritme mencakup berbagai jenis algoritme yang dapat dilatih pada data untuk membuat model. Kerangka kerja seperti scikit-learn, TensorFlow, PyTorch, dan lainnya disiapkan dan dikelola di sini. Kerangka kerja ini mencakup algoritme yang mencakup simulasi dan optimasi, deteksi anomali, pengenalan pola, prediksi, penambangan pengetahuan, dan kemungkinan lainnya.



**Gambar 12.1** Komponen arsitektur untuk platform AI.

*Generator fitur* membuat dan mengevaluasi fitur turunan yang dihasilkan dengan menggabungkan fitur data yang ada. Misalnya, jika terdapat dua fitur dasar yang disebut tanggal mulai dan tanggal berakhir, generator fitur membuat fitur (durasi) yang merupakan selisih antara keduanya. Generator fitur kemudian menentukan apakah fitur turunan ini prediktif terhadap target atau tidak. Generator fitur serupa dengan penjelajah sinyal, tetapi menghasilkan dan mengevaluasi fitur turunan, alih-alih hanya mengerjakan fitur dasar.

*Pelatihan model* mencakup berbagai cara untuk mengonfigurasi model dan berbagai jenis pembelajaran yang dapat dimanfaatkan. Pemodelan mencakup konfigurasi aturan untuk RPA dan mesin inferensi, pelatihan batch untuk model pembelajaran mesin dan pembelajaran mendalam, pelatihan daring menggunakan pembelajaran penguatan, pembelajaran terdistribusi (atau terfederasi) untuk perangkat edge, atau kombinasi dari semuanya. Regulator bias digunakan dalam proses pemodelan selama pelatihan untuk memberikan penalti pada model karena bias tinggi dan akurasi rendah. Proses ini memastikan bahwa bias terkelola dan model tidak hanya mengoptimalkan daya prediktif terlepas dari bias. Ini adalah bagian dari langkah-langkah tata kelola risiko model yang dibahas dalam Bab 10. Pemilih model membantu membandingkan kinerja berbagai eksperimen untuk memilih yang optimal.

Pengoptimal model membantu menyetel hiperparameter model AI untuk kinerja terbaik. Ini menjalankan pelatihan dan evaluasi model beberapa kali. Ini juga melacak metadata dan menggunakan metode optimasi untuk menemukan pengaturan hiperparameter yang optimal untuk kinerja model. Tanpa komponen ini, banyak waktu dapat dihabiskan untuk menyetel hiperparameter secara manual, dan ini dapat menyulitkan untuk menemukan parameter berkinerja terbaik.

*Komponen manajemen* model membantu melacak model dan perubahannya untuk mendokumentasikan eksperimen model mana yang telah dilakukan sebelumnya dan cara mereproduksi hasil eksperimen secara akurat. Komponen ini menggunakan alat manajemen konfigurasi untuk melacak hal-hal seperti fitur apa yang diuji dan dibuang, modifikasi apa yang dilakukan pada jalur data, dan sumber daya komputasi apa yang tersedia untuk mendukung pelatihan yang memadai. Bersama dengan informasi konfigurasi, manajemen model mempercepat penerapan layanan AI yang konsisten sekaligus membantu mengurangi pekerjaan yang redundan. Sumber model adalah repositori untuk semua kode model. Alur model merupakan komponen tempat konfigurasi alur model, mulai dari data, pemodelan, hingga penerapan, dikelola.

### **Aktivator Inferensi**

Aktivator inferensi menerapkan model ke dalam produksi dan mengaktifkan AI selama proses inferensi run-time. Model biasanya dipanggil melalui panggilan API dari aplikasi bisnis menggunakan data yang bukan bagian dari pelatihan dan pengujian model. Kemudian, model dieksekusi untuk membuat prediksi. Pipeline DevOps integrasi berkelanjutan dan penerapan berkelanjutan (CI/CD) menampung berbagai alat yang mengelola penerapan model, termasuk pengujian otomatis dan jaminan model. Komponen pengujian dan jaminan menjalankan pengujian otomatis. Pengujian ini mencakup validasi model untuk memastikan kinerja model sesuai harapan di setiap lingkungan baru; pengujian menyeluruh untuk memvalidasi bahwa alur data dapat diakses dan API untuk model berfungsi; dan uji risiko model, seperti pengujian kewajaran, kondisi batas, dan sensitivitas, yang dibahas di Bab 10.

Komponen penerapan model mengemas dan menerapkan model AI, biasanya dalam kontainer untuk digunakan melalui API untuk penggunaan waktu nyata, atau dalam alur data untuk digunakan dalam mode batch. Kontainer adalah standar terbuka untuk mengemas dan mendistribusikan aplikasi, termasuk semua dependensi perangkat lunak, sehingga model dapat berjalan dengan cepat dan andal di lingkungan apa pun. Kontainer digunakan sehari-hari saat ini, baik untuk penerapan model maupun untuk penerapan dan penskalaan jenis perangkat lunak lainnya. Komponen penerapan fitur menerapkan kode untuk setiap transformasi fitur yang diperlukan pada data masukan. Terkadang, komponen ini terintegrasi ke dalam API model, tetapi terkadang terpisah.

Untuk AI dalam waktu proses, terdapat gateway API yang mengelola serangkaian API yang dapat digunakan aplikasi untuk memanggil atau memicu model AI yang telah dilatih sebelumnya. Ini mencakup model AI yang telah diterapkan dan dibangun di atas platform serta layanan AI publik yang tersedia dari penyedia cloud besar, seperti *Microsoft Cognitive Services*, berbagai perusahaan rintisan, dan banyak perusahaan kecil. Terakhir, bot RPA juga dapat

diakses melalui API sesuai kebutuhan. API dipanggil oleh aplikasi bisnis terhubung yang menggunakan model atau dari platform kecerdasan bisnis (BI) untuk mengumpulkan dan menyajikan informasi dalam laporan dan dasbor. Aplikasi ini biasanya memiliki pengguna manusia. API juga dapat diaktifkan untuk berbagai jenis mesin, seperti sistem otonom yang sinkron atau asinkron tetapi tidak memiliki pengguna manusia. Terkadang antarmuka pengguna aplikasi itu sendiri memerlukan API AI, seperti permintaan dari antarmuka pengguna percakapan di asisten virtual atau untuk kontrol gestur, pelacakan mata, pengenalan biometrik, dan antarmuka ethmoid (berbasis gerakan).

Koreografer dikonfigurasi ketika beberapa model perlu bekerja sama. Terkadang hal ini ditangani di dalam aplikasi bisnis di lapisan atas; di lain waktu, lebih baik membiarkan interaksi model tetap berada di lapisan inferensi, tergantung pada kasus penggunaan spesifik. Beberapa "produk cerdas" merupakan koreografi komposit dari beberapa API tingkat rendah misalnya, analisis sentimen dari berkas suara dapat mencakup koreografi antara API ucapan-ke-teks dan berbagai API pemrosesan bahasa alami (NLP).

Komponen failsafe memungkinkan administrator untuk mengonfigurasi mekanisme failsafe untuk model yang dianggap perlu menggunakan mekanisme tersebut. Komponen manajemen kontainer membantu mengelola kontainer yang diterapkan di belakang API model dan mengatur infrastruktur penyimpanan, jaringan, dan komputasi untuk memastikan aplikasi tersedia. Di sinilah konfigurasi dikelola untuk menskalakan jumlah kontainer untuk suatu model berdasarkan penggunaan CPU atau GPU, menyelesaikan tugas-tugas seperti memulai ulang atau mengganti kontainer yang gagal.

### **Manajer Kinerja**

Manajer kinerja ini memantau model AI run-time model yang sedang aktif digunakan. Komponen pergeseran fitur mengikuti distribusi data dari semua data masukan untuk setiap fitur dan membandingkannya dengan distribusi yang diketahui pada saat pelatihan model. Jika distribusi ini cukup berbeda (yaitu, melebihi ambang batas diferensial tertentu), maka model mungkin perlu dilatih ulang. Pergeseran merupakan indikator utama dari potensi masalah dalam akurasi model. Komponen kinerja model memantau apakah keluaran dan prediksi model masih memenuhi ambang batas yang ditentukan. Misalnya, apakah akurasi model masih pada atau di atas 90%? Terkadang, terdapat jeda waktu antara saat model memprediksi sesuatu dan saat peristiwa yang diprediksi terjadi, sehingga kinerja model sering kali merupakan indikator tertinggal dari kesesuaian model untuk digunakan.

Yang telah kita bahas sejauh ini adalah semua komponen dan subkomponen dari platform AI yang tangguh. Anda dapat menganggap ini sebagai tampilan statis arsitektur. Selanjutnya kita akan membahas tampilan dinamis.

## **12.2 POLA TEKNIS**

Bagian ini akan membahas lebih mendalam berbagai pola teknis untuk beberapa aplikasi umum AI dan pembelajaran mesin, serta bagaimana aplikasi ini dapat diimplementasikan dalam lingkungan produksi. Pola teknis yang menjelaskan bagaimana

platform digunakan ini sering muncul dalam banyak kasus penggunaan. Anda dapat menganggapnya sebagai tampilan dinamis dari arsitektur tersebut.

### **Asisten Virtual Cerdas**

Asisten virtual cerdas adalah perangkat lunak yang memiliki antarmuka pengguna berbasis percakapan atau pesan yang mendukung pelanggan dan karyawan dalam melakukan berbagai tugas. Contoh umum termasuk Siri dari Apple dan Alexa dari Amazon. Lebih banyak lagi yang sedang dikembangkan oleh bisnis untuk membantu karyawan melakukan hal-hal seperti mendapatkan jawaban langsung atas pertanyaan investasi, dan untuk mendukung pelanggan melakukan hal-hal seperti mendapatkan akses percakapan ke rekening bank dan memberikan saran.

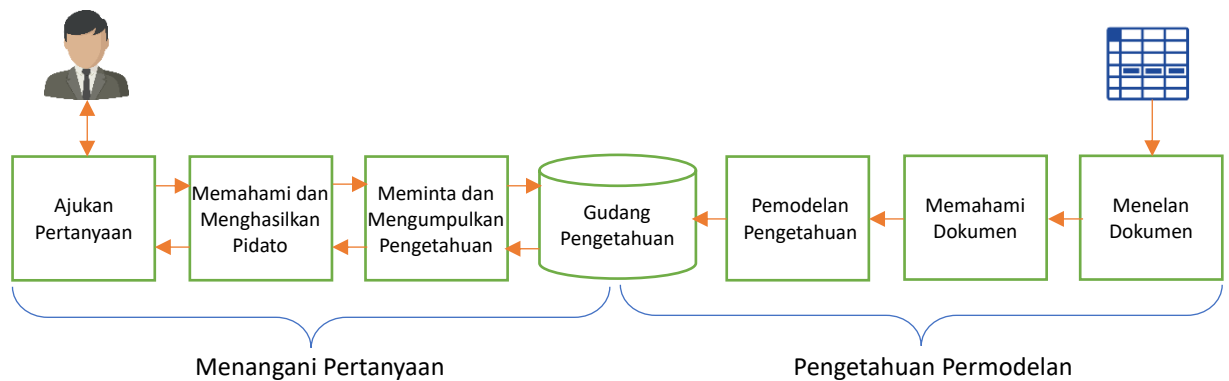
Agar berfungsi dengan baik, asisten virtual memerlukan beberapa komponen inti. Pertama, kemampuan untuk memahami perintah atau pertanyaan pengguna. Jika input berasal dari suara, perintah suara perlu diterjemahkan menjadi teks melalui model ucapan-ke-teks. Kemudian, pemrosesan bahasa alami perlu diterapkan pada teks untuk memahami maksud pembicara. Misalnya, jika seorang pembicara bertanya, "Berapa harga saham Microsoft?", asisten virtual harus mampu menafsirkan maksud pengguna harga saham dan entitas yang terkait dengan maksud tersebut Microsoft.

Kerumitan dalam memahami pertanyaan meningkat seiring dengan permintaan pengguna. Pertanyaan tersebut dapat memiliki beberapa entitas, seperti "Berapa harga saham Microsoft dan Google?" atau dapat memiliki entitas gabungan seperti "sektor teknologi". Permintaan juga dapat memiliki beberapa maksud, seperti "Kapan penerbangan berikutnya ke Chicago tersedia, dan bagaimana cuaca di sana?"

Setelah sistem memahami permintaan, langkah selanjutnya adalah menentukan respons atau memberikan jawaban. Bahasa Indonesia: Jika permintaan tersebut adalah perintah seperti "Kecikkan volume," maka API dapat dipanggil untuk menindaklanjutinya. Jika itu adalah pertanyaan yang memerlukan jawaban, hal itu dapat ditangani dengan berbagai cara. Untuk pencarian data terstruktur, seperti harga saham Google, panggilan API atau pencarian basis data mungkin memberikan solusi yang memadai. Jika jawabannya terletak pada data yang tidak terstruktur, maka itu perlu berasal dari solusi pemodelan pengetahuan. Contoh model pengetahuan adalah grafik pengetahuan yang menyusun teks dari dokumen. Anda dapat melihat hasil dari grafik pengetahuan jika Anda mencari sesuatu seperti "Berapa tinggi Barack Obama?" Meskipun mesin pencari Google menyediakan tautan ke dokumen yang berisi pertanyaan ini atau kata kunci darinya, di bagian atas halaman, ia juga memberikan jawaban dari grafik pengetahuannya.

Pemodelan pengetahuan melewati beberapa langkah (lihat Gambar 12.2) untuk mengembangkan repositori pengetahuan yang dapat dibaca mesin. Proses ini dapat mencakup memecah dokumen asli menjadi segmen-segmen yang lebih kecil, kadang-kadang disebut "unit jawaban"; mengekstraksi topik diskusi dan menandai bagian tersebut dengan topik-topik tersebut; mengekstraksi entitas dan konsep dari setiap kalimat dalam dokumen; dan membangun grafik pengetahuan entitas dan konsep tersebut beserta hubungannya. Dengan

pendekatan ini, Anda dapat menemukan jawaban yang lebih tepat dalam grafik pengetahuan atau unit jawaban.



**Gambar 12.2** Sistem tanya jawab yang dibangun berdasarkan pemodelan pengetahuan.

Terakhir, agar asisten virtual berfungsi, interaksi pengguna perlu dikelola. Fungsi ini mencakup mengajukan pertanyaan klarifikasi, mengonfirmasi apa yang akan dilakukan asisten virtual untuk menangani permintaan, dan memutuskan kapan harus berbicara atau merespons atau bagaimana menyajikan informasi kembali. Dalam beberapa kasus, asisten perlu membantu pengguna melalui percakapan terpandu misalnya, jika nasabah mengajukan pinjaman hipotek. Dalam kasus ini, asisten virtual juga perlu mengetahui langkah apa yang telah diambil dan apa yang perlu dilakukan selanjutnya.

Tujuan pola asisten virtual cerdas adalah berinteraksi dengan pengguna secara alami, menggunakan bahasa alami melalui suara atau teks dan sentuhan. Dalam aplikasi yang lebih baru, perusahaan juga menggabungkan gestur, realitas tertambah, dan pelacakan mata sebagai mode input. Mengingat pengguna dapat mengajukan pertanyaan apa pun dengan berbagai cara, perusahaan perlu membangun solusi di mana pelatihan NLP dapat berlangsung cepat dan berkelanjutan tanpa harus menyusun pemodelan pengetahuan secara manual. Pembelajaran aktif oleh sistem harus diintegrasikan ke dalam solusi agar dapat dilatih dengan cepat untuk pertanyaan apa pun yang belum dipahaminya.

### **Mesin Personalisasi dan Rekomendasi**

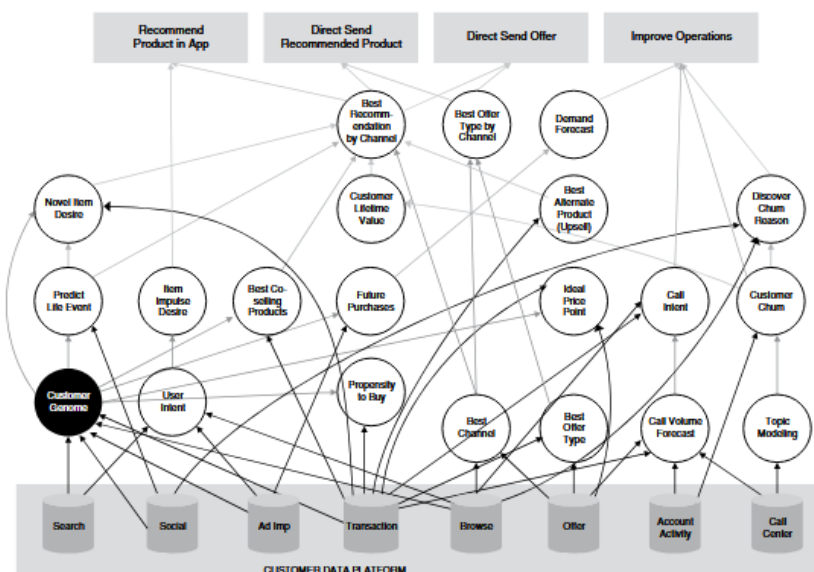
Personalisasi adalah tentang menciptakan pengalaman individual untuk setiap pengguna yang berinteraksi dengan bisnis, biasanya secara digital. Dulu, hal ini hanya tentang mengirimkan barang yang tepat kepada orang yang tepat, tetapi kini, jauh lebih banyak yang dapat dilakukan. Kini, personalisasi umumnya digambarkan sebagai menyediakan barang yang tepat (pesan, konten, atau produk) kepada orang yang tepat pada waktu yang tepat menggunakan saluran yang tepat.

Agar hiperpersonalisasi semacam ini berhasil, perusahaan perlu mengambil beberapa langkah. Pertama, berbagai jenis informasi dikumpulkan untuk setiap pelanggan sehingga wawasan dapat diekstraksi untuk memahami pelanggan tersebut dengan cara yang lebih bermanfaat. Data ini dapat mencakup profil dan demografi, data transaksi historis, data dari penggunaan aplikasi seluler atau situs web perusahaan oleh pelanggan, data tayangan dari iklan, dan kemungkinan jenis data pihak ketiga lainnya. Semua data ini perlu digabungkan

menggunakan ID pelanggan. Jenis penyimpanan data ini disebut platform data pelanggan (CDP).

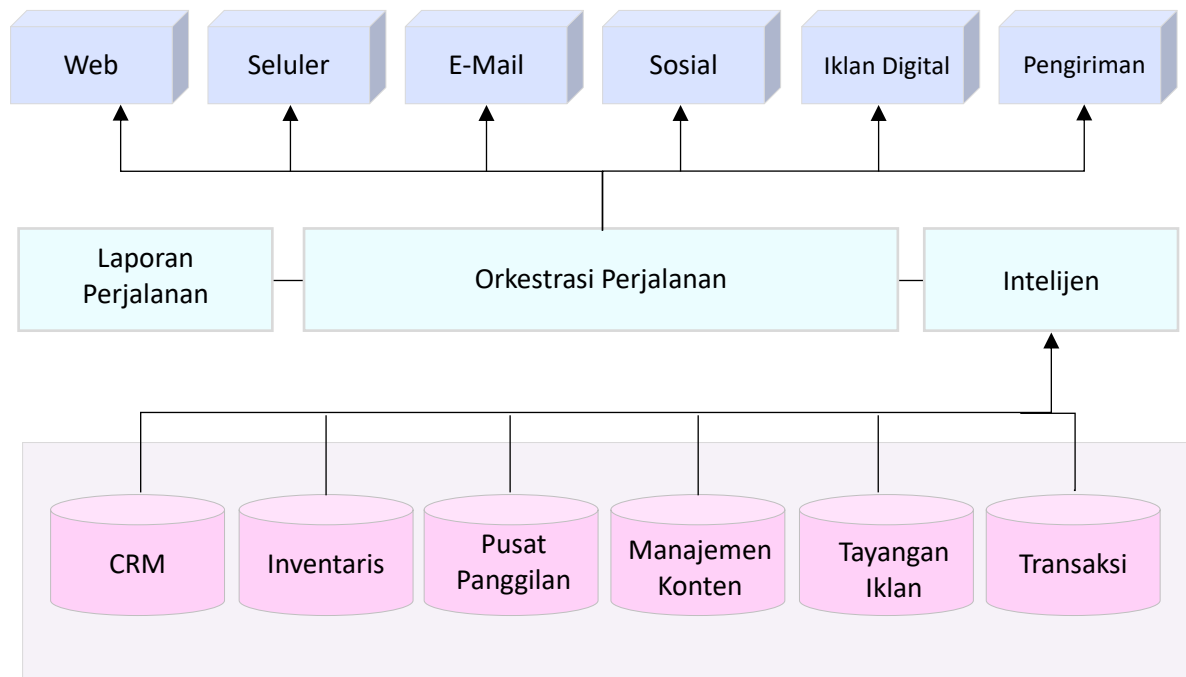
Berikutnya, beberapa model AI dikembangkan dan digabungkan untuk meningkatkan pemahaman bisnis tentang setiap pelanggan. Model-model ini mungkin mencakup prediksi tentang produk mana yang kemungkinan besar akan dibeli konsumen. Apa sensitivitas mereka terhadap harga produk? Apakah ada produk alternatif yang mereka sukai? Apakah ada produk tambahan yang mereka sukai, mengingat apa yang mereka beli? Saluran apa yang paling baik merekaanggapi email, iklan, pesan? Kapan waktu terbaik untuk menghubungi mereka, dan kapan mereka paling mungkin merespons secara positif? Bahasa apa dalam iklan atau email yang menghasilkan tingkat respons terbaik? Model-model ini digunakan bersama-sama; misalnya, tergantung pada produk yang akan direkomendasikan (dari model 1), saluran yang optimal (model 2), waktu (model 3), dan bahasa yang akan digunakan dalam pesan (model 4), pesan yang sesuai ditampilkan kepada pengguna. Gambar 12.3 menunjukkan contoh serangkaian model (kotak hitam) dalam bisnis ritel yang digunakan untuk menciptakan interaksi yang tepat dengan pelanggan (kotak abu-abu di bagian atas).

Akhirnya, interaksi dengan pelanggan diatur melalui titik sentuh yang paling relevan melalui aplikasi orkestrasi perjalanan pelanggan. Model AI menghasilkan wawasan dan membuat rekomendasi untuk langkah selanjutnya yang paling tepat. Orkestrasi mengambil tindakan atas hal ini dengan menerima wawasan atau rekomendasi dan memicu respons yang tepat dalam sistem lain (lihat Gambar 12.4). Ini dapat mencakup hal-hal seperti mengirim email kepada pelanggan dengan penawaran atau menampilkan rekomendasi produk ketika pelanggan sedang menjelajahi situs web atau aplikasi perusahaan. Karena ada banyak model yang diekspos melalui API dan banyak titik sentuh perjalanan pelanggan untuk diatur, tidak menggunakan orkestrasi perjalanan menciptakan ratusan koneksi titik-ke-titik yang menjadi mustahil untuk dipertahankan dalam jangka panjang. Untuk alasan ini, solusi orkestrasi perjalanan sangat ideal.



**Gambar 12.3** Memanfaatkan berbagai model untuk hiperpersonalisasi.

Cara paling umum untuk membangun rekomendasi adalah dengan pemrosesan batch. Namun, metode ini memiliki kekurangan. Salah satunya adalah data dapat menjadi usang di antara sesi. Kekurangan lainnya adalah rekomendasi mungkin didasarkan pada seluruh riwayat pelanggan, alih-alih hanya sesi mereka saat ini, yang dapat menunjukkan apa yang sedang mereka inginkan saat itu. Karena alasan ini, semakin banyak perusahaan yang membangun mesin rekomendasi real-time. Pertanyaan tentang batch versus real-time dibahas di Bab 9.



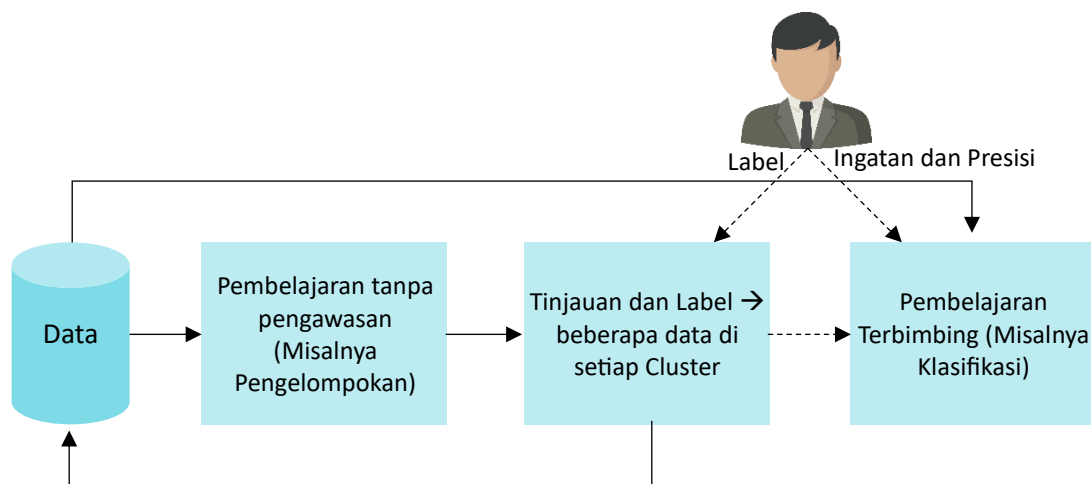
**Gambar 12.4** Mengorkestrasi interaksi personalisasi.

### **Deteksi Anomali**

Deteksi anomali secara teratur digunakan untuk mengidentifikasi data tertentu yang berbeda secara signifikan dari data lainnya dalam suatu kumpulan data. Mendeteksi penipuan, mendiagnosis tumor kanker, melakukan analisis risiko, dan mengidentifikasi pola dalam data untuk memfasilitasi wawasan hanyalah beberapa aplikasi di mana deteksi anomali digunakan. Untuk dapat menemukan anomali, diperlukan deskripsi yang spesifik dan terperinci tentang seperti apa data "normal". Kemudian, skor anomali dihitung untuk setiap sampel data baru, dengan membandingkan sampel ini dengan norma model. Jika deviasi melebihi ambang batas yang telah ditentukan, sampel data tersebut dianggap sebagai outlier atau anomali.

Mendeteksi anomali dimulai sebagai tugas tanpa pengawasan, karena anomali muncul dari pola data yang tidak diketahui, sehingga tidak ada data berlabel untuk dipelajari. Namun, deteksi anomali tanpa pengawasan seringkali gagal memenuhi tingkat deteksi yang disyaratkan dalam banyak kasus penggunaan bisnis. Dalam kasus tersebut, data berlabel diperlukan untuk menyempurnakan model. Hal ini dapat menjadi sangat penting; Dalam beberapa kasus, seperti deteksi kanker, model yang sangat akurat dapat menentukan hidup dan mati.

Inti dari keberhasilan deteksi anomali terletak pada penggunaan alur kerja (lihat Gambar 12.5) yang dimulai dengan pembelajaran tanpa pengawasan, pengelompokan berbagai kelompok data, serta peninjauan dan pelabelan beberapa item dalam kluster. Kemampuan untuk melakukan iterasi cepat melalui proses ini menyediakan data berlabel yang cukup untuk memungkinkan pembelajaran terawasi untuk pola anomali yang telah diidentifikasi dan diberi label sebelumnya, sementara proses tanpa pengawasan terus menemukan pola anomali baru. Platform yang memungkinkan pelabelan data yang cepat, eksperimen, dan penerapan model idealnya secara otonom sangat penting untuk penskalaan deteksi anomali yang sukses.



**Gambar 12.5** Aktivitas untuk deteksi anomali.

Salah satu tujuan deteksi anomali adalah menghindari alarm palsu sekaligus mengidentifikasi sebanyak mungkin anomali yang sah. Untuk menentukan akurasi deteksi, kami menggunakan dua ukuran: recall (jumlah titik anomali sejati yang diprediksi dibagi dengan jumlah total titik anomali sejati baik positif sejati maupun negatif palsu) dan presisi (jumlah titik anomali sejati yang diprediksi dibagi dengan jumlah titik anomali yang diprediksi baik positif sejati maupun positif palsu). Pengukuran recall dan presisi perlu dikalibrasi secara akurat dari waktu ke waktu untuk mendeteksi anomali secara efektif. Untuk memperluas fungsi ini agar dapat digunakan dalam deteksi anomali waktu nyata, kami membuat "skor anomali" yang menunjukkan "jarak" (dalam ruang data) data masuk dari data standar maupun anomali. Dengan menggunakan jarak ini, data masuk dapat diberi label anomali atau normal misalnya, dalam mendeteksi penipuan kartu kredit pada saat transaksi.

### ***Penginderaan Sekitar dan Kontrol Fisik***

Ada banyak contoh penginderaan sekitar saat ini, yang semakin banyak mengendalikan sistem fisik. Dalam bidang manufaktur, kita dapat menggunakan jenis kemampuan ini untuk skenario kesehatan dan keselamatan yang dapat mencakup menjaga keselamatan seseorang dengan memberi tanda ketika ia terlalu dekat dengan kondisi yang tidak aman, atau mengurangi beban kehilangan barang dengan memberi tanda di mana barang yang hilang mungkin berada beserta petunjuk tentang cara menemukannya. Namun, banyak aplikasi yang

melampaui lingkungan industri. Kemampuan ini memungkinkan terciptanya beberapa pengalaman ritel yang menarik, termasuk belanja siap pakai dan bantuan langsung di tempat saat dibutuhkan. Ada banyak skenario menarik yang dimungkinkan oleh pola ini yang baru saja muncul.

Langkah-langkah yang perlu diambil bisnis saat menerapkan AI untuk penginderaan ambien dan kontrol fisik lebih kompleks daripada langkah-langkah yang diperlukan untuk menyiapkan sistem AI bagi perusahaan lain. Hal ini dikarenakan diperlukan integrasi yang jauh lebih erat antara aspek digital dan fisik dari operasional. Pola ini mencakup penggunaan pembelajaran mesin dan pendekatan AI lainnya di mana sistem berinteraksi dengan dunia fisik dengan cara tertentu, memanfaatkan *Internet of Things* (IoT). IoT adalah kumpulan perangkat yang berisi perangkat elektronik, sensor, aktuator, dan perangkat lunak yang menghubungkan, mengumpulkan, dan bertukar data. Perangkat otomatisasi rumah, termasuk aksesoris seperti termostat Nest, merupakan contoh IoT yang sangat baik.

Sistem IoT terdiri dari perangkat edge perangkat yang ditempatkan lebih dekat dengan mesin (misalnya robot di pabrik mobil) daripada dekat dengan server di jaringan inti perusahaan. Perangkat ini memiliki sensor yang mendeteksi suhu, kelembapan, tekanan, gas, cahaya, suara, identifikasi frekuensi radio (RFID), yang memungkinkan informasi yang tersimpan pada label yang terpasang pada suatu objek untuk dibaca, dan komunikasi jarak dekat (NFC) yang memungkinkan perangkat pintar berkomunikasi dengan perangkat elektronik lainnya. Perangkat lain seperti sensor ultrasonik, pengukur aliran, dan kamera juga dapat digunakan.

Aliran informasi dapat satu arah, seperti kamera cerdas yang melakukan pengenalan wajah di lokasi (di tepi jaringan), atau dapat dua arah, seperti perangkat termostat IoT yang mengukur suhu sekitar dan menyesuaikan termostat. Perangkat tepi juga dapat mencakup aktuator untuk mengendalikan mesin. Aktuator ini bekerja bersama dengan sakelar, relai, komputer industri yang diadaptasi untuk digunakan dalam pengendalian proses manufaktur yang dikenal sebagai pengontrol logika terprogram (PLC), dan motor. Terkadang, perangkat tepi bertindak sebagai sensor sekaligus aktuator.

Penggunaan paling umum dari pola ini adalah untuk memungkinkan sistem menyelesaikan tugas, mencapai tujuan, atau berinteraksi dengan lingkungan sekitarnya dengan keterlibatan manusia yang minimal atau tanpa keterlibatan manusia. Pola ini dapat digunakan untuk mengendalikan perangkat keras sistem secara otonom, sehingga meminimalkan tenaga kerja manusia, atau untuk membuat keputusan atau penyesuaian dengan cepat dalam sistem yang membutuhkan kecerdasan di mana intervensi manusia akan terlalu lambat. Termostat pintar, kendaraan otonom, mesin otomatis di dalam pabrik, serta mesin dan robot lainnya adalah contoh penggunaan pola ini, dan sejumlah produk pintar lainnya telah diproduksi, mulai dari yang biasa (misalnya, tempat makan hewan peliharaan) hingga yang eksotis (misalnya, rangka luar manusia). Semakin banyak perangkat, baik besar maupun kecil, yang bergabung dengan dunia IoT setiap tahun.

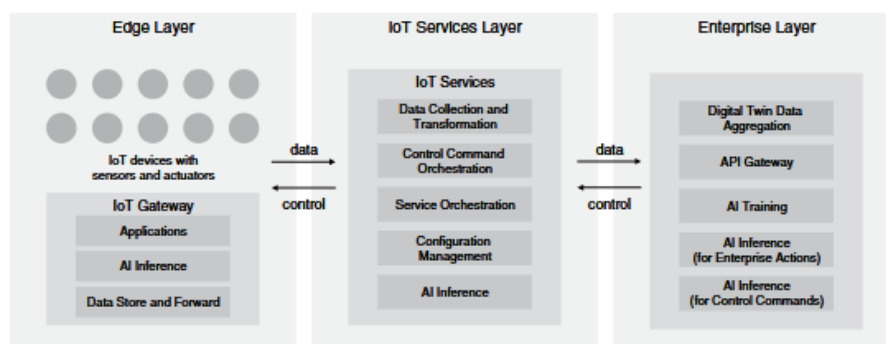
Bagian penting dari pola ini adalah menentukan sifat suatu objek ketika diberikan beberapa bentuk data terstruktur atau tidak terstruktur. Data ini dapat mencakup gambar,

video, audio, atau data sensor IoT; tujuannya adalah agar beberapa aspek data diberi label dan ditandai dengan mengidentifikasi, mengenali, atau mengklasifikasikannya. Kasus penggunaan meliputi pengenalan objek dan gambar, pengenalan wajah, pengenalan suara dan audio, pengenalan tulisan tangan dan teks, deteksi gestur, dan klasifikasi perilaku apa pun yang berbasis gerakan.

Arsitektur IoT secara keseluruhan terdiri dari tiga lapisan (lihat Gambar 12.6): lapisan tepi, lapisan layanan IoT, dan lapisan perusahaan. Lapisan tepi terdiri dari perangkat tepi, seperti sensor, dan gateway IoT: perangkat atau program perangkat lunak yang menghubungkan cloud ke pengontrol, sensor, dan mesin cerdas. Perangkat dapat didistribusikan di berbagai lokasi tergantung pada kasus penggunaan, yang membatasi komunikasinya ke jaringan jarak pendek karena keterbatasan konektivitas, pemrosesan, dan daya.

Gerbang IoT berperan ketika perangkat perlu berkomunikasi dengan seluruh dunia. Gerbang IoT biasanya berisi penyimpanan data untuk data perangkat IoT, satu atau beberapa layanan untuk menganalisis data yang datang langsung dari perangkat atau dari penyimpanan data, dan mengontrol tindakan berdasarkan data yang masuk. Lapisan IoT juga mungkin memiliki kecerdasan tepi dengan berbagai tingkat kemampuan pemrosesan, yang memungkinkan model pembelajaran mesin untuk menarik inferensi jika tersedia daya komputasi yang cukup.

Data dari lapisan tepi dikonsolidasikan, diproses, dan dianalisis oleh lapisan layanan IoT. Lapisan ini juga mengelola perangkat, aset fisik, dan proses serta meneruskan perintah kontrol dari lapisan perusahaan ke lapisan tepi. Data dari lapisan tepi dan layanan IoT diterima oleh lapisan perusahaan, yang merupakan lapisan yang telah kita fokuskan di bab-bab sebelumnya. Model yang terlatih di lapisan perusahaan kemudian memicu reaksi terhadap data sensor baru yang masuk dan mengirimkan tindakan kembali ke aktuator, mengirimkan pemicu sederhana seperti peringatan. Pemicu yang lebih kompleks termasuk mengeluarkan perintah kontrol ke dua lapisan lainnya yang memodifikasi instruksi, tugas, atau proses utama atau subsistem.



**Gambar 12.6** Pola interaksi untuk IoT dan perangkat edge.

Model pembelajaran mesin paling sering dilatih di lapisan perusahaan, setelah data dikumpulkan di sana. Inferensi model juga dapat terjadi di lapisan ini, seperti yang dijelaskan sebelumnya, atau model dapat diinstal pada perangkat edge untuk merespons di lingkungan

lokal. Pendekatan baru, yang disebut pembelajaran terfederasi, saat ini sedang dieksplorasi untuk melatih model ML di seluruh perangkat edge tanpa membawa semua data kembali ke lokasi pusat. Di lapisan perusahaan, sangat penting untuk memastikan metadata tetap akurat. Mengetahui, misalnya, sensor mana yang terhubung ke komponen apa, komponen mana yang merupakan bagian dari mesin yang sama, dan mesin mana yang merupakan bagian dari pabrik yang sama diperlukan untuk mengembangkan kembar digital.

### **Tenaga Kerja Digital**

Pola ini menggunakan kombinasi otomatisasi proses robotik (RPA) dan pembelajaran mesin untuk menciptakan tenaga kerja virtual. Karyawan virtual ini mengikuti proses yang telah ditentukan dan didokumentasikan dengan tepat, tanpa kesalahan, kelalaian, atau penyimpangan. Memanfaatkan tenaga kerja digital semacam ini mengurangi biaya operasional dan pengerjaan ulang, serta memungkinkan terciptanya tenaga kerja sesuai permintaan seiring dengan perubahan beban kerja akibat perubahan permintaan atau pasokan yang disebabkan oleh musim, cuaca, atau efek serupa lainnya.

Tenaga kerja digital terdiri dari bot yang melakukan aktivitas bisnis tertentu untuk mengotomatiskan proses bisnis. Setelah masuk ke sistem TI perusahaan, layaknya pengguna, menggunakan desktop virtual, bot ini dapat menjalankan beberapa proses bisnis paralel secara bersamaan. Umumnya, ada tiga jenis pekerjaan yang dilakukan bot tenaga kerja digital: otomatisasi tindakan, otomatisasi analisis, dan otomatisasi keputusan.

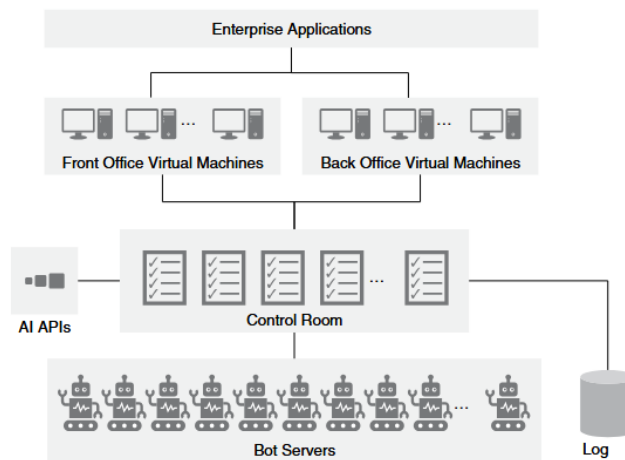
*Otomatisasi tindakan* terkadang disebut sebagai "jari di atas tombol." Bot yang melakukan otomatisasi tindakan dapat terhubung ke sistem perangkat lunak lain seperti sistem perencanaan sumber daya perusahaan (ERP), biasanya melalui layar front-end. Bot ini berfokus pada aktivitas bervolume tinggi, berulang, dan berbasis aturan yang dipicu secara digital, seperti mengambil tindakan saat berkas faktur diterima. Ini mungkin terjadi selama penutupan aktivitas hari kerja untuk organisasi perdagangan komoditas, misalnya, ketika bot dapat digunakan untuk mengeksekusi dan memantau serangkaian proses di beberapa sistem seperti penilaian, simulasi, pembuatan laporan, dan rekonsiliasi data.

*Otomatisasi analisis* adalah pengumpulan dan interpretasi informasi, seperti komentar bentuk bebas bahasa alami dalam sistem transaksi. Bot dapat mencerna data pasar dari berbagai sumber nonstandar dalam format terstruktur (misalnya data dari web atau umpan) atau data tidak terstruktur (misalnya email broker, dokumen Word, lampiran PDF). Bot kemudian mengekstrak dan mengubah data menjadi templat informasi standar dan memasukkannya ke dalam sistem hilir. Dalam beberapa kasus, bot dapat memanggil API AI untuk pengenalan gambar, pengenalan karakter optik (OCR), atau pemrosesan bahasa alami sesuai kebutuhan.

*Otomatisasi keputusan* adalah otomatisasi keputusan atas nama pengguna berdasarkan beberapa kebijakan yang telah ditentukan, atau kemampuan bot untuk memberikan rekomendasi kepada pengguna yang kemudian dapat membuat keputusan akhir. Contohnya adalah ketika faktur diterima dari rekanan perdagangan komoditas melalui email. Bot membandingkan jumlah yang ditagih dengan jumlah dalam sistem perdagangannya. Jika selisih nilai kurang dari jumlah dolar tertentu (ditetapkan oleh kebijakan), bot akan membuat

penyesuaian tunai dan menandai faktur untuk pembayaran. Jika tidak, bot akan menyerahkan faktur tersebut untuk ditinjau oleh analis manusia. Ini adalah contoh sederhana dari sebuah "keputusan", tetapi dalam situasi yang lebih kompleks, bot juga dapat memanggil API pembelajaran mesin untuk membantu pengambilan keputusan, seperti dalam proses persetujuan aplikasi pinjaman kecil.

Gambar 12.7 menunjukkan arsitektur tenaga kerja digital yang umum. Server bot menyimpan konfigurasi bot RPA dan tugas spesifik yang harus dijalankan bot dalam sistem. Ruang kontrol bot mengelola dan mengendalikan beberapa bot untuk menyelesaikan proses bisnis otomatis yang terkait dengan grup bot tersebut.



**Gambar 12.7** Arsitektur tenaga kerja digital berbasis RPA.

Ruang kontrol menetapkan proses bisnis yang akan dijalankan ke bot dalam grup berdasarkan ketersediaan bot, serta pada jadwal proses yang ditentukan dan dikonfigurasi di ruang kontrol. Ada dua jenis pemicu bot front office dan back office. Pengguna memicu bot front office saat mereka membutuhkannya. Informasi yang dihasilkan kemudian dikirim kembali ke ruang kontrol. Ruang kontrol itu sendiri memicu bot back-office, yang juga melaporkan status kembali ke ruang kontrol.

### 12.3 KESIMPULAN

Setelah kita meninjau struktur arsitektur platform AI, bab selanjutnya akan dikhususkan untuk mengilustrasikan cara kerja proses pembangunan model. Kita akan melakukannya dengan mengkaji aplikasi pembelajaran mesin spesifik yang dirancang untuk memahami churn pelanggan di industri telekomunikasi, tetapi metode yang sama berlaku untuk kasus penggunaan lainnya.

## **BAB 13**

### **PROSES PEMODELAN AI**

Dalam bab ini, kita akan membahas contoh sederhana aplikasi pembelajaran mesin untuk memberikan pemahaman yang lebih baik kepada pembaca tentang cara kerja proses pembangunan model, memperluas siklus hidup AI yang dibahas di Bab 8. Untuk contoh kasus penggunaan kita, kita akan menggunakan perusahaan telekomunikasi hipotetis yang sedang menyiapkan model AI untuk memecahkan masalah bisnis spesifik berupa churn pelanggan. Kita akan membahas penjelasan langkah demi langkah tentang cara kerja proses pemodelan, mengapa langkah-langkah spesifik penting, dan cara menghindari potensi jebakan. Bab ini tidak dimaksudkan untuk menunjukkan kepada pembaca cara menjadi ahli pemodelan AI, melainkan untuk membantu para manajer dan eksekutif yang akan mengawasi atau berinteraksi dengan tim AI mereka untuk lebih memahami apa yang dilakukan tim ini dan mengapa, serta keputusan bisnis apa yang terlibat dalam proses pemodelan.

#### **13.1 MENDEFINISIKAN KASUS PENGGUNAAN DAN TUGAS AI**

Ada beberapa pertanyaan kunci yang perlu dijawab untuk mendefinisikan kasus penggunaan AI dengan baik. Pertama, apa yang seharusnya menjadi keluaran dari sebuah model AI atau rangkaian model yaitu, prediksi apa yang harus dibuat atau tugas apa yang harus diselesaikan oleh model tersebut? Kedua, keputusan atau aktivasi apa dalam alur bisnis yang akan dimungkinkan oleh tugas AI agar dapat bernilai bagi bisnis? Terlalu sering, tim yang mengerjakan proyek AI hanya akan menjawab pertanyaan pertama dan bukan yang kedua, tetapi aktivasilah yang memastikan proyek AI tidak akan ditinggalkan. Maraknya bukti konsep (PoC) yang dibuang di ratusan perusahaan merupakan bukti bahwa pertanyaan tentang aktivasi tidak diajukan sejak dini dan cukup sering.

Bisnis menghabiskan banyak upaya dan uang untuk mendapatkan pelanggan baru dengan membangun dan mempertahankan tim penjualan dan pemasaran, membayar iklan, dan memberikan diskon atau insentif lain kepada individu untuk menjadi pelanggan. Oleh karena itu, bisnis di sebagian besar industri umumnya menganggap mempertahankan pelanggan lebih murah daripada mendapatkan pelanggan baru. Namun, dengan semakin banyaknya merek yang menawarkan layanan dan produk unggulan, sebagian besar perusahaan menghadapi tantangan churn pelanggan.

Churn pelanggan mengacu pada pelanggan yang berhenti menggunakan produk atau layanan perusahaan. Biasanya diukur sebagai persentase dalam jangka waktu tertentu, misalnya satu tahun. Memprediksi cara mempertahankan pelanggan ternyata menjadi keberhasilan awal yang penting dari pembelajaran mesin modern. Churn adalah masalah dalam perusahaan telekomunikasi imajiner kita; para eksekutif ingin mengetahui kelompok pelanggan mana yang cenderung beralih ke pesaing dan kemudian mencari cara untuk mempertahankan pelanggan ini dengan biaya yang lebih rendah daripada yang diperlukan untuk mendapatkan pelanggan baru. Oleh karena itu, tugas AI perusahaan telekomunikasi

untuk kasus penggunaan ini adalah melihat data pelanggan dan memprediksi, dengan akurasi yang cukup tinggi, ambang batas di mana pelanggan mereka kemungkinan akan churn katakanlah, dalam tiga bulan ke depan.

Masalah ini dapat diperlakukan sebagai masalah klasifikasi (pelanggan yang akan churn versus pelanggan yang tidak akan), atau sebagai masalah regresi (probabilitas churn untuk setiap pelanggan). Perlu diingat bahwa tugas AI ini (model) akan memprediksi apakah pelanggan akan churn atau tidak. Itu tidak akan merekomendasikan tindakan terbaik untuk bagaimana memperlakukan pelanggan yang berpotensi pergi. Aktivasi dalam kasus kami adalah untuk mengambil daftar pelanggan yang sangat mungkin untuk churn dan membuat kampanye pemasaran yang ditargetkan kepada pelanggan ini, dengan pesan atau penawaran untuk membuat kasus bagi mereka untuk tetap tinggal. Salah satu contohnya adalah menawarkan diskon kepada pelanggan ini pada tagihan berikutnya untuk memberi insentif kepada mereka untuk tetap dengan perusahaan telekomunikasi kami, sebagai penawaran melalui email langsung, atau sebagai panggilan keluar melalui pusat panggilan, atau sebagai pesan pada bagian penagihan dan pembayaran dari aplikasi telekomunikasi, di mana mereka dapat menerimanya dengan satu klik.

Saat mengimplementasikan kasus penggunaan apa pun untuk AI, sebuah tim sering kali memunculkan ide-ide tambahan untuk kasus penggunaan berdasarkan pengalaman kami, rata-rata empat ide baru per kasus penggunaan. Dalam skenario telekomunikasi kami, kasus penggunaan tambahan mungkin mencakup prediksi jenis insentif apa yang paling diminati setiap pelanggan (misalnya, tarif diskon di masa mendatang, akhir pekan gratis, atau pesan tentang perusahaan yang ramah lingkungan). Cara lainnya adalah memprediksi saluran komunikasi apa yang disukai setiap pelanggan, yaitu saluran mana yang akan meningkatkan kemungkinan respons positif (misalnya, email, panggilan agen, atau surat). Cara ketiga adalah memprediksi nilai seumur hidup pelanggan sehingga insentif hanya diberikan kepada pelanggan yang akan berharga dalam jangka panjang. Pada setiap proyek, kasus penggunaan baru seperti ini harus ditambahkan ke daftar kasus penggunaan yang telah diidentifikasi.

Pertanyaan berikutnya yang perlu dijawab untuk kasus penggunaan tertentu adalah tentang nilai. Nilai bisnis apa yang kita ciptakan, atau tujuan bisnis apa yang kita layani? Sangat penting untuk membuat identifikasi nilai tingkat tinggi selama definisi use case, baik untuk prioritas relatif (lihat Bab 8) dan untuk memastikan bahwa ada nilai aktual jika use case dikembangkan. Ini membantu mendorong adopsi aktivasi. Dalam contoh kami, jika dibutuhkan upaya periklanan dan pemasaran sebesar Rp 2.000.000 untuk mendapatkan pelanggan telekomunikasi baru dan perusahaan memiliki 10.000.000 pelanggan dengan churn tahunan sebesar 15%, mengurangi churn sebesar dua poin persentase akan mempertahankan 200.000 pelanggan setiap tahun, yang mengarah ke penghematan tahunan sebesar Rp 400 miliar. Mempertahankan pelanggan mungkin memerlukan insentif rata-rata Rp 500.000 per pelanggan, yang akan membuat biaya retensi menjadi Rp 100 miliar, menghasilkan penghematan bersih sebesar Rp 300 miliar per tahun. Contoh ini membuat banyak asumsi penyederhanaan dan tidak dimaksudkan untuk menjelaskan kasus bisnis secara rinci; untuk

penentuan kasus bisnis yang sebenarnya, detail yang cukup harus diberikan untuk membangun kasus yang lebih kuat sebelum investasi substansial dilakukan.

### 13.2 MEMILIH DATA YANG DIBUTUHKAN

Tidak selalu jelas di awal proses pemodelan AI data apa yang harus dikumpulkan untuk mengembangkan model. Sebaiknya cantumkan alasan potensial churn dan gunakan ini sebagai indikator data yang harus diperoleh. Misalnya, churn bisa jadi karena tagihan pelanggan yang terlalu tinggi, sehingga mendorongnya mencari penawaran yang lebih baik (misalnya, mendapatkan data penagihan) atau karena ia mengalami masalah dengan layanannya (misalnya, mendapatkan data penghentian layanan atau data tentang frekuensi panggilan ke pusat panggilan). Praktik terbaik lainnya adalah tidak berasumsi bahwa ini adalah alasan yang sebenarnya ini hanya menyediakan cara untuk melakukan brainstorming data apa yang harus dikumpulkan. Ini dapat diuji sebagai hipotesis menggunakan data tersebut. Setelah Anda memiliki data, biarkan data tersebut berbicara sendiri.

Dalam contoh kita, kita akan menggunakan kumpulan data sampel yang tersedia untuk umum tentang pelanggan di sebuah perusahaan telekomunikasi anonim. Kemungkinan besar data ini berasal dari beberapa sistem dan dihubungkan dengan identifikasi pelanggan. Sumber yang baik untuk kumpulan data publik seperti ini adalah Kaggle. Kumpulan data kami memiliki sekitar 3.500 baris data (observasi), dengan setiap baris mewakili satu pelanggan. Kolom-kolomnya adalah sebagai berikut:

1. CUSTOMER\_ID: nomor identifikasi pelanggan
2. STATE: negara bagian AS tempat pelanggan berada
3. AREA\_CODE: kode area telepon pelanggan
4. PHONE\_NUMBER: nomor telepon pelanggan
5. ACCOUNT\_LENGTH: durasi bulan pelanggan memiliki akun
6. INTL\_PLAN: apakah pelanggan memiliki paket internasional
7. VMAIL\_PLAN: apakah pelanggan memiliki paket pesan suara
8. VMAIL\_MSG: jumlah pesan suara yang diterima pelanggan dalam bulan tertentu
9. DAY\_MINS: jumlah menit siang hari yang digunakan pelanggan dalam bulan tertentu
10. DAY\_CALLS: jumlah panggilan yang dilakukan pelanggan pada siang hari dalam bulan tertentu
11. DAY\_CHARGE: total biaya dolar untuk panggilan siang hari dalam bulan tertentu
12. EVE\_MINS: jumlah menit malam hari yang digunakan pelanggan dalam bulan tertentu
13. EVE\_CALLS: jumlah panggilan yang dilakukan pelanggan pada malam hari dalam bulan tertentu
14. EVE\_CHARGE: total biaya dolar untuk panggilan malam hari dalam bulan tertentu
15. NIGHT\_MINS: jumlah menit malam hari yang digunakan pelanggan dalam bulan tertentu
16. NIGHT\_CALLS: jumlah panggilan yang dilakukan pelanggan pada malam hari dalam bulan tertentu
17. NIGHT\_CHARGE: total biaya dolar untuk panggilan malam hari dalam bulan tertentu

18. INTL\_MINS: jumlah menit internasional yang digunakan pelanggan dalam bulan tertentu
19. INTL\_CALLS: jumlah panggilan internasional yang dilakukan pelanggan dalam bulan tertentu
20. INTL\_CHARGE: total biaya dolar untuk panggilan internasional dalam bulan tertentu
21. CUST\_SERV\_CALLS: jumlah panggilan layanan pelanggan dalam bulan tertentu
22. CHURN: kolom ya atau tidak yang menunjukkan apakah pelanggan berhenti berlangganan atau tidak dalam jangka waktu tiga bulan dari periode yang dicakup oleh data di kolom lainnya

### 13.3 MENYIAPKAN LINGKUNGAN NOTEBOOK DAN MENGIMPOR DATA

Kami akan mengembangkan model kami menggunakan lingkungan notebook Jupyter sumber terbuka dan bahasa pemrograman Python, yang saat ini merupakan bahasa pemrograman paling populer untuk AI dan ilmu data. Banyak pustaka yang ada dapat digunakan untuk proyek ini pustaka untuk manipulasi data, aljabar matriks, grafik, dan sebagainya. Kami akan menggunakan pustaka pembelajaran mesin sumber terbuka scikit-learn (sklearn). Gambar 13.1 menunjukkan pengimporan berbagai pustaka dari sklearn dan juga pengimporan pustaka sumber terbuka lainnya seperti SMOTE.

Dengan menggunakan kode pada Gambar 13.2, kami mengimpor data yang terdapat dalam berkas nilai yang dipisahkan koma (CSV) di folder pada drive C. Kami melakukan ini dengan menggunakan struktur data Python DataFrame, dan kami menyebutnya `imp_data` (untuk data yang diimpor).

```

1 # import math and data libraries
2 import pandas as pd
3 import numpy as np
4 from scipy.stats import uniform, randint
5
6 # import visualization libraries
7 import matplotlib.pyplot as plt
8 import seaborn as sns
9 from mpl_toolkits.mplot3d import Axes3D
10
11 import missingno as msno
12
13 from imblearn.over_sampling import SMOTE
14
15 # import sklearn machine learning libraries
16 from sklearn.preprocessing import LabelBinarizer, LabelEncoder, \
17     StandardScaler
18
19 from sklearn.compose import ColumnTransformer
20 from sklearn.pipeline import Pipeline
21 from sklearn.impute import SimpleImputer
22
23 from sklearn.model_selection import train_test_split, cross_val_score, \
24     GridSearchCV, KFold, StratifiedKFold, RandomizedSearchCV
25
26 # import the necessary model types
27 from sklearn.linear_model import LogisticRegression, LinearRegression
28 from sklearn.naive_bayes import GaussianNB
29 from sklearn.svm import LinearSVC
30 from sklearn.ensemble import RandomForestClassifier
31 import xgboost as xgb
32
33 # import model performance tools
34 from sklearn import metrics
35 from sklearn.metrics import precision_recall_curve, roc_curve, auc, \
36     accuracy_score, make_scorer, recall_score, \
37     precision_score, confusion_matrix

```

**Gambar 13.1** Mengimpor pustaka relevan yang akan digunakan.

```
1 # set the folder and file names from where you want to get data
2 folderName = 'gdrive/My Drive/Colab Notebooks/Data/'
3 fileName = 'customer_churn.csv'
4
5 # create dataframe and read file into dataframe
6 imp_data = pd.read_csv(folderName + fileName)
7 imp_data.shape
```

(3333, 22)

Gambar 13.2 Mengimpor data untuk churn pelanggan.

```
1 imp_data.head()
```

	CUSTOMER_ID	STATE	AREA_CODE	PHONE_NUMBER	ACCOUNT_LENGTH	INTL_PLAN	VMAII
0	10001	KS	415	382-4657	128	no	
1	10002	OH	415	371-7191	107	no	
2	10003	NJ	415	358-1921	137	no	
3	10004	OH	408	375-9999	84	yes	
4	10005	OK	415	330-6626	75	yes	

5 rows x 22 columns

Gambar 13.3 Melihat beberapa baris teratas data.

Dua angka di bagian bawah Gambar 13.2 menunjukkan bahwa terdapat 3.333 baris dan 22 kolom data. 21 kolom pertama adalah fitur, dan kolom terakhir dalam kasus kita, CHURN – adalah target yang ingin kita prediksi. Pada Gambar 13.3, kita dapat melihat beberapa baris pertama data.

#### 13.4 MEMBERSIHKAN DAN MEMPERSIAPKAN DATA

Langkah pertama kita adalah membersihkan data (dibahas di Bab 8). Ini seringkali merupakan tugas yang memakan waktu. Mencari tahu cara menangani nilai yang hilang merupakan bagian penting dari proses ini, karena model yang berbeda bisa lebih atau kurang sensitif terhadap kondisi ini, sehingga model tersebut kurang atau bahkan tidak dapat memprediksi dengan andal kapan data yang tersedia tidak lengkap. Inilah sebabnya kita pertama-tama mencari nilai yang hilang. Kita dapat memvisualisasikannya dengan menggambar peta panas, yang ditunjukkan pada Gambar 13.4. Dalam contoh dataset kita, seperti yang Anda lihat, terdapat beberapa data yang hilang di kolom PHONE\_NUMBER. Ini bukan sesuatu yang akan kita perbaiki, karena kita akan menghapus kolom ini. Namun, jika terdapat nilai yang hilang dari fitur yang lebih penting, ada beberapa cara untuk mengatasinya.

Salah satu opsi adalah menghapus semua baris dengan nilai kosong. Misalnya, jika data yang hilang untuk beberapa baris ada di kolom target, maka mungkin lebih baik untuk menghapus seluruh baris data tersebut. Masalahnya adalah menghapus baris-baris tersebut berisiko kehilangan informasi penting. Cara yang lebih baik untuk menangani nilai yang hilang mungkin dengan menggantinya dengan nilai default yang berasal dari dataset. Dalam beberapa kasus, penggunaan nilai dari baris sebelumnya atau setelahnya dapat menyelesaikan

masalah ini. Dalam kasus lain, hal ini dilakukan dengan imputasi data interpolasi dari seluruh dataset dan menggunakan, misalnya, nilai rata-rata. Interpolasi ini tidak selalu harus berupa nilai rata-rata angka dalam dataset tertentu; bisa juga berupa keluaran dari model pembelajaran mesin lain. Dalam contoh kita, kita akan menggunakan imputer dari framework sklearn.



**Gambar 13.4** Peta panas nilai yang hilang. Jika ada, akan ditampilkan sebagai bilah putih untuk baris dan kolom tersebut.

Langkah persiapan data berikutnya adalah mengonversi nilai kategorikal menjadi nilai numerik. Konversi ini diperlukan karena banyak model pembelajaran mesin tidak bekerja dengan nilai data teks. Dalam contoh kami, kami menggunakan teknik yang disebut pengkodean label, yang mengubah nilai ya/tidak dan benar/salah menjadi 0 dan 1 dengan menggunakan fungsi sklearn LabelEncoder (lihat Gambar 13.5). Namun, konversi ini hanya dapat dilakukan ketika berhadapan dengan dua kategori. Pengkodean, katakanlah, satu set nama negara bagian menjadi data numerik dapat menimbulkan kesulitan, karena nama negara bagian adalah data kategorikal tanpa hubungan apa pun satu sama lain. Masalah muncul ketika model AI mengasumsikan hubungan atau urutan di antara angka-angka yang ditetapkan yang berada di kolom yang sama dan memutuskan, misalnya, bahwa 0 kurang dari 1, di mana nol mungkin merujuk ke Alaska, dan satu mungkin merujuk ke Florida.

Untuk mengatasi masalah ini, kami menggunakan apa yang disebut pengkodean one-hot. Pengkodean one-hot mengambil kolom dengan data kategorikal yang dikodekan-label dan membaginya menjadi beberapa kolom di mana angka yang ada diganti dengan 0 dan 1. Dalam contoh kita, kita membuat tiga kolom baru: katakanlah, New York, California, dan Michigan. Untuk baris data pelanggan yang negara bagiannya adalah New York, kolom yang sesuai dengan New York diatur ke 1; kolom yang sesuai dengan negara bagian lainnya diatur ke 0.

```

1 # drop features that are low impact
2 proc_data = imp_data.drop(columns=['AREA_CODE', 'PHONE_NUMBER'])
3 proc_data.shape
4
5 # transforming categorical data to numerical values
6 target_features = ['INTL_PLAN', 'VMAIL_PLAN', 'CHURN']
7 for i, target_feature in enumerate(target_features):
8     print(target_feature + " : ", proc_data[target_feature].unique())
9
10 # use encoder and transform
11 encoder = LabelEncoder()
12 for i, target_feature in enumerate(target_features):
13     encoded_values = encoder.fit_transform(proc_data[target_feature].values)
14     proc_data[target_feature] = pd.Series(encoded_values, index=imp_data.index)
15     # proc_data[target_feature] = proc_data[target_feature].astype('float64')
16     print(target_feature + " : ", proc_data[target_feature].unique())

```

```

INTL_PLAN : ['no' 'yes']
VMAIL_PLAN : ['yes' 'no']
CHURN : ['False.' 'True.']
INTL_PLAN : [0 1]
VMAIL_PLAN : [1 0]
CHURN : [0 1]

```

**Gambar 13.5** Transformasi data teks kategorikal menjadi nilai numerik.

```

1 # one hot encode categorical values that have more than 2 categories
2
3 proc_data = pd.get_dummies(proc_data, columns=['STATE'])
4 proc_data.shape

```

```

(3333, 71)

```

**Gambar 13.6** Pengkodean one-hot negara bagian AS.

Jika baris data pelanggan adalah untuk warga California, kolom yang sesuai dengan California akan menjadi 1; kolom untuk New York dan Michigan keduanya akan diatur ke 0. Kita sekarang menggunakan fungsi `get_dummies` Python (lihat Gambar 13.6) untuk mengubah set negara bagian menjadi 51 kolom (untuk 50 negara bagian ditambah Distrik Columbia), dengan 1 di kolom untuk setiap baris yang mengacu pada negara bagian yang diwakili oleh kolom itu, dan 0 untuk semua kolom lainnya. Dengan mengubah nilai kategorikal menjadi nilai numerik dengan cara ini, kita dapat mengeksplorasi data dengan lebih bermanfaat. Anda dapat melihat di bagian bawah Gambar 13.6 bahwa sekarang terdapat 71 kolom. Ini berasal dari 22 kolom awal, dikurangi 2 kolom yang kita hilangkan (untuk kode area dan nomor telepon), dan 51 kolom negara bagian baru yang kita tambahkan.

### 13.5 MEMAHAMI DATA MENGGUNAKAN ANALISIS DATA EKSPLORATORI

Jika Anda ingin membuat prediksi AI yang baik, penting untuk memiliki pemahaman yang baik tentang set data pelatihan Anda (yang berkualitas tinggi). Jika model pembelajaran mesin gagal memprediksi nilai masa depan yang akurat sebuah kejadian yang umum terjadi dalam proyek AI hal ini seringkali disebabkan oleh pemahaman data yang kurang tepat; kualitas hasil dari model AI berkorelasi langsung dengan kualitas set data pelatihan dan seberapa baik pemahamannya. Memperbaiki situasi ini dapat dilakukan secara iteratif, tetapi akan lebih mudah dan efektif untuk benar-benar memahami data, merencanakan transformasi yang diperlukan, dan kemudian mengulangi proses pemodelan. Tingkat pemahaman ini diperoleh melalui apa yang dikenal sebagai analisis data eksploratori. Pemahaman yang mendalam tentang data inilah yang seringkali membedakan ilmuwan AI yang baik dari yang biasa-biasa saja. Hal ini tidak terkait dengan teknis maupun pemrograman: pemahaman ini

melibatkan kemampuan untuk membuat keputusan yang tepat tentang data dan memilih model yang paling relevan untuk situasi tertentu.

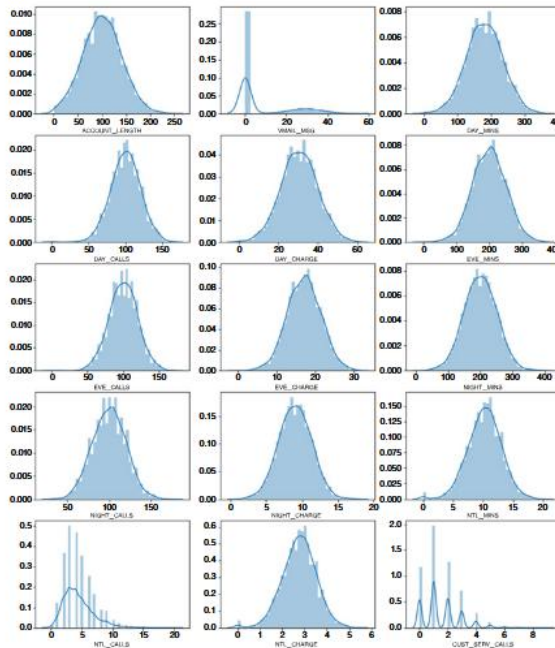
Salah satu teknik untuk melakukan analisis eksploratori adalah dengan menghitung statistik dasar tentang data, misalnya, menentukan rerata dan deviasi standar setiap fitur. Gambar 13.7 menunjukkan kode untuk memplot distribusi data beberapa kolom. Dalam hasilnya, yang ditunjukkan pada Gambar 13.8, kita dapat dengan jelas melihat bahwa sebagian besar fitur numerik tampak terdistribusi normal, meskipun VMAIL\_MSG, INT\_CALLS, dan CUST\_SERV\_CALLS tidak.

Selanjutnya, kita melihat dependensi antar fitur. Dalam matriks korelasi pada Gambar 13.9, kita dapat melihat seberapa kuat setiap fitur bervariasi dengan fitur lainnya dan bagaimana atribut individual berkorelasi dengan atribut target (baris terakhir pada Gambar 13.9). Misalnya, kotak gelap di sel untuk VMAIL\_MSG (pada sumbu vertikal) dan VMAIL\_PLAN (pada sumbu horizontal) memberi tahu kita bahwa paket pesan suara dan jumlah pesan suara sangat berkorelasi. Hal itu dapat dimengerti karena Anda tidak dapat menerima pesan suara jika tidak memiliki paket pesan suara. Matriks ini akan diperlukan saat kita membahas rekayasa fitur.

Kita juga dapat mencari outlier dalam analisis eksploratori kita, dalam hal ini, menggunakan diagram kotak. Setiap diagram kotak menunjukkan median kolom data, yang ditunjukkan sebagai garis di tengah setiap kotak abu-abu pada Gambar 13.10. Kotak abu-abu menunjukkan rentang data dari kuartil pertama (yang merupakan median dari separuh bagian bawah kumpulan data) hingga kuartil ketiga (median dari separuh bagian atas kumpulan data). Titik-titik di bagian atas dan bawah adalah nilai akhir yang berpotensi menjadi outlier. Dalam contoh kita, kita akan mengabaikan outlier dalam eksperimen pemodelan pertama dan memasukkan semua data.

```
1 # look at distribution of numeric data sets
2 col_names = ['ACCOUNT_LENGTH', 'VMAIL_MSG', 'DAY_MINS', 'DAY_CALLS', \
3             'DAY_CHARGE', 'EVE_MINS', 'EVE_CALLS', 'EVE_CHARGE', \
4             'NIGHT_MINS', 'NIGHT_CALLS', 'NIGHT_CHARGE', 'INTL_MINS', \
5             'INTL_CALLS', 'INTL_CHARGE', 'CUST_SERV_CALLS']
6
7 fig, axs = plt.subplots(5,3, figsize=(14,17))
8 for i, col_val in enumerate(col_names):
9     sns.distplot(proc_data[col_val], hist=True, ax=axs.flat[i])
10    axs.flat[i].set_xlabel(col_val, fontsize=8)
11    #axs.flat[i].set_ylabel('Count', fontsize=8)
```

**Gambar 13.7** Plotting frekuensi kumpulan data.

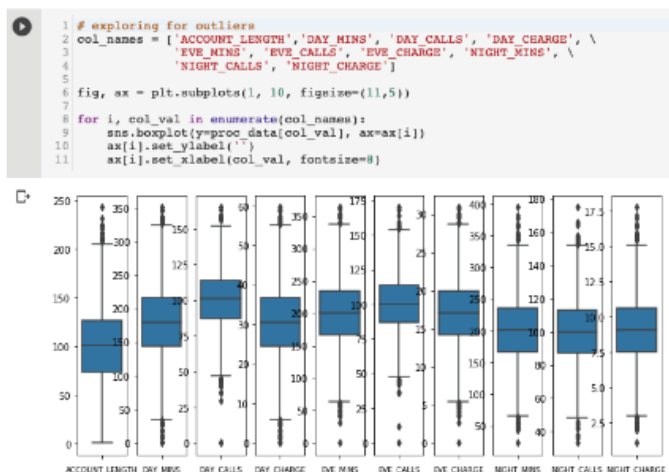


**Gambar 13.8** Distribusi frekuensi data beberapa kolom.

Analisis eksploratori juga perlu mencari ketidakseimbangan data. Pada Gambar 13.11, kita dapat melihat bahwa sekitar 14,5% pelanggan telekomunikasi berhenti berlangganan (churn), sementara 85,5% tidak. Artinya, jika kita membuat "model" yang hanya menyatakan bahwa, apa pun inputnya, prediksinya adalah tidak ada churn pelanggan, model ini akan akurat 85,5% dari waktu, dan salah 14,5%.



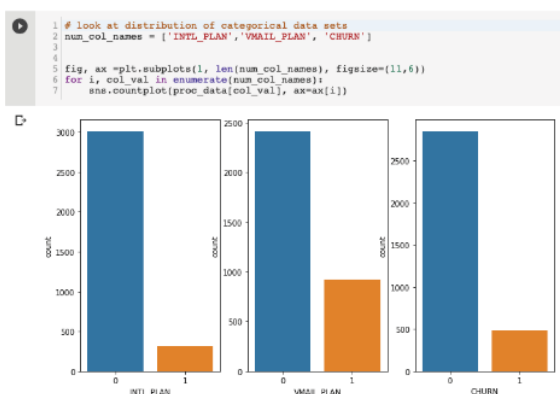
**Gambar 13.9** Peta panas korelasi beberapa kolom utama satu sama lain.



**Gambar 13.10** Mencari outlier.

Namun, model ini tidak akan bermanfaat meskipun memiliki akurasi tinggi, yaitu 85,5%. Melihat ketidakseimbangan data seperti ini akan mengingatkan Anda bahwa model Anda perlu memprediksi siapa yang akan churn, alih-alih hanya memprediksi 15% pelanggan yang akan churn, yang akan membantu menentukan apakah model yang telah dilatih dapat digunakan atau tidak.

Hal lain yang perlu diperhatikan adalah memastikan bahwa nilai dari berbagai fitur diskalakan dengan benar, karena pasti akan ada beberapa fitur dalam satu set data. Seringkali, model AI akan menentukan bahwa nilai yang lebih tinggi lebih penting daripada nilai yang lebih rendah. Hal ini akan menyebabkan fitur dengan magnitudo yang lebih besar mendominasi model. Jika fitur VMAIL\_MSG berkisar antara 0 hingga 51, sedangkan fitur DAY\_MINS berkisar antara 0 hingga 351, model mungkin berasumsi bahwa DAY\_MINS lebih penting daripada VMAIL\_MSG, yang mungkin bukan kasusnya.



**Gambar 13.11** Ketidakseimbangan data label atau target.

Penskalaan memungkinkan kita untuk memperlakukan semua fitur secara setara. Sebagaimana dibahas dalam Bab 8, ada dua cara untuk menskalakan fitur. Normalisasi menskalakan angka agar berada di antara nol dan satu, dan standardisasi menskalakan angka sehingga nilai fitur mengikuti distribusi normal dengan nilai rata-rata nol dan deviasi standar

satu. Kami menggunakan standardisasi (lihat Gambar 13.12) dalam contoh kami karena mempertahankan outlier, yang mungkin berisi informasi penting yang tidak ingin kami hilangkan.

Seperti yang Anda lihat pada Gambar 13.13, kami menstandarisasi semua atribut numerik kecuali CUST\_SERV\_CALLS. Kami melakukan ini karena pelanggan yang puas cenderung tidak menghubungi layanan pelanggan secara teratur, sehingga jumlah panggilan ke layanan pelanggan mungkin berkorelasi tinggi dengan churn. Karena kami tidak menskalakan fitur ini, model kami akan menganggapnya lebih signifikan, tetapi tidak akan terlalu mendominasi atribut fitur lainnya karena memiliki nilai rata-rata sekitar 1,56. Jika suatu fitur tertentu memiliki nilai yang lebih tinggi, misalnya 7, fitur tersebut akan terlihat signifikan.

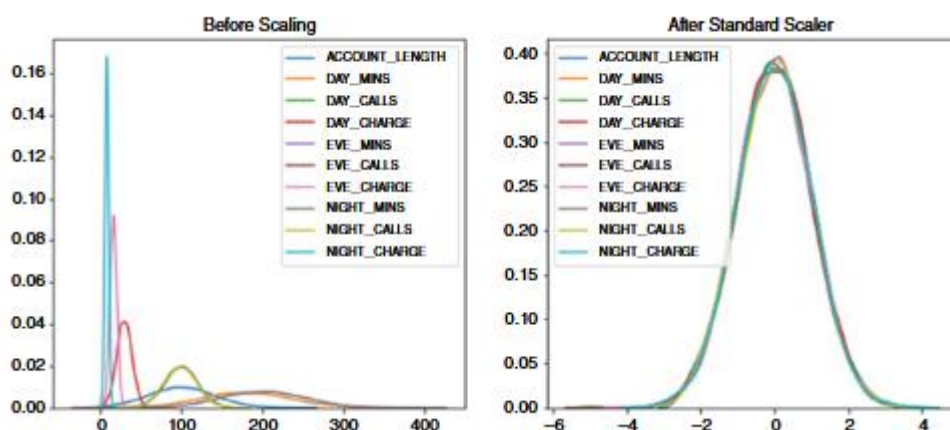
```
1 # scaling the features
2 scaler = StandardScaler()
3 scale_cols = ['ACCOUNT_LENGTH', 'DAY_CALLS', 'DAY_CHARGE', \
4             'EVE_MINS', 'EVE_CALLS', 'EVE_CHARGE', 'NIGHT_MINS', \
5             'NIGHT_CALLS', 'NIGHT_CHARGE',
6             'VMAIL_MSG', 'INTL_MINS', 'INTL_CALLS', 'INTL_CHARGE']
7 scaled_data = scaler.fit_transform(proc_data[scale_cols])
8 scaled_data = pd.DataFrame(scaled_data, columns=scale_cols)
9 scaled_full_data = proc_data.drop(scale_cols, axis=1)
10 scaled_full_data = pd.concat([scaled_full_data, scaled_data], \
11                             axis=1, sort=False)
12 scaled_full_data.shape
13
14 fig, (ax1, ax2) = plt.subplots(ncols=2, figsize=(12, 5))
15
16 ax1.set_title('Before Scaling')
17 sns.kdeplot(proc_data['ACCOUNT_LENGTH'], ax=ax1)
18 sns.kdeplot(proc_data['DAY_MINS'], ax=ax1)
19 sns.kdeplot(proc_data['DAY_CALLS'], ax=ax1)
20 sns.kdeplot(proc_data['DAY_CHARGE'], ax=ax1)
21 sns.kdeplot(proc_data['EVE_MINS'], ax=ax1)
22 sns.kdeplot(proc_data['EVE_CALLS'], ax=ax1)
23 sns.kdeplot(proc_data['EVE_CHARGE'], ax=ax1)
24 sns.kdeplot(proc_data['NIGHT_MINS'], ax=ax1)
25 sns.kdeplot(proc_data['NIGHT_CALLS'], ax=ax1)
26 sns.kdeplot(proc_data['NIGHT_CHARGE'], ax=ax1)
27
28 ax2.set_title('After Standard Scaler')
29 sns.kdeplot(scaled_data['ACCOUNT_LENGTH'], ax=ax2)
30 sns.kdeplot(scaled_data['DAY_MINS'], ax=ax2)
31 sns.kdeplot(scaled_data['DAY_CALLS'], ax=ax2)
32 sns.kdeplot(scaled_data['DAY_CHARGE'], ax=ax2)
33 sns.kdeplot(scaled_data['EVE_MINS'], ax=ax2)
34 sns.kdeplot(scaled_data['EVE_CALLS'], ax=ax2)
35 sns.kdeplot(scaled_data['EVE_CHARGE'], ax=ax2)
36 sns.kdeplot(scaled_data['NIGHT_MINS'], ax=ax2)
37 sns.kdeplot(scaled_data['NIGHT_CALLS'], ax=ax2)
38 sns.kdeplot(scaled_data['NIGHT_CHARGE'], ax=ax2)
39
40 plt.show()
```

**Gambar 13.12** Penskalaan Kolom Data yang Relevan.

### 13.6 REKAYASA FITUR

Beberapa fitur mungkin tidak sepenting yang kita bayangkan, sementara fitur lainnya signifikan. Menambahkan fitur yang tidak perlu tidak hanya memperlambat pelatihan tetapi juga dapat menyebabkan overfitting model pembelajaran mesin (lihat Bab 8). Selain itu, mempertahankan dua fitur berbeda yang sangat berkorelasi satu sama lain dapat memberikan pengaruh yang terlalu besar pada hasil.

Dari matriks korelasi pada Gambar 13.9, kita dapat melihat kolom berkorelasi tinggi mana yang harus kita pertahankan sebagai fitur. Berdasarkan informasi ini, kita dapat mengganti DAY\_CHARGE dengan DAY\_MINS, EVE\_CHARGE dengan EVE\_MINS, NIGHT\_CHARGE dengan NIGHT\_MINS, dan INTL\_CHARGE dengan INTL\_MINS (lihat Gambar 13.14). Jika perlu, kita dapat bereksperimen dengan memasukkan dan mengecualikannya serta melihat dampaknya terhadap kinerja model. Secara umum, kami akan menggunakan teknik seperti analisis komponen utama untuk melakukan hal ini, tetapi kami akan melewatkannya di sini karena lebih rumit.



**Gambar 13.13** Memvisualisasikan distribusi data sebelum penskalaan (kiri) dan penskalaan setelahnya (kanan).

```

1 # remove columns with higher correlations, etc.
2 scaled_full_data['TOTAL_CHARGE'] = scaled_full_data['DAY_CHARGE'] + \
3   scaled_full_data['EVE_CHARGE'] + scaled_full_data['NIGHT_CHARGE'] + \
4   scaled_full_data['INTL_CHARGE']
5 scaled_full_data = scaled_full_data.drop(['DAY_CHARGE', 'EVE_CHARGE', \
6   'NIGHT_CHARGE', 'INTL_CHARGE'], axis = 1)

```

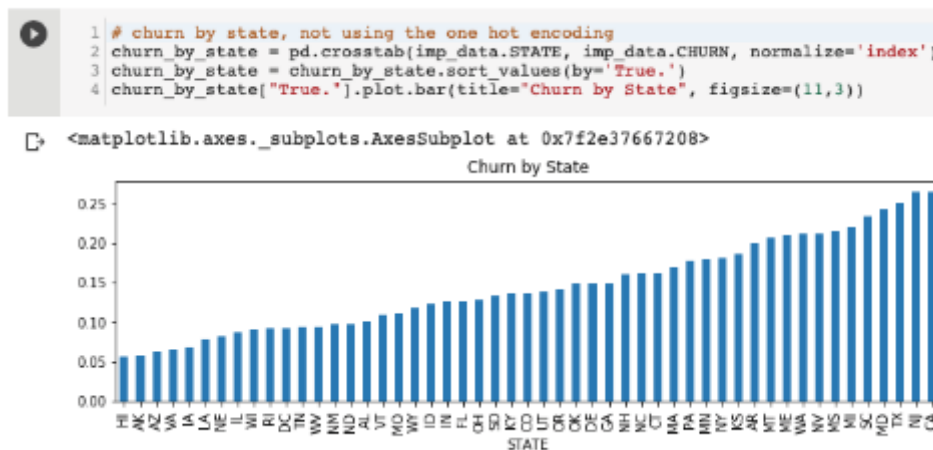
**Gambar 13.14** Menghilangkan kolom muatan individual dan menambahkan kolom muatan total.

Kita juga dapat melihat variabilitas data untuk memutuskan variabel mana yang akan disertakan dalam model. Umumnya merupakan ide yang baik untuk membuang variabel jika Anda yakin variabel tersebut tidak berpengaruh pada atribut target dalam hal ini, churn karena ini menghilangkan noise yang harus ditangani oleh algoritma pelatihan model.

Misalnya, bagaimana kita dapat memutuskan apakah menggunakan informasi STATE dalam model AI kita akan membantu? Kita dapat menentukan apakah STATE yang dikaitkan dengan seseorang berkorelasi dengan peningkatan kemungkinan churn. Jika persentase penduduk New York yang churn tinggi dalam dataset, kita dapat berasumsi bahwa penduduk New York lebih mungkin untuk churn daripada yang lain. Ini akan menjadi atribut yang berguna untuk membuat prediksi, jadi STATE harus disertakan dalam kasus ini. Jika tidak ada korelasi antara STATE dan churning, kita mungkin ingin menghilangkan STATE dari dataset kita, karena kemungkinan tidak akan membantu dalam memprediksi churn. Pada Gambar 13.15, kita dapat melihat bahwa terdapat varians yang cukup dari satu negara bagian ke negara bagian lain dalam churn dari 5 hingga 25% agar STATE bermanfaat, sehingga kami mempertahankannya sebagai fitur dalam kumpulan data kami.

Contoh lain fitur yang mungkin ingin Anda hilangkan dari kumpulan data Anda adalah variabilitas yang rendah. Katakanlah Anda melihat statistik untuk STATE dan AREA\_CODE dan menemukan bahwa meskipun terdapat 51 kolom STATE yang berbeda, hanya terdapat tiga kolom AREA\_CODE. Salah satu alasannya mungkin karena untuk melindungi privasi klien, nomor telepon mereka, dan oleh karena itu kode area mereka, dihilangkan dari kumpulan

data. Ini menunjukkan bahwa Anda mungkin sebaiknya menghilangkan fitur AREA\_CODE dan PHONE\_NUMBER (lihat Gambar 13.5), karena data fiktif tidak akan berkontribusi pada akurasi hasil Anda.



**Gambar 13.15** Menganalisis tingkat churn berdasarkan negara bagian.

Terkadang, membuat fitur baru dari satu atau beberapa fitur yang sudah ada dalam kumpulan data Anda bermanfaat. Misalnya, kita dapat memutuskan untuk membuat fitur baru dengan menggabungkan keempat fitur CHARGE, karena kemungkinan besar total biaya yang menyebabkan seseorang churn, bukan biaya berdasarkan waktu. Kita juga dapat membuat fitur baru yang merangkum beberapa fitur asli dengan mengalikan atau membaginya. Misalnya, kita membagi DAY\_MINS dengan DAY\_CALLS dan menghasilkan fitur baru, DAY\_AVG\_CALL, yaitu rata-rata lama panggilan dalam sehari.

### 13.7 MEMBUAT DAN MEMILIH MODEL YANG OPTIMAL

Model dilatih pada dataset, dan setelah pelatihan, Anda memerlukan dataset yang berbeda untuk menguji model Anda. Ingatlah bahwa tujuan pembelajaran mesin adalah untuk membuat model dengan akurasi prediktif tertinggi pada data yang belum pernah dilihatnya. Jika dataset pelatihan dan pengujian Anda sama, model Anda dapat mengingat data pelatihan, yang merupakan kasus ekstrem dari apa yang dikenal sebagai overfitting (lihat Bab 8), dan dengan demikian tidak dapat digeneralisasi ke data baru. Mendeteksi overfitting adalah alasan untuk memisahkan dataset pelatihan dari dataset pengujian. Biasanya, data harus dibagi secara acak. Sebuah alat yang dapat membantu, yang disebut Pustaka Pemilihan Model, tersedia di pustaka sklearn. Di dalam Pustaka Pemilihan Model terdapat kelas bernama train\_test\_split. Kelas ini memungkinkan kita untuk membagi dataset menjadi dataset pelatihan dan pengujian dalam proporsi yang kita pilih.

Cara kerjanya relatif sederhana. Parameter test\_size, yang diberikan sebagai pecahan, menentukan seberapa banyak dataset yang akan digunakan dalam pengujian. Jika, misalnya, test\_size adalah 0,5, 50% dari dataset akan dipisahkan sebagai data pengujian. Jika Anda tidak menentukan parameter ini, Anda memiliki opsi untuk menggunakan train\_size, yang

beroperasi dengan cara yang sama. Jika Anda memilih 0,5 sebagai nilai, 50% dari dataset akan digunakan sebagai set pelatihan. Jika Anda ingin menentukan elemen mana yang dipilih untuk pelatihan dan pengujian secara acak, Anda dapat menggunakan parameter `random_state` dengan memilih bilangan bulat untuk dijadikan seed bagi generator bilangan acak selama pemisahan. Dalam contoh kami, kami menggunakan 75% data untuk pelatihan, dan 25% untuk pengujian (lihat Gambar 13.16).

Kami telah memilih model regresi logistik sebagai algoritma yang ingin kami latih untuk menjadi model AI awal kami. Ingatlah bahwa ketika kami menggunakan istilah algoritma, kami mengacu pada algoritma generik, seperti algoritma regresi linier. Ketika kami menggunakan istilah model, yang kami maksud adalah model yang dilatih secara eksplisit pada data spesifik untuk hasil spesifik, seperti model churn kami. Model ini akan dilatih pada data churn pelanggan historis untuk memprediksi apakah pelanggan akan meninggalkan perusahaan berdasarkan input data (fitur) yang diberikan. Model ini didasarkan pada algoritma regresi logistik, yang merupakan pengklasifikasi biner atau multikelas umum. Apakah Anda memilih algoritma regresi atau algoritma klasifikasi tergantung pada masalah yang Anda pecahkan. Jika kita ingin memprediksi pelanggan mana yang mungkin pergi, kita menggunakan algoritma klasifikasi. Jika kita, sebaliknya, ingin memprediksi nilai umur pelanggan dari setiap pelanggan, kita menggunakan algoritma regresi untuk mendapatkan keluaran variabel kontinu.

```
1 # split the features from the target variable
2
3 sourcevars = scaled_full_data.drop(['CHURN', 'CUSTOMER_ID'], axis=1)
4 targetvar = scaled_full_data['CHURN']
5
6 # split the training and validation datasets
7
8 xTrain, xTest, yTrain, yTest = train_test_split(sourcevars, targetvar, \
9         test_size = 0.25, random_state = 0)
10
11 sourcevars.shape, targetvar.shape
```

↳ ((3333, 65), (3333,))

**Gambar 13.16** Membagi data untuk pelatihan dan pengujian dalam rasio 75:25.

Anda dapat melihat dari Gambar 13.17 bahwa model kami memprediksi dengan akurasi 84,8%. Hal ini mungkin tampak cukup baik, tetapi perlu diingat bahwa sebelumnya kami menunjukkan bahwa sekitar 14,5% dari dataset sumber pelanggan telah berhenti berlangganan, sedangkan 85,5% tidak (dalam dataset lengkap yang terdiri dari 3.333 pelanggan). Jika model kami memprediksi bahwa tidak ada pelanggan yang akan berhenti berlangganan, akurasi akan mencapai 85,5%. Oleh karena itu, model kami sebenarnya berkinerja lebih buruk daripada jika memprediksi tidak ada yang akan berhenti berlangganan. Perhatikan bahwa jika kami hanya menggunakan data uji (25% dari data) untuk akurasi prediksi, kami akan mendapatkan 86,2% (lihat Gambar 13.18). Angka ini sedikit berbeda dari 85,5% karena kami menggunakan sebagian data untuk pengujian.

Mari kita telaah metrik kinerja model lebih cermat. Pertama, kita lihat matriks konfusi (lihat Gambar 13.19), yang digunakan untuk menggambarkan kinerja model klasifikasi. Dalam kasus kami, ada dua kemungkinan kelas yang diprediksi “diprediksi tidak ada churn” dan

“diprediksi churn” dan data memiliki dua kategori “benar tidak ada churn” dan “benar churn.” Matriks kebingungan menunjukkan bahwa 687 pelanggan true\_no\_churn diprediksi secara akurat sebagai pred\_no\_churn, tetapi 32 pelanggan actual\_no\_churn salah diprediksi sebagai pred\_churn.

```

1 # try classification models
2 model = LogisticRegression(solver = 'lbfgs')
3
4 # train the algorithm on training data and predict using the testing data
5 model.fit(xTrain, yTrain)
6 predictions = model.predict(xTest)
7 print("Accuracy : ",accuracy_score(yTest, predictions, normalize = True))

```

Accuracy : 0.8477218225419664

**Gambar 13.17** Menyiapkan model regresi logistik untuk klasifikasi biner.

```

1 # this is the accuracy if you assume NO customers will churn
2 1 - yTest.mean()

```

0.8621103117505995

**Gambar 13.18** Persentase pelanggan yang tidak berhenti berlangganan dalam set data validasi.

Demikian pula, 20 pelanggan actual\_churn diprediksi dengan benar sebagai pred\_churn, dan 95 pelanggan actual\_churn salah diprediksi sebagai pred\_no\_churn. Akurasi model adalah total prediksi yang benar (687 + 20) sebagai persentase dari semua pelanggan (687 + 20 + 32 + 95), yang, seperti yang kita lihat sebelumnya, adalah 84,8% yang tidak terlalu berguna. Untuk mengatasi masalah ini, kita perlu melihat metrik di luar akurasi.

Beberapa ukuran penting lainnya, terutama dalam kasus data yang tidak seimbang seperti milik kami, disebut recall, precision, dan skor F1. *Recall*, juga disebut sensitivitas atau rasio positif sejati, adalah jumlah positif sejati (20) dibagi dengan semua yang sebenarnya positif (20 + 95 = 115). Dalam contoh kami, recall adalah 17% (20/115).

```

1 # print(metrics.confusion_matrix(yTest, predictions))
2 print(pd.DataFrame(confusion_matrix(yTest, predictions),
3                    columns=['pred_no_churn', 'pred_churn'],
4                    index=['actual_no_churn', 'actual_churn']))

```

	pred_no_churn	pred_churn
actual_no_churn	687	32
actual_churn	95	20

```

[173] 1 # look at performance metrics
2 print(metrics.classification_report(yTest, predictions))

```

	precision	recall	f1-score	support
0	0.88	0.96	0.92	719
1	0.38	0.17	0.24	115
accuracy			0.85	834
macro avg	0.63	0.56	0.58	834
weighted avg	0.81	0.85	0.82	834

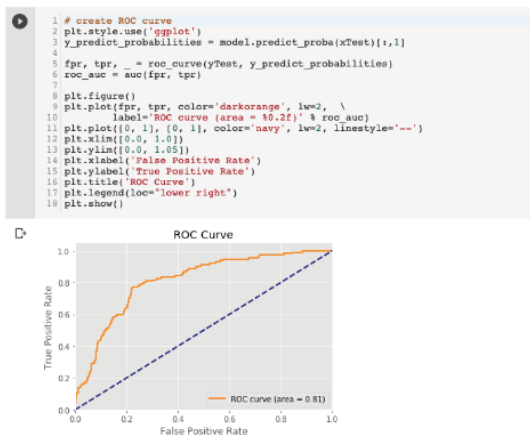
**Gambar 13.19** Melihat matriks kebingungan dan presisi, ingatan, dan skor F1.

Ini berarti kami hanya memprediksi secara akurat 17% pelanggan yang churn. Itu sangat rendah sampai pada titik di mana model tersebut tidak dapat digunakan jika tujuannya adalah untuk memprediksi secara akurat pelanggan mana yang akan churn. Precision adalah seberapa sering prediksi itu benar ketika telah memprediksi ya (yaitu prediksi churn). Ini adalah jumlah positif sejati (20) dibagi dengan semua yang diprediksi positif (20 + 32). Dalam contoh kami, recall adalah 38% (20/52). Ini berarti bahwa setiap kali kami memprediksi churn, hanya 38% dari prediksi ini yang akan benar.

Skor F1 adalah rata-rata tertimbang dari recall dan presisi: yaitu, recall dikalikan presisi dibagi recall ditambah presisi. Skor ini merepresentasikan keseimbangan antara presisi dan recall. Skor F1 umumnya digunakan ketika recall dan presisi sama-sama penting, dan kita perlu menemukan keseimbangan di antara keduanya. Dalam kasus kita, kasus bisnis yang dibahas sebelumnya didasarkan pada recall yang lebih tinggi, bukan pada akurasi, presisi, atau skor F1. Recall lebih penting karena akan memberi tahu kita pelanggan mana yang harus dipertahankan. Jika recall rendah, kita tidak akan mengidentifikasi pelanggan yang cukup untuk dipertahankan agar latihan ini bermanfaat. Pemilihan metrik ini merupakan alasan lain mengapa kasus bisnis harus diuraikan di awal: penting untuk mendefinisikan apa yang perlu dilakukan model AI dan bagaimana outputnya akan digunakan. Kita sekarang tahu bahwa meskipun kita memiliki akurasi model yang cukup baik, model tersebut tidak memadai untuk memenuhi tujuan bisnis kita.

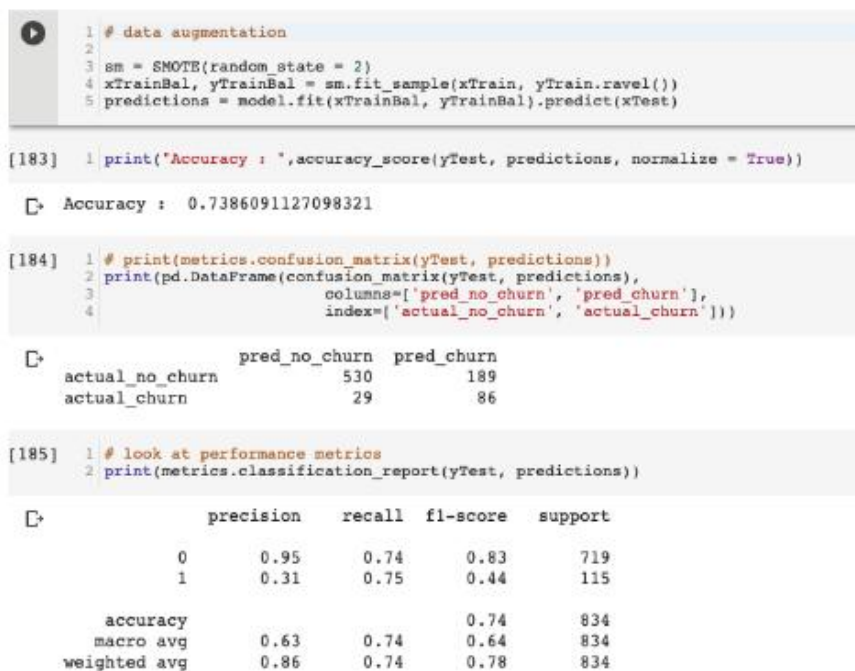
Cara lain untuk memahami seberapa baik kinerja suatu model adalah dengan melihat kurva karakteristik operasi penerima (ROC)-nya (lihat Gambar 13.20). Kurva ROC adalah plot rasio positif benar terhadap rasio positif salah. Model ideal akan memiliki 100% positif benar dan tanpa positif salah. Area di bawah kurva (AUC) adalah ukuran seberapa baik suatu model. Semakin dekat area tersebut ke 1, semakin baik modelnya. Dalam kasus kami, AUC adalah 0,81.

Dengan menerapkan beberapa teknik yang dijelaskan sebelumnya, kami telah menentukan bahwa terdapat masalah dengan model yang kami gunakan. Meskipun tampaknya model tersebut cukup baik dalam memprediksi berapa banyak pelanggan yang akan churn, model tersebut kurang baik dalam memprediksi pelanggan mana yang akan churn. Seperti yang mungkin Anda ingat, penggunaan histogram untuk menganalisis data kami mengungkapkan bahwa terdapat ketidakseimbangan dalam data churn, yang merupakan indikasi bahwa mungkin terdapat masalah dalam memprediksi siapa yang akan churn. Untuk mengurangi ketidakseimbangan ini, ada berbagai teknik yang dapat kami terapkan. Salah satunya adalah dengan menggunakan augmentasi data.



**Gambar 13.20** Kurva karakteristik operasi penerima (ROC) dan area di bawah kurva (AUC).

Pada Gambar 13.21, kita dapat melihat hasil penggunaan metode yang dikenal sebagai *Synthetic Minority Over-Sampling Technique* (SMOTE). Dengan menggunakan SMOTE, kita dapat menghasilkan data pelatihan tambahan untuk pelanggan yang melakukan churn. Kita melakukan ini dengan membuat sampel lain yang serupa dengan data yang ada yang kurang terwakili dalam kasus kita, untuk pelanggan yang melakukan churn. Kita menerapkan gangguan acak kecil pada kolom yang dipilih dan menggunakannya untuk membuat baris baru. Ketika kita menggunakan augmentasi data ini, kita dapat melihat akurasi model telah menurun menjadi 73,9%, tetapi recall telah meningkat secara signifikan dari 17% menjadi 75%. Presisi telah turun sedikit dari 38% menjadi 31%, dan skor F1 telah ditingkatkan dari 0,24 menjadi 0,44.



**Gambar 13.21** Augmentasi data minoritas.

Sebagaimana dibahas di Bab 8, merupakan ide yang baik untuk mencoba berbagai algoritma selama proses pemodelan. Sebagaimana disebutkan sebelumnya, dalam contoh ini, kami menggunakan algoritma regresi logistik. Kami juga dapat menggunakan jaringan saraf tiruan dua lapis atau jaringan saraf tiruan dalam. Kami dapat memilih untuk menggunakan model regresi Proses Gaussian, yang memplot semua fitur dalam kurva Gaussian, atau kurva lonceng, yang memungkinkan kami menggunakan korelasi antar fitur untuk membangun model prediktif. Kami dapat memutuskan untuk menggunakan rata-rata beberapa algoritma, yang dikenal sebagai model ensemble, untuk meningkatkan akurasi prediksi kami.

Dalam contoh kami, kami akan mencoba pendekatan pembelajaran mesin lain yang baru-baru ini populer: peningkatan gradien ekstrem, atau XGBoost. Ternyata ketika kami mencoba ini dengan mengubah baris kode yang menetapkan model kami sebelumnya ke regresi logistik dan mengalihkannya ke Pengklasifikasi XGBoost, kami mendapatkan hasil yang jauh lebih baik.

```

1 # try classification models
2 # model = LogisticRegression(solver = 'lbfgs')
3 model = xgb.XGBClassifier(objective="binary:logistic", random_state=42)
4
5 # train the algorithm on training data and predict using the testing data
6 model.fit(xTrain, yTrain)
7 predictions = model.predict(xTest)
8 print("Accuracy : ",accuracy_score(yTest, predictions, normalize = True))

```

Accuracy : 0.960431654676259

```

[189] 1 # print(metrics.confusion_matrix(yTest, predictions))
2 print(pd.DataFrame(confusion_matrix(yTest, predictions),
3                    columns=['pred_no_churn', 'pred_churn'],
4                    index=['actual_no_churn', 'actual_churn']))

```

	pred_no_churn	pred_churn
actual_no_churn	710	9
actual_churn	24	91

```

[190] 1 # look at performance metrics
2 print(metrics.classification_report(yTest, predictions))

```

	precision	recall	f1-score	support
0	0.97	0.99	0.98	719
1	0.91	0.79	0.85	115
accuracy			0.96	834
macro avg	0.94	0.89	0.91	834
weighted avg	0.96	0.96	0.96	834

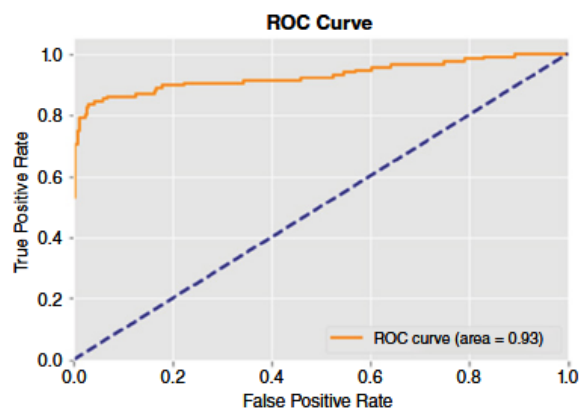
**Gambar 13.22** Mencoba algoritma yang berbeda hanya baris 2 dan 3 di blok pertama yang diubah untuk memilih model yang berbeda.

Pada Gambar 13.22, kita dapat melihat bahwa penggunaan XGBoost telah meningkatkan akurasi model menjadi 96% dari 74%, dan sedikit meningkatkan perolehan data, dari 75% menjadi 79%. Presisi telah meningkat secara signifikan, dari 31% menjadi 91%, dan skor F1 kini berada di angka 0,85, naik dari 0,44. Kurva ROC juga terlihat lebih baik, dengan AUC di angka 0,93, seperti yang ditunjukkan pada Gambar 13.23.

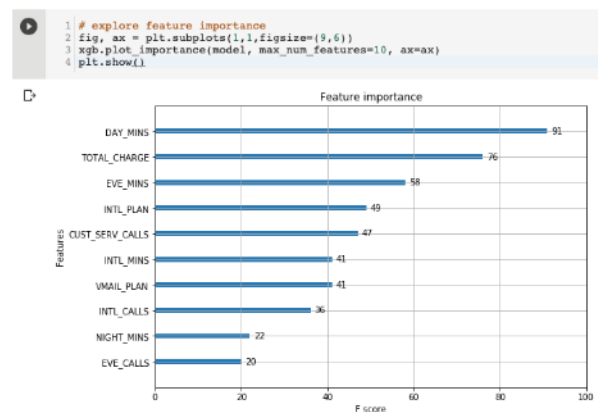
Dengan model ini yang sekarang tampak bermanfaat, kita dapat mencoba dan memahami model sedikit lebih baik dengan melihat fitur apa yang memiliki bobot paling besar

dalam model akhir. Pada Gambar 13.24, kita dapat melihat 10 fitur teratas berdasarkan kepentingan fitur-fitur ini memiliki dampak tertinggi pada prediksi dari model. Model ini menunjukkan bahwa penggunaan dan biaya adalah faktor pendorong terbesar.

Satu hal lain yang mungkin telah kita lakukan adalah menggunakan algoritma regresi linier, yang akan memberikan keluaran angka antara 0 dan 1, alih-alih keluaran biner. Untuk melakukan ini, kita harus menetapkan ambang batas hiperparameter (antara 0 dan 1) di atas yang mana kita akan menganggap keluaran sebagai "churn" dan di bawahnya sebagai "tidak churn." Ini awalnya dapat ditetapkan pada 0,5.



**Gambar 13.23** kurva ROC dan AUC menggunakan XGBoost.



**Gambar 13.24** Pentingnya fitur untuk 10 fitur teratas dalam model.

Menggunakan optimasi hiperparameter (lihat Bab 8) dalam proses pemodelan akan memungkinkan kita untuk meningkatkan akurasi model dan mengurangi positif palsu dan negatif palsu. Ini dapat diterapkan pada XGBoost untuk berpotensi meningkatkan akurasi dan recall; Kami meninggalkannya sebagai latihan bagi pembaca yang tertarik.

Dalam skenario dunia nyata, tim yang membangun model untuk memprediksi churn juga harus mempertimbangkan beberapa potensi masalah lainnya. Misalnya, beberapa pelanggan akan churn terlepas dari insentif retensi yang mereka terima. Selain itu, ada bentuk churn lain misalnya, pelanggan yang menonaktifkan fitur berbayar atau beralih ke paket

dengan harga lebih rendah yang mungkin ingin dimodelkan oleh tim. Sebaiknya juga memodelkan perubahan perilaku pelanggan dari waktu ke waktu. Peningkatan panggilan layanan pelanggan dan penurunan penggunaan mungkin merupakan indikasi kepuasan pelanggan yang lebih rendah.

Dalam contoh ini, kami berfokus pada proses inti pembuatan model dan tidak membahas risiko model dan masalah kewajaran yang perlu ditangani sebelum dan selama proses pembuatan model. Setelah model, seperti model churn XGBoost kami, memenuhi metrik kinerja (misalnya  $\text{recall} > 0,70$  dan  $\text{AUC} > 0,90$ ), kami akan menjalani proses penerapan model, memaparkan model melalui API, lalu mengintegrasikan API ke dalam alur kerja bisnis agar pengguna dapat memanfaatkan wawasan.

Contoh ini menggunakan pembelajaran mesin klasik. Tentu saja, ketika menggunakan pembelajaran mendalam atau pemrosesan bahasa alami untuk mendapatkan hasil, prosesnya akan lebih kompleks dan akan menggunakan algoritma serta kerangka kerja AI yang berbeda. Setelah model dibangun dan divalidasi, model tersebut harus melalui langkah-langkah selanjutnya untuk diterapkan melalui API seperti yang dijelaskan di Bab 8, dengan API yang terintegrasi ke dalam sistem bisnis yang menggunakan keluaran untuk aktivasi yang sesuai.

Terlepas dari kompleksitas pembangunan model AI, pemahaman tentang prosesnya akan memberikan para manajer dan eksekutif pemahaman yang baik tentang apa yang dilakukan tim AI mereka dan alasannya. Ini juga akan membantu mereka membuat keputusan yang lebih baik dan mengetahui metrik mana yang relevan untuk jenis hasil model yang mana. Bab selanjutnya akan berfokus pada masa depan, melihat teknologi AI yang sedang berkembang dan beberapa dampak teknologi yang kini mendunia ini terhadap pekerjaan dan masyarakat.

## **BAB 14**

### **MASA DEPAN MASYARAKAT, PEKERJAAN, DAN AI**

Pada tahun 2014, mendiang Stephen Hawking membuat pernyataan di atas dalam sebuah wawancara dengan BBC News. Kutipan yang agak menakutkan ini sering dianggap merujuk pada hal-hal seperti robot berteknologi AI yang pada akhirnya dapat menguasai dunia. Namun, banyak yang tidak menyadari apa yang dikatakan Hawking selanjutnya. "Saya seorang yang optimis, dan saya percaya bahwa kita dapat menciptakan AI untuk kebaikan dunia," lanjutnya. "Bahwa AI dapat bekerja selaras dengan kita. Kita hanya perlu menyadari bahayanya, mengidentifikasinya, menerapkan praktik dan manajemen terbaik, dan mempersiapkan konsekuensinya jauh-jauh hari."

AI memiliki potensi untuk meningkatkan kualitas hidup setiap orang di planet ini, meningkatkan tingkat pendapatan kita, dan membantu kita hidup lebih lama dan lebih sehat. Diperkirakan lebih dari 90% dari semua interaksi pelanggan dengan bisnis kemungkinan akan memanfaatkan AI dalam beberapa cara di masa depan.

Dengan menggunakan AI, orang akan dapat menyesuaikan produk dan layanan yang mereka butuhkan di berbagai bidang, mulai dari perbankan hingga perawatan kesehatan. Teknologi AI kemungkinan akan digunakan secara luas di lembaga pemerintah dan sistem hukum di seluruh dunia. Ilmuwan AI akan terus mengembangkan perangkat lunak yang lebih canggih untuk memungkinkan kendaraan dikendalikan secara otonom, dan robot serta perangkat lain yang dikendalikan AI akan terus berkembang dalam kecanggihan.

Pada saat yang sama, ada berbagai tantangan yang akan kita hadapi seiring dengan semakin meluasnya penggunaan teknologi AI. Masalah tata kelola, sebagaimana disebutkan dalam Bab 10, merupakan salah satu rintangan kritis yang kita hadapi saat kita mempertimbangkan masa depan AI; kebijakan dan regulasi yang pragmatis namun efektif diperlukan agar AI tidak mengancam otonomi, agensi, atau kemampuan manusia. Regulasi juga dapat mendorong industri AI yang sehat dan berkembang. Hawking sendiri mengutip beberapa pekerjaan legislatif yang sedang dilakukan di Eropa, khususnya yang melibatkan aturan untuk mengatur AI dan robotika, sebagai perkembangan positif.

Agar masa depan AI cerah agar kita dapat memaksimalkan AI dan mengembangkan teknologinya lebih lanjut dengan cara yang tepat kita perlu melakukan berbagai hal. Kita harus menerapkan teknologi AI yang ada serta teknologi baru yang muncul. Kita harus mengelola risiko yang melekat pada AI, termasuk potensi hilangnya pekerjaan; bias kodifikasi; penggunaan AI yang berbahaya, seperti manipulasi jejaring sosial melalui cara-cara seperti deep-fake atau serangan siber; dan konsekuensi tak terduga lainnya. Dan kita harus meningkatkan dan mengembangkan teknologi AI melalui penelitian aktif, baik fundamental maupun terapan.

Dalam bab-bab sebelumnya, kita telah membahas cara menerapkan teknologi AI yang ada ke semua aspek bisnis kita. Dalam bab ini, kita melihat masa depan masyarakat dan pekerjaan, dan bagaimana perkembangan AI dalam waktu dekat yaitu, perkembangan

sebelum kedatangan kecerdasan umum buatan akan memengaruhi kemajuan yang kita harapkan maupun tantangan yang akan kita hadapi.

#### **14.1 AI DAN MASA DEPAN MASYARAKAT**

Di tahun-tahun mendatang, AI akan muncul dalam berbagai aplikasi, mulai dari robotika hingga peningkatan produktivitas hingga teknologi komputasi kuantum yang sedang berkembang. Di bidang robotika, perusahaan seperti Intuitive Surgical telah mengembangkan teknologi, peralatan, dan layanan yang dibantu robot untuk operasi bedah. Keyence dan Daifuku sedang mengembangkan solusi AI dalam otomatisasi pabrik, dan Nvidia serta perusahaan lain sedang mengembangkan mobil tanpa pengemudi. Ada juga inisiatif yang sedang berlangsung di perusahaan-perusahaan seperti Microsoft, Google, IBM, dan lainnya dalam pembelajaran mesin kuantum.

Perawatan medis tentu akan menjadi salah satu bidang di mana AI akan mengubah kehidupan. Robot cerdas akan menjadi semakin cepat dan akurat dan kemungkinan akan memungkinkan dokter untuk melakukan prosedur canggih yang dibantu robot yang mustahil dilakukan saat ini. Robot AI akan merawat para lansia; para ilmuwan telah mengembangkan "robot-kucing" yang akan mengingatkan pasien geriatri untuk minum obat mereka.

AI akan membantu pesawat ruang angkasa dalam perjalanan mereka ke bintang-bintang dan dapat membantu meringankan perubahan iklim di Bumi. Untuk mendukung perkembangan ilmu lingkungan, Microsoft baru-baru ini berkomitmen sebesar Rp 500 miliar untuk program "AI for Earth" yang baru dibuatnya. Yayasan AI for Good nirlaba adalah contoh lain dari sebuah perusahaan yang berusaha untuk bergerak melampaui aplikasi komersial AI "untuk membantu memecahkan masalah sosial, ekonomi, dan lingkungan untuk memberi manfaat bagi masyarakat semaksimal mungkin."

Namun, meskipun mungkin ada perdebatan tentang apa sebenarnya yang akan dicapai AI di masa depan, yang tidak dipertanyakan adalah banyaknya kekhawatiran tentang penggunaan AI yang meluas. Akuntabilitas adalah salah satunya. Gugatan hukum saat ini atas kerugian investasi sebesar Rp 200 miliar karena "kesalahan" algoritmik saat ini sedang diproses di pengadilan. Karena tidak dapat menuntut superkomputer tersebut, investor tersebut mencoba meminta pertanggungjawaban orang yang menjualnya atas kerugiannya. Dan tanggung jawab algoritma telah diuji di Tempe, Arizona, di mana pada Maret 2018, sebuah Uber yang mengemudi sendiri menabrak dan menewaskan seorang wanita yang sedang berjalan dengan sepedanya di seberang jalan. Setahun kemudian, Uber dibebaskan dari segala tuntutan pidana. Sebaliknya, pengemudi keselamatan yang duduk di kursi pengemudi menghadapi tuntutan. Dalam kedua kasus tersebut, diputuskan bahwa algoritma AI dan penciptanya tidak dapat dimintai pertanggungjawaban.

Apa yang akan terjadi, misalnya, jika apa yang dikenal sebagai tata kelola algoritmik diterapkan secara legal, yang memungkinkan, katakanlah, kepolisian untuk melakukan pengawasan terhadap semua warganya, secara otomatis melacak perilaku? Di Tiongkok, hal ini sudah terjadi: Sistem Kredit Sosial (SCS) Tiongkok dirancang untuk melacak segala hal, mulai dari aktivitas internet seseorang hingga apakah ia cukup menghormati orang tuanya, dan

peringkat akhir seseorang digunakan untuk menentukan tunjangan sosialnya. Meskipun pengawasan semacam ini meningkat di lokasi lain, ada juga tempat-tempat seperti San Francisco yang telah melarang penggunaan pengenalan wajah oleh polisi. Di dunia yang terus menyusut, bagaimana peraturan yang saling bertentangan ini akan terus hidup berdampingan?

Kemungkinan AI menangani semua masalah lingkungan kita juga masih bisa diperdebatkan. Satu algoritma terjemahan mesin siap pakai memiliki kebutuhan sumber daya komputasi yang begitu besar sehingga menciptakan jejak karbon yang serupa dengan jumlah bahan bakar yang akan dikonsumsi lima kendaraan selama masa pakainya. Dan ada potensi kerugian lain dari penggunaan AI yang ekstensif. Seiring kita semakin bergantung pada penggunaan pembelajaran mesin dalam ilmu-ilmu dasar, misalnya, kita berisiko hanya memprediksi hasil tanpa benar-benar memahaminya, mengabaikan pengembangan teori-teori yang mendasari untuk menjelaskan fenomena ini sejak awal. Dalam AI, biaya tersirat ini menyebabkan apa yang disebut Jonathan Zittrain sebagai utang intelektual AI. Hal ini serupa dengan utang teknis dalam rekayasa perangkat lunak: mengambil rute pemrograman yang "mudah" tanpa mempertimbangkan pendekatan yang lebih baik yang mungkin membutuhkan waktu lebih lama untuk diimplementasikan tetapi akan lebih bermanfaat di masa mendatang, sehingga menghemat pengerjaan ulang tambahan untuk pemeliharaan dan kelincahan.

## 14.2 AI DAN MASA DEPAN PEKERJAAN

Buku ini telah membahas banyak aspek positif AI di tempat kerja, baik saat ini maupun di masa depan. Ini mencakup hal-hal seperti potensi AI untuk meningkatkan keselamatan kerja dengan menghindari kesalahan yang disebabkan oleh kelalaian manusia, menghilangkan tugas-tugas yang berulang dan membosankan sehingga karyawan dapat bebas melakukan pekerjaan yang lebih kreatif dan memuaskan, serta memanfaatkan chatbot secara lebih efektif untuk memberikan layanan pelanggan yang lebih baik dan mendukung tim yang berhadapan langsung dengan pelanggan.

Tempat kerja yang digerakkan oleh AI tentu akan terlihat berbeda. Mesin akan berinteraksi satu sama lain menggunakan algoritma AI dan membuat keputusan tentang rantai produksi tanpa perlu campur tangan manusia. Bahkan selama revolusi industri, Karl Marx, dalam bab tentang permesinan dan industri modern dari bukunya *Capital. A Critique of Political Economy*, menulis, "Instrumen tenaga kerja, ketika berbentuk mesin, langsung menjadi pesaing pekerja itu sendiri." Namun, "instrumen tenaga kerja" tersebut akan tetap ada, dan begitu pula dengan kecerdasan buatan. Layaknya komputer pribadi dan ponsel pintar, AI sudah menjadi bagian dari kehidupan kita sehari-hari. Kita tak bisa menghentikannya. Jika kita ingin menghadapi kebangkitan AI dengan sukses, ada baiknya mengingat beberapa situasi di mana teknologi, dalam satu bentuk atau lainnya, memengaruhi tempat kerja, dan dampak yang ditimbulkannya.

Ketika mobil ditemukan, kuda bukan satu-satunya yang kehilangan pekerjaan. Pandai besi, perawat kuda, kusir, pedagang pakan ternak, pemilik kandang kuda, pembuat pelana, pembuat roda, pembuat cambuk, pembersih jalan, dan dokter hewan termasuk di antara

orang-orang yang kehilangan pekerjaan. Di sisi lain, lapangan kerja baru tercipta di pabrik-pabrik mobil, dealer mobil, tempat parkir, bengkel, dan pabrik-pabrik di mana orang-orang harus membuat semua komponen yang dibutuhkan mobil agar berfungsi, termasuk ban, pengapian, baterai, dan karburator.

Para pekerja di awal abad ke-20 mencele otomatisasi pabrik, tetapi generasi-generasi selanjutnya akhirnya diuntungkan oleh akhir pekan dua hari tersebut. Orang-orang memandang kemunculan anjungan tunai mandiri (ATM) pada tahun 1970an sebagai bencana bagi para pekerja di industri perbankan ritel. Namun, seiring dengan penurunan biaya cabang, pekerjaan di cabang perbankan meningkat seiring waktu, menjadi kurang bersifat transaksional dan lebih berfokus pada pengelolaan hubungan nasabah.

Hal serupa terjadi saat ini. Dulu ada 600 pedagang ekuitas di kantor pusat Goldman Sachs, tetapi sekarang hanya tinggal dua. Hal ini terjadi karena Goldman menggunakan program perdagangan AI untuk mengotomatiskan perdagangan mata uang dan berjangka. Goldman juga menggunakan AI untuk mengambil alih banyak dari 146 langkah yang diperlukan dalam penawaran umum perdana. Namun, AI telah menciptakan lebih banyak pekerjaan daripada yang hilang, dan banyak yang tidak hilang begitu saja, melainkan justru bertransformasi. Faktanya, menurut McKinsey Global Institute, pada tahun-tahun sebelum 1950, industri otomotif menciptakan 6,9 juta lapangan kerja baru di Amerika Serikat, sementara 623.000 lapangan kerja hilang. Perusahaan konsultan TI Gartner memperkirakan bahwa pada tahun 2020, AI akan menciptakan 2,3 juta lapangan kerja sementara menghilangkan 1,8 juta lapangan kerja. Dan seiring tempat kerja menjadi lebih efisien, manusia akan bebas untuk fokus menggunakan perangkat AI guna meningkatkan efisiensi rantai pasok, pengembangan produk, dan tugas-tugas lainnya. Peningkatan efisiensi ini berpotensi menurunkan harga, sehingga barang-barang tersedia bagi konsumen berpenghasilan rendah. Dan jika janji AI dan otomatisasi sedemikian rupa sehingga seluruh populasi memiliki lebih sedikit pekerjaan, mungkin sudah saatnya untuk mempertimbangkan akhir pekan tiga hari.

Tentu saja ada masalah yang perlu ditangani di sepanjang perjalanan. Terlepas dari apakah jumlah pekerjaan bersih bertambah atau berkurang, kelas pekerjaan tertentu akan sangat terdampak, dan meskipun kita pernah menghadapi masalah serupa di masa lalu, AI menyebabkan berbagai hal terjadi dengan kecepatan yang lebih cepat. Pergeseran ini membutuhkan pendekatan solusi holistik untuk melatih ulang tenaga kerja agar populasi dapat bergerak menuju penciptaan lapangan kerja dan menjauh dari otomatisasi pekerjaan. Kesulitan yang dihadapi oleh mereka yang pekerjaannya memungkinkan industri AI juga mudah diabaikan. Pelabelan data pelatihan dan algoritma pelatihan ulang memakan waktu dan tenaga. Seiring model AI meluas ke bisnis lain, kondisi tenaga kerjanya kemungkinan perlu diatur untuk mencegah munculnya "pabrik-pabrik" AI.

### **14.3 MENGATUR DATA DAN KECERDASAN BUATAN**

Ketika kita memikirkan regulasi dan AI, terkadang kita hanya berpikir dalam hal algoritma. Hal ini tentu perlu diatur untuk melindungi karyawan dan konsumen. Namun,

aturan tentang penggunaan data sama pentingnya, bahkan mungkin lebih penting. Mengingat besarnya jumlah data yang sekarang disimpan secara pribadi, di mana kita tidak memiliki regulasi seputar kerahasiaan atau tanggung jawab, hampir pasti akan ada masalah dalam membagikan informasi tersebut di masa mendatang, yang menghambat pengembangan teknologi terobosan di berbagai bidang seperti perawatan kesehatan. Lebih buruk lagi, pengumpulan dan penggunaan data tersebut dapat menyebabkan penyalahgunaan data atau masalah signifikan lainnya.

Paul Nemitz, salah satu perancang Peraturan Perlindungan Data Umum (GDPR) Uni Eropa, pernah dikutip mengatakan bahwa kita sedang bergerak menuju "dunia di mana teknologi seperti AI menjadi sangat luas dan benar-benar menggabungkan dan menjalankan aturan yang sebagian besar kita jalani" dan bahwa "ketiadaan kerangka seperti itu untuk ekonomi internet telah menyebabkan budaya mengabaikan hukum yang meluas dan membahayakan demokrasi, skandal Facebook Cambridge Analytica hanyalah peringatan terbaru."

Pada Mei 2019, 42 negara dari Organisasi untuk Kerja Sama Ekonomi dan Pembangunan (OECD) menyetujui seperangkat pedoman kebijakan baru untuk pengembangan sistem AI, yang disebut "Rekomendasi Dewan Kecerdasan Buatan." Pedoman ini mempromosikan lima prinsip untuk pengembangan AI yang bertanggung jawab:

1. Pertumbuhan inklusif, pembangunan berkelanjutan dan kesejahteraan untuk memberi manfaat bagi masyarakat.
2. Nilai-nilai yang berpusat pada manusia dan keadilan untuk menghormati supremasi hukum, hak asasi manusia, dan prinsip-prinsip demokrasi.
3. Transparansi dan kemudahan dijelaskan untuk memastikan orang-orang memahami ketika mereka berinteraksi dengan AI atau hasil berbasis AI dan dapat menggugatnya jika perlu.
4. Ketahanan, keamanan, dan keselamatan untuk terus menilai dan memitigasi risiko AI di sepanjang siklus hidup AI.
5. Akuntabilitas untuk meminta pertanggungjawaban perusahaan dan individu yang mengembangkan dan menerapkan teknologi atas berfungsinya sistem ini dengan baik.

Pada KTT tahunan Kelompok Tujuh (G7) (Kanada, Prancis, Jerman, Italia, Jepang, Britania Raya, dan Amerika Serikat) yang diadakan di Prancis pada tahun 2018, Perdana Menteri Prancis Emmanuel Macron dan Perdana Menteri Kanada Justin Trudeau membuat pengumuman bersama tentang pembentukan sekelompok pakar internasional tentang AI yang disebut Panel Internasional tentang Kecerdasan Buatan. Panel tersebut akan dimodelkan pada Panel Internasional tentang Perubahan Iklim. Informasi lebih lanjut tentang panel ini dibagikan dalam sesi pada bulan Mei 2019 yang menjadi tuan rumah bagi para menteri urusan digital G7.

Salah satu RUU pertama Amerika Serikat untuk mengatur AI diperkenalkan pada bulan April 2019 oleh Senator Cory Booker dan Ron Wyden, dengan RUU setara DPR yang disponsori oleh Perwakilan Yvette Clarke. Dikenal sebagai Undang-Undang Akuntabilitas Algoritmik, undang-undang ini akan memerlukan audit sistem pembelajaran mesin untuk bias dan

diskriminasi serta audit semua proses yang melibatkan data sensitif. Perusahaan perlu memastikan tindakan korektif dalam jangka waktu yang wajar ketika masalah tersebut ditemukan dalam audit ini. Data yang dilindungi mencakup informasi identitas pribadi, biometrik, dan genetik. Komisi Perdagangan Federal AS (FTC) akan bertanggung jawab untuk mengawasi kepatuhan, karena FTC juga bertanggung jawab atas perlindungan konsumen dan regulasi antimonopoli. RUU lain yang diperkenalkan di Amerika Serikat pada April 2019 akan melarang praktik desain manipulatif yang mereka tuduhkan terkadang digunakan oleh raksasa teknologi seperti Facebook dan Google untuk membuat pelanggan menyerahkan data mereka.

Negara-negara lain telah merancang atau mengesahkan undang-undang serupa yang dirancang untuk meminta pertanggungjawaban hukum perusahaan teknologi atas algoritma mereka. Mengingat Amerika Serikat merupakan rumah bagi Silicon Valley dan Silicon Alley di New York, Amerika Serikat memiliki peran penting dalam pengembangan regulasi AI internasional, yang semakin menjadi alasan bagi para legislator dan pembuat kebijakan untuk memiliki pemahaman yang mendalam dan bernuansa tentang teknologi AI.

Amerika Serikat juga sangat prihatin tentang mengatur disinformasi dan "deep-fakes," teknologi AI yang memungkinkan pembuatan gambar dan video yang mudah yang tampak nyata tetapi tidak yaitu, visual dari suatu peristiwa yang sebenarnya tidak pernah terjadi, dibuat dengan memanipulasi gambar dengan cara yang semakin canggih. Masalah dengan deep-fakes ada dua. Pertama, mereka dapat mewakili hal-hal yang tidak pernah terjadi, seperti situasi memalukan yang melibatkan tokoh masyarakat yang tidak pernah terjadi. Selain itu, keberadaan mereka memungkinkan beberapa orang untuk meragukan apa yang sebenarnya terjadi di masa lalu. Orang yang ragu dengan pendaratan di bulan tahun 1969, misalnya, mungkin membenarkan keyakinan mereka berdasarkan keberadaan teknologi yang dapat dengan mudah memalsukan perjalanan ke bulan. Beberapa orang mengklaim bahwa deep-fakes telah dibuat untuk tujuan hiburan saja, dan beberapa tentu saja telah dibuat. Tetapi kemampuan yang sangat canggih ini untuk menipu sebagian besar orang pada suatu waktu memiliki implikasi yang mengerikan di berbagai bidang seperti kejahatan dunia maya.

Undang-undang telah diperkenalkan sejak tahun 2018 untuk mengendalikan deepfakes, tetapi belum dapat dipastikan apakah undang-undang tersebut akan berhasil atau bahkan dapat ditegakkan. Regulasi teknologi AI mungkin didorong oleh agenda politik, sehingga individu harus menyusunnya dengan cermat dengan pemahaman yang mendalam tentang AI dan hukum. Menurut sebuah artikel di *Columbia Journalism Review*, Electronic Freedom Foundation khawatir bahwa RUU yang sedang dibahas di Kongres yang dikenal sebagai Deepfakes Accountability Act berpotensi menimbulkan beberapa masalah Amandemen Pertama.

Ada banyak tantangan lain dalam meregulasi AI. Misalnya, belum ada kesepakatan tentang apa itu AI. Teknologi informasi mudah dipindahkan; data dapat berada di satu negara, algoritma di negara lain, dan pengguna di negara ketiga. Akankah regulasi tingkat negara berhasil? Akankah regulasi AI membutuhkan aliansi regional seperti yang dilakukan antara Prancis dan Kanada yang disebutkan sebelumnya? Atau akankah regulasi tersebut mengikuti rekomendasi yang dibuat oleh entitas supranasional seperti OECD? Kebijakan seperti

rekomendasi kebijakan Juni 2019 oleh *American Medical Association* (AMA) untuk memastikan pengawasan dan akuntabilitas bagi peningkatan kecerdasan dalam layanan kesehatan mungkin juga perlu didukung oleh kebijakan badan-badan lain. Inisiatif Global tentang Etika Sistem Otonom dan Cerdas, yang diselenggarakan oleh asosiasi profesional teknis terbesar di dunia, *Institute of Electrical and Electronics Engineers* (IEEE), baru-baru ini meluncurkan "Desain yang Selaras secara Etis, Edisi Pertama: Sebuah Visi untuk Memprioritaskan Kesejahteraan Manusia dengan Sistem Otonom dan Cerdas." Mereka menyebutnya "risalah global yang paling komprehensif dan bersumber dari masyarakat mengenai Etika Sistem Otonom dan Cerdas yang tersedia saat ini."

Transparansi pemangku kepentingan, serta perlunya memastikan bahwa sistem tersebut tidak melanggar hak asasi manusia, hanyalah dua dari sekian banyak isu yang diuraikan dalam risalah yang perlu diselesaikan seiring waktu. Masih banyak pekerjaan yang perlu dilakukan di bidang ini. Sebuah studi terbaru melaporkan, "meskipun terdapat kesepakatan yang jelas bahwa AI harus 'etis', terdapat perdebatan tentang apa yang dimaksud dengan 'AI etis' dan persyaratan etika, standar teknis, serta praktik terbaik apa yang diperlukan untuk mewujudkannya."

#### **14.4 MASA DEPAN AI: MENINGKATKAN TEKNOLOGI AI**

Pertumbuhan ekonomi didorong oleh inovasi teknologi, khususnya penciptaan apa yang disebut sebagai teknologi umum: inovasi seperti kemampuan menghasilkan uap dan tenaga listrik serta mesin pembakaran internal. AI adalah yang terbaru yang dianggap sebagai teknologi umum. Untuk memajukan penggunaannya, kita perlu terus meningkatkan dan mengembangkan teknologi itu sendiri melalui penelitian fundamental dan terapan.

Seiring dengan berkembangnya kemampuan AI, AI akan menemukan lebih banyak aplikasi dalam bisnis dan kehidupan kita sehari-hari. Di antara jenis-jenis AI yang mungkin banyak digunakan di masa depan adalah pembelajaran penguatan, pembelajaran adversarial generatif, pembelajaran terfederasi, pemrosesan bahasa alami, jaringan kapsul, dan pembelajaran mesin kuantum.

##### ***Pembelajaran Penguatan***

Selain pembelajaran mesin terawasi, tanpa pengawasan, atau semi-terawasi (lihat Bab 2), terdapat pendekatan lain: pembelajaran penguatan (RL). Pembelajaran penguatan menyerupai apa yang dilakukan Pavlov saat melatih tikus. Sebagaimana Pavlov memberi hewan-hewannya hadiah berupa pelet makanan ketika mereka berhasil melewati labirin, tindakan atau perilaku optimal dalam sistem pembelajaran mesin diperkuat dengan hadiah yaitu, nilai numerik yang diberikan kepada mereka karena melakukan tindakan tertentu pada waktu tertentu. Algoritme ditugaskan untuk mencoba berbagai urutan tindakan untuk menemukan tindakan optimal yang akan memaksimalkan hadiahnya. Hadiah ini terkait dengan keberhasilan beberapa tujuan yang ingin dicapai oleh algoritma misalnya, memungkinkan robot berjalan dengan sukses atau mobil mengemudi dengan aman.

Yang membuat hal ini sangat menarik adalah pembelajaran penguatan tidak memerlukan data yang telah ada sebelumnya untuk membuat model. Dengan beberapa

instruksi, algoritma ini memungkinkan komputer menganalisis suatu situasi dan kemudian menghasilkan data melalui uji coba berdasarkan situasi tersebut. Jika masalahnya kompleks, algoritma pembelajaran penguatan dapat beradaptasi seiring waktu, jika perlu, untuk meningkatkan imbalannya. Namun, sulit untuk menggunakan pembelajaran penguatan dengan sukses kecuali masalah memiliki struktur imbalan yang jelas dan terukur, dan lingkungan tempat RL akan beroperasi dijelaskan dengan mudah.

Pembelajaran penguatan mendalam menggunakan jaringan saraf dalam bersama dengan kerangka kerja penguatan untuk mencapai kinerja yang hampir setara manusia dalam aktivitas tertentu. Google telah melaporkan bahwa dengan menggunakan teknologi ini, komputer telah belajar bermain gim di konsol Atari 2600 menggunakan sistem penghargaan untuk permainan yang baik. Algoritme ini mencapai kinerja seperti manusia di hampir setengah dari semua gim yang menerapkannya. Pembelajaran penguatan menjadi populer karena keberhasilan sistem AlphaGo dan AlphaZero dari Google DeepMind. Alasan mengapa pembelajaran penguatan (RL) sebagian besar berhasil di bidang yang melibatkan gim dan simulasi adalah karena dalam lingkungan ini, pembelajaran dapat dengan mudah dilakukan melalui coba-coba. Jika, alih-alih simulasi, pembelajaran tersebut berinteraksi dengan pelanggan, pasien, atau mesin fisik yang sebenarnya, ini akan menjadi cara belajar yang sangat mahal.

Pembelajaran penguatan merupakan bagian aktif dari penelitian, dengan penggunaan komersial yang baru-baru ini muncul. Kegunaannya akan terus meningkat seiring perusahaan mulai menggunakannya dalam aplikasi spesifik dalam situasi di mana data berlabel historis tidak tersedia secara luas. Pembelajaran penguatan juga memiliki implikasi dunia nyata yang kuat. Google menggunakan model serupa untuk mengurangi konsumsi daya hingga 40% di pusat datanya. Di perusahaan lain, dua area penerapan RL adalah konfigurasi proses dan pengambilan keputusan berurutan. Misalnya, dengan menganalisis dan mengoptimalkan urutan, RL dapat menentukan konfigurasi lalu lintas yang optimal, yang menentukan secara tepat kapan dan di mana sinyal lalu lintas harus diubah. Model ini juga digunakan untuk mengoptimalkan sumber daya di pusat data dan mengendalikan lalu lintas jaringan. Pada akhirnya, harapannya adalah RL akan mampu memecahkan berbagai macam masalah bisnis. Namun, untuk mencapainya, aplikasi bisnis memerlukan skenario dengan uji coba yang rendah biaya, idealnya menggunakan lingkungan simulasi yang realistis untuk melatih dan menguji agen RL.

### ***Pembelajaran Adversarial Generatif***

Baru-baru ini, sebuah bentuk baru pembelajaran mesin telah dikembangkan yang mampu menghasilkan gambar, ucapan, atau teks baru yang tidak dapat dibedakan dari gambar sebenarnya dalam data pelatihan. Bentuk pembelajaran mesin ini dikenal sebagai pembelajaran adversarial generatif. Model yang digunakan disebut jaringan adversarial generatif (GAN).

GAN, yang pertama kali diperkenalkan oleh Ian Goodfellow pada tahun 2014, sangat berguna karena menawarkan cara baru untuk melakukan pembelajaran tanpa pengawasan. GAN terdiri dari dua jaringan saraf sebuah generator yang menerima masukan dan

menghasilkan sampel baru, dan sebuah diskriminator yang mempelajari seperti apa masukan asli dan bertugas membedakan yang asli dari yang palsu. Ketika kita menggunakan model diskriminatif untuk pembelajaran klasifikasi atau regresi, teknik yang digunakan melibatkan pembelajaran dari sampel berlabel dan mengamati sampel yang tidak berlabel untuk menentukan labelnya. Sebagai alternatif, model diskriminatif dapat menentukan apakah sampel yang tidak berlabel ini merupakan bagian dari kelas berlabel atau tidak. Di sisi lain, model generatif menggunakan data pelatihan dan belajar untuk menghasilkan data yang tampak serupa dengan data pelatihan ini. Hal ini dilakukan dengan menentukan kombinasi fitur mana yang membuat suatu sampel tampak serupa. Tantangannya, yang sering direpresentasikan sebagai permainan, adalah bagi generator untuk menipu diskriminator agar berpikir bahwa data yang dihasilkannya nyata, dan bagi diskriminator untuk berhasil membedakan satu dari yang lain (maka nama "adversarial"). Seiring waktu, generator menjadi lebih baik dalam melakukan tugasnya dan dapat menipu diskriminator. Model GAN, misalnya, dapat mengambil data pelatihan yang terdiri dari gambar wajah dan menghasilkan gambar sintetis yang menyerupai wajah asli, meskipun orang yang seharusnya mereka wakili sebenarnya tidak ada. Ada berbagai aplikasi akademis yang menggunakan GAN, dan teknologinya berkembang pesat, tetapi sejauh ini, aplikasi bisnis yang lebih luas belum muncul.

### ***Pembelajaran Terfederasi***

Aplikasi pembelajaran mesin pada umumnya mengharuskan data dikumpulkan di satu komputer atau di pusat data yaitu, di lingkungan yang relatif kecil dan terpusat. Komputasi tepi memungkinkan kita memproses dan menganalisis data di dekat sumbernya. Ketika data digunakan secara lokal di tepi oleh perangkat seperti termostat pintar aliran data yang dibutuhkan lebih sedikit, sehingga mengurangi lalu lintas jaringan dan waktu respons.

*Pembelajaran terfederasi* adalah istilah yang merujuk pada kemampuan perangkat *Internet of Things* (IoT), seperti ponsel atau drone, untuk berbagi wawasan tanpa membagikan datanya melalui cloud. Hal ini memungkinkan perangkat tepi untuk melatih model, alih-alih membutuhkan model yang telah dilatih sebelumnya. Dalam arti tertentu, pembelajaran terfederasi menghadirkan pembelajaran mesin ke komputasi tepi dengan cara yang belum pernah ada sebelumnya. Model misalnya, termostat Anda mengunduh perangkat lunak terbaru dan kemudian belajar dari data yang dikumpulkannya. Model terenkripsi yang telah dipelajarinya adalah satu-satunya hal yang diunggahnya ke cloud. Data asli tidak pernah meninggalkan termostat Anda. Kemudian, model-model yang dipelajari dari setiap termostat diagregasi (misalnya, dengan merata-ratakan) menjadi satu model, dan model baru didistribusikan ulang ke edge. Dengan cara iteratif ini, model akan semakin baik seiring waktu.

Pembelajaran terfederasi akan memiliki banyak aplikasi di masa mendatang, memainkan peran penting dalam berbagai bidang. Perangkat seluler akan membutuhkan lebih sedikit waktu untuk menghasilkan wawasan baru dan menindaklanjutinya. Keamanan akan meningkat, karena wawasan dapat dibagikan tanpa membagikan data asal wawasan tersebut. Privasi data akan terjaga, karena pengguna tidak perlu lagi mengirimkan informasi sensitif ke berbagai perusahaan melalui jaringan milik penyedia layanan. Efisiensi akan meningkat seiring perangkat edge menjalankan tugas yang paling sesuai, dan ini akan memungkinkan aplikasi

pembelajaran mesin yang lebih privat dan hemat biaya. Pembelajaran terfederasi kemungkinan akan menjadi lebih penting karena kekhawatiran privasi data pengguna individu menjadi lebih umum dan perangkat seluler serta edge menjadi lebih canggih.

### ***Pemrosesan Bahasa Alami***

Pemrosesan bahasa alami (NLP) memungkinkan komputer untuk mengekstrak dan menganalisis informasi dari teks bahasa alami, kemudian menjawab pertanyaan, mengambil informasi, menghasilkan teks, dan menerjemahkan satu bahasa ke bahasa lain. Karena tujuan utama AI adalah memungkinkan komputer dan perangkat pintar untuk memahami dan menerapkan bahasa lisan dan tulisan serta memecahkan masalah, NLP telah menjadi bidang penelitian yang penting. Ada begitu banyak data teks tak terstruktur di dunia sehingga penguraian dengan cepat dan mudah menjadi keharusan. Selain itu, perangkat pintar kecil seperti ponsel mendorong penggunaan bahasa alami sebagai bentuk masukan dan keluaran yang termudah dan paling intuitif.

Kendala utama yang dihadapi NLP adalah keragaman makna yang dapat dimiliki bahasa. Pendekatan NLP saat ini melibatkan representasi kata dan teks sebagai vektor, yang masing-masing merupakan sekumpulan bilangan riil. Hal ini memungkinkan untuk menangkap hubungan suatu kata dengan kata lain, kata apa yang muncul bersama atau di sampingnya, dan seberapa sering. Vektor-vektor ini, yang disebut penyisipan kata, mendukung penerjemahan bahasa serta pencarian dan pengambilan informasi. Namun, penggunaan penyisipan kata memiliki beberapa kekurangan. Salah satu masalahnya adalah penyisipan kata kurang sensitif terhadap konteks; di mana pun sebuah kata muncul, representasinya sama. Ilmuwan komputer baru-baru ini menggunakan enkoder sekuens neural untuk menambahkan informasi kontekstual, yang telah sangat meningkatkan penyisipan kata tradisional.

Hanya dalam beberapa tahun terakhir, kemajuan besar telah memungkinkan model bahasa yang tidak hanya mendukung penambahan data, tetapi juga mensimulasikan jenis penalaran yang sangat mendasar, yang memungkinkan program-program ini untuk memahami materi dan menyajikannya dengan cara yang bermakna kepada manusia. Kemajuan ini baru-baru ini membantu sistem AI bernama AristoBERT lulus ujian sains kelas delapan. Namun, bahkan kemajuan terbaru ini pun belum memungkinkan teknologi NLP berada pada tingkat kinerja yang sama dengan teknologi visi komputer atau pengenalan gambar saat ini. NLP tetap menjadi bidang penelitian yang sangat aktif, dengan investigasi yang dilakukan di area struktur jaringan saraf baru, pembelajaran transfer, cara memanfaatkan pengetahuan dan akal sehat dalam pemahaman bahasa alami, dan metode penambahan data seperti memperkenalkan pengetahuan domain atau menggunakan kamus dan sinonim untuk meningkatkan kinerja.

### ***Jaringan Kapsul***

Jaringan saraf dalam membutuhkan data dalam jumlah besar untuk pelatihan. Dalam banyak kasus, data ini mungkin tidak tersedia atau mungkin mahal untuk diperoleh. Tantangan ini mendorong Geoffrey Hinton dari Google dan para mahasiswanya untuk mengembangkan ide jaringan kapsul. Jaringan kapsul memanfaatkan kelompok neuron kecil, yang dikenal sebagai kapsul, untuk memodelkan hubungan hierarkis dengan lebih baik. Salah satu masalah yang diatasi oleh jaringan kapsul terkadang disebut sebagai "masalah Picasso" dalam

pengenalan gambar. Jika gambar wajah manusia dipotong dan dikolase kembali (seperti lukisan wajah Picasso), jaringan saraf dalam kemungkinan besar akan tetap mengklasifikasikannya sebagai wajah "manusia", karena meskipun dengan data pelatihan dalam jumlah besar, jaringan tersebut tidak sepenuhnya merangkum hubungan hierarkis hidung, mata, mulut, alis, wajah, rambut, dan kepala.

Di sisi lain, jaringan kapsul akan mampu mengenali ini sebagai bukan wajah, dengan data pelatihan yang jauh lebih sedikit, karena ia mempelajari hubungan hierarkis antara bagian-bagian kepala dan kepala itu sendiri, sementara ia tetap mampu mengenali gambar kepala dari berbagai sudut atau sudut pandang. Saat ini, jaringan kapsul ini bekerja sedikit lebih lambat daripada jaringan saraf tradisional, tetapi Hinton berpendapat bahwa pada akhirnya jaringan ini dapat menyediakan cara untuk memecahkan masalah secara lebih efisien dengan menggunakan lebih sedikit data untuk pelatihan.

### ***Pembelajaran Mesin Kuantum***

Komputasi kuantum adalah teknologi mutakhir yang, jika berhasil, dapat secara radikal mengubah tidak hanya kecepatan perhitungan tertentu, tetapi juga paradigma komputasi itu sendiri. Komputasi ini didasarkan pada mekanika kuantum: teori terbaru dan paling diterima tentang cara kerja dunia fisik, yang memungkinkan banyak teknologi modern, mulai dari chip komputer hingga pemutar DVD hingga pembangkit listrik tenaga nuklir. Dikembangkan oleh Erwin Schrödinger dan Werner Heisenberg, mekanika kuantum kemudian dipopulerkan oleh Schrödinger dalam deskripsinya tentang seekor kucing di dalam kotak yang secara bersamaan dapat mati (atau nol) dan hidup (satu).

Komputasi saat ini didasarkan pada bit; bit-bit tersebut memiliki nilai pasti satu atau nol. Dalam komputasi kuantum, alih-alih sebuah bit yang secara definitif bernilai satu atau nol, ia merupakan kombinasi keduanya seperti kucing Schrödinger. Komputasi kuantum menyebut bit-bit kuantum ini, atau qubit, dan hal yang menarik tentang mereka adalah bahwa mereka dapat berada dalam beberapa keadaan pada saat yang bersamaan. Untuk beroperasi di ranah kuantum misterius ini, molekul didinginkan hingga mendekati nol mutlak Kelvin, yang lebih dingin daripada di luar angkasa. Ketika qubit-qubit kemudian terjatuh, meskipun mereka tidak berdekatan satu sama lain, mereka tetap memengaruhi perilaku satu sama lain.

Komputer kuantum menggunakan efek kuantum, seperti koherensi kuantum, untuk memproses informasi. Komputasi kuantum secara teoritis dapat melakukan jauh lebih baik daripada komputer saat ini dalam memecahkan beberapa masalah, termasuk mencari basis data yang tidak diurutkan dan membalikkan matriks yang jarang. Saat ini, itu hanya terjadi di laboratorium pada skala yang melibatkan sangat sedikit qubit, dan hanya masalah tertentu yang dapat dilakukannya. Google baru-baru ini menerbitkan sebuah makalah yang menunjukkan bahwa prosesor kuantum mereka "membutuhkan waktu sekitar 200 detik" untuk menyelesaikan suatu tugas dan bahwa tolok ukur mereka menunjukkan bahwa "tugas yang setara untuk superkomputer klasik yang canggih akan memakan waktu sekitar 10.000 tahun." Jika eksperimen terus berhasil, dampaknya pada AI, dan komputasi secara umum, akan sangat besar. Gagasan menggunakan komputasi kuantum untuk pembelajaran mesin adalah bidang minat yang muncul. Sebuah tim dari IBM telah melakukan uji pembelajaran mesin

sederhana, pertama tanpa menjerat qubit, dan kemudian dengan mereka yang terjat. Dalam pengujian pertama, tingkat kesalahan adalah 5%; pada yang kedua, tingkat kesalahannya adalah 2,5%. Ini mungkin bukan perbedaan yang signifikan, tetapi merupakan indikator bahwa komputasi kuantum dapat mengubah AI di masa depan.

#### **14.5 DAN INI BARU PERMULAAN**

Mengingat intensitas fokus pada kecerdasan buatan, lebih banyak kemajuan akan terus dicapai dalam penelitian fundamental, penelitian terapan, dan perangkat yang lebih baik, dan AI akan terus digunakan untuk sejumlah kasus penggunaan tambahan. Tentu saja ada cukup banyak minat di bidang ini saya baru-baru ini menghadiri kelas tingkat PhD tingkat lanjut tentang pembelajaran mendalam yang diikuti 120 mahasiswa jumlah yang sangat besar untuk departemen pascasarjana mana pun.

Di masa depan, AI akan dibuat kurang bergantung pada data, lebih transparan dan dapat diinterpretasikan, dan kurang bias. Area seperti jaringan kapsul dan pemodelan pengetahuan akan terus berkembang, dan solusi AI yang lebih canggih akan dikembangkan bersama dengan teknologi lain, seperti robotika, pembelajaran mesin kuantum, realitas tertambah, dan IoT. AI akan lebih sering digunakan untuk pengambilan keputusan berkecepatan tinggi dan bergerak menuju operasi otonom. Ini bukan hanya tempat keluaran model AI akan ditampilkan dalam laporan atau diintegrasikan ke dalam aplikasi untuk konsumsi manusia dan pengambilan keputusan, tetapi juga tempat AI akan merasakan apa yang terjadi, membuat keputusan, dan mengambil tindakan berdasarkan hal tersebut. Kita mungkin akan menikmati akhir pekan tiga hari tersebut paling cepat pada tahun 2030, dengan kesehatan yang lebih baik dan standar hidup yang lebih tinggi bagi semua orang berkat AI.

Banyak bidang yang melibatkan teknologi AI akan dipenuhi tantangan yang harus kita atasi, baik sebagai pemimpin bisnis maupun teknologi, maupun sebagai warga negara. Hal ini akan membutuhkan banyak penyesuaian, tetapi AI siap untuk mendorong transformasi bisnis terbesar dalam sejarah kita. Kita sangat beruntung dapat bekerja di masa ketika penggunaan AI dapat memberikan dampak positif yang luar biasa besar bagi bisnis dan kehidupan kita. Saatnya untuk terjun dan mulai mengadaptasi organisasi Anda untuk memanfaatkannya.

## DAFTAR PUSTAKA

- Ai, Y., Chi, Z., Sun, G., Zhou, H., & Kong, T. (2023). The research on non-linear relationship between enterprise digital transformation and stock price crash risk. *The North American Journal of Economics and Finance*, 68, 101984.
- Al-Alawi, A. I., Munir, M. M., & Munir, F. M. (2024, December). Digital Transformation and Competitive Strategies: Insights Into Innovation, SMEs, and AI Startups. In *2024 International Conference on Decision Aid Sciences and Applications (DASA)* (pp. 1-7). IEEE.
- Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2024). AI-powered innovation in digital transformation: Key pillars and industry impact. *Sustainability*, 16(5), 1790.
- Alghamdi, N. A., & Al-Baity, H. H. (2022). Augmented analytics driven by AI: A digital transformation beyond business intelligence. *Sensors*, 22(20), 8071.
- Arugula, B. (2024). AI-Powered Code Generation: Accelerating Digital Transformation in Large Enterprises. *International Journal of AI, BigData, Computational and Management Studies*, 5(2), 48-57.
- Bettoni, A., Matteri, D., Montini, E., Gładysz, B., & Carpanzano, E. (2021). An AI adoption model for SMEs: A conceptual framework. *IFAC-PapersOnLine*, 54(1), 702-708.
- Bevanda, V. (2024). Tailoring Managers' Journeys Through the AI Transformation of SMEs. *Engineering Proceedings*, 79(1), 10.
- Bi, S., Xiao, J., & Deng, T. (2024, December). The Role of AI in Financial Forecasting: ChatGPT's Potential and Challenges. In *Proceedings of the 4th Asia-Pacific Artificial Intelligence and Big Data Forum* (pp. 1064-1070).
- Burström, T., Parida, V., Lahti, T., & Wincent, J. (2021). AI-enabled business-model innovation and transformation in industrial ecosystems: A framework, model and outline for further research. *Journal of Business Research*, 127, 85-95.
- Cooper, R. G. (2024). The AI transformation of product innovation. *Industrial Marketing Management*, 119, 62-74.
- Costa, I., Riccotta, R., Montini, P., Stefani, E., de Souza Goes, R., Gaspar, M. A., ... & Larieira, C. L. C. (2022). The degree of contribution of digital transformation technology on company sustainability areas. *Sustainability*, 14(1), 462.
- Davenport, T. H., & Mittal, N. (2023). How companies can prepare for the coming "AI-first" world. *Strategy & Leadership*, 51(1), 26-30.
- Deep, S., & Zanke, P. (2024). Digital Transformation Strategy with CRM and AI for SMB's Sustainable Growth. *ESP J. Eng. Technol. Adv.(ESP-JETA)*, 4, 9-22.

- Devaraj, S. M. (2024). Cloud, Ai, And Digital Transformation: A Winning Combination. *International Journal Of Computer Engineering And Technology (Ijcet)*, 15(5), 1020-1032.
- Deveau, R., Griffin, S. J., & Reis, S. (2023). AI-powered marketing and sales reach new heights with generative AI. *McKinsey & Company*.
- Dittmar, E. C. (2025). AI as a catalyst for organizational learning: moving beyond tool implementation to learning transformation. *Development and Learning in Organizations: An International Journal*.
- El Khatib, M. (2024). Digital Transformation and Digital Disruption Technologies: The Effect of AI and Machine Learning on Managing Projects. In *Technology Innovation for Business Intelligence and Analytics (TIBIA) Techniques and Practices for Business Intelligence Innovation* (pp. 63-82). Cham: Springer Nature Switzerland.
- Farmakis, T., Lounis, S., Georgoutsou, I. S., Koukopoulous, A., & Mourtos, I. (2024, June). Innovations in Manufacturing Business Models Enabled by Digital and AI-driven Transformation: A Conceptual Framework. In *2024 IEEE International Conference on Engineering, Technology, and Innovation (ICE/ITMC)* (pp. 1-8). IEEE.
- Fenwick, A., Molnar, G., & Frangos, P. (2024). The critical role of HRM in AI-driven digital transformation: a paradigm shift to enable firms to move from AI implementation to human-centric adoption. *Discover Artificial Intelligence*, 4(1), 34.
- Gołęb-Andrzejak, E. (2023). AI-powered digital transformation: Tools, benefits and challenges for marketers—case study of LPP. *Procedia computer science*, 219, 397-404.
- Holmström, J. (2022). From AI to digital transformation: The AI readiness framework. *Business horizons*, 65(3), 329-339.
- Hossin, M. E., Rahman, M. M., Hossain, S., Siddiqa, K. B., Rozario, E., Khair, F. B., ... & Mahmud, F. (2025). Digital Transformation in the USA Leveraging AI and Business Analytics for IT Project Success in the Post-Pandemic Era. *Journal of Posthumanism*, 5(4), 958-976.
- Iansiti, M., & Nadella, S. (2022). Democratizing transformation. *Harvard Business Review*, 100(5-6), 42-49.
- Jan, I. U., Ji, S., & Kim, C. (2023). What (de) motivates customers to use AI-powered conversational agents for shopping? The extended behavioral reasoning perspective. *Journal of Retailing and Consumer Services*, 75, 103440.
- Jay, R. (2024). Enterprise Transformation with AI in the Cloud.
- Jiang, Z. E., Huang, F., & Wu, Q. (2025). The Impact of AI on Corporate Green Transformation: Empirical Evidence from China. *Sustainability*, 17(17), 7782.
- Joel, O. S., Oyewole, A. T., Odunaiya, O. G., Soyombo, O. T., Joel, O. S., Oyewole, A. T., ... & Soyombo, O. T. (2024). The impact of digital transformation on business development

strategies: Trends, challenges, and opportunities analyzed. *World Journal of Advanced Research and Reviews*, 21(3), 617-624.

- Krishna, S. H., Kumar, G. P., Reddy, Y. M., Ayarekar, S., & Lourens, M. (2024, May). Generative AI in Business Analytics by Digital Transformation of Artificial Intelligence Techniques. In *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)* (pp. 1532-1536). IEEE.
- Kulkov, I. (2021). The role of artificial intelligence in business transformation: A case of pharmaceutical companies. *Technology in Society*, 66, 101629.
- Lakarasu, P. (2022). AI-Driven Data Engineering: Automating Data Quality, Lineage, And Transformation In Cloud-Scale Platforms. *Lineage, and Transformation in Cloud-scale Platforms (December 10, 2022)*.
- Lamarre, E., Smaje, K., & Zimmel, R. (2023). *Rewired: the McKinsey Guide to Outcompeting in the Age of Digital and AI*. John Wiley & Sons.
- Landeta Echeberria, A. (2022). AI integration in the digital transformation strategy. In *Artificial intelligence for business: innovation, tools and practices* (pp. 115-140). Cham: Springer International Publishing.
- Ledro, C., Nosella, A., & Dalla Pozza, I. (2023). Integration of AI in CRM: Challenges and guidelines. *Journal of Open Innovation: Technology, Market, and Complexity*, 9(4), 100151.
- Lee, B. G. (2024). Digital and AI Transformation of Industry. In *Understanding the Digital and AI Transformation* (pp. 163-184). Singapore: Springer Nature Singapore.
- Lee, B. G. (2025). *Understanding the Digital and AI Transformation*. Springer Nature Singapore Pte Limited.
- Manduva, V. C. M. (2022). Leveraging AI, ML, and DL for Innovative Business Strategies: A Comprehensive Exploration. *International Journal of Modern Computing*, 5(1), 62-77.
- Mardani, A. D. (2024). The rise of AI in Business: Uncharted Avenues for Digital Transformation. *Equator Journal of Management and Entrepreneurship*, 25-34.
- Mehmood, F., & Zaki, S. (2024). AI Transformation at Philips: Revolutionizing TV Picture Quality.
- Mihai, F., Aleca, O. E., & Gheorghe, M. (2023). Digital transformation based on AI technologies in European union organizations. *Electronics*, 12(11), 2386.
- Oyekunle, D., & Boohene, D. (2024). Digital transformation potential: The role of artificial intelligence in business. *International Journal of Professional Business Review: Int. J. Prof. Bus. Rev.*, 9(3), 1.
- Podgorelec, V., Karakatič, S., Fister Jr, I., Brezočnik, L., Pečnik, Š., & Vrbančič, G. (2022, May). Digital transformation using artificial intelligence and machine learning: an electrical energy consumption case. In *International Conference "New Technologies,*

*Development and Applications*” (pp. 498-504). Cham: Springer International Publishing.

- Quy, V. K., Thanh, B. T., Chehri, A., Linh, D. M., & Tuan, D. A. (2023). AI and digital transformation in higher education: Vision and approach of a specific university in Vietnam. *Sustainability*, 15(14), 11093.
- Sayed-Mouchaweh, M. (2021). *Explainable AI Within the Digital Transformation and Cyber Physical Systems*. Springer International Publishing.
- Stukalina, Y., & Zervina, O. (2023). Business digital transformation in the data-driven economy: enhancing value with ai services.
- Sundar, I. S., Aljohani, N., Choudhary, N., Pokhariya, H. S., Malathy, V., & Arri, H. S. (2023, December). The Advantages of AI-Based Transformation Projects and their Effects on Company Performance. In 2023 IEEE International Conference on ICT in Business Industry & Government (ICTBIG) (pp. 1-5). IEEE.
- Taherizadeh, A., & Beaudry, C. (2023). An emergent grounded theory of AI-driven digital transformation: Canadian SMEs’ perspectives. *Industry and Innovation*, 30(9), 1244-1273.
- Tominc, P., Oreški, D., & Rožman, M. (2023). Artificial intelligence and agility-based model for successful project implementation and company competitiveness. *Information*, 14(6), 337.
- Wang, Y., & Su, X. (2021). Driving factors of digital transformation for manufacturing enterprises: a multi-case study from China. *International Journal of Technology Management*, 87(2-4), 229-253.
- Xiao, C. (2023). How Has the Digital Transformation Affected the Organization and Operation of a Business. *Law and Economy*, 2(10), 41-48.
- Xu, P., & Li, T. (2025). The effect of technological core executives on industrial AI transformation: evidence from Chinese manufacturing enterprises. *Chinese Management Studies*.

# TRANSFORMASI AI (Artificial Intelligence)

PADA

# PERUSAHAAN

**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 978-634-7227-41-6 (PDF)



9

786347

227416