

Bab 02 — Peta AI, ML, DL, dan RL

Status: Draft lengkap v0.3 untuk review editorial Bagian buku: Level A — Orientasi AI Target pembaca: Pembaca yang sudah memahami Bab 1 dan ingin punya peta mental yang rapi sebelum masuk Python, matematika, ML, DL, dan RL.

Cara membaca bab ini

Bab 1 menjawab “apa itu AI?”. Bab 2 menjawab pertanyaan lanjutan yang sering membuat pemula tersesat:

- AI, Machine Learning, Deep Learning, Data Science, Generative AI, dan Reinforcement Learning itu saling tumpang tindih atau berbeda?
- Kapan memakai rule-based, supervised learning, unsupervised learning, deep learning, generative model, atau RL?
- Mengapa tidak semua masalah harus diselesaikan dengan model paling canggih?

Bab ini memakai pola subbab belajar. Setiap subbab punya ilustrasi, catatan kata kunci, dan tes cepat.

Catatan biru — tujuan bab:

Setelah bab ini, kamu punya “peta kota AI”. Kamu belum menguasai semua jalan, tetapi sudah tahu area mana yang disebut AI, ML, DL, Generative AI, Data Science, dan RL.

Pendalaman awal — taksonomi sebagai peta keputusan teknis

Taksonomi AI bukan sekadar daftar istilah. Ia adalah peta untuk memilih pendekatan. Ketika seseorang berkata “pakai AI”, kita harus bertanya: apakah masalahnya butuh aturan simbolik, analisis data, supervised learning, unsupervised learning, deep learning, generative model, atau reinforcement learning? Salah peta berarti salah alat.

Secara set sederhana, kita bisa membayangkan:

AI = payung sistem yang mengejar perilaku cerdas
ML \subset AI = sistem belajar pola dari data
DL \subset ML = ML dengan neural network berlapis
RL \subset ML/AI = belajar aksi dari reward

Tetapi relasi ini tidak selalu sempurna seperti lingkaran buku teks. Ada AI symbolic yang bukan ML. Ada data science yang bukan AI. Ada generative AI yang sering berbasis deep learning, tetapi penggunaannya bisa sebagai alat kreatif, pencarian, atau agent.

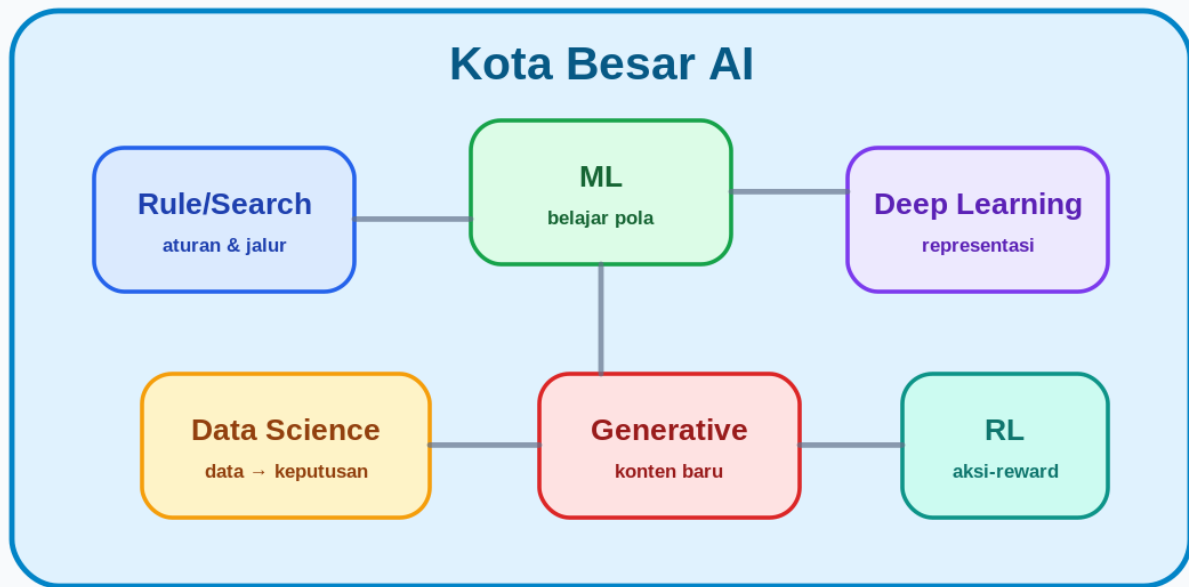
Peta yang baik harus menjawab tiga pertanyaan:

Input apa yang tersedia?
Output/aksi apa yang dibutuhkan?
Feedback apa yang dimiliki: label, struktur, reward, atau tidak ada?

Subbab 1 — Mengapa peta AI dibutuhkan?

Bayangkan kamu baru pertama kali ke Jakarta. Jika hanya diberi nama tempat — Monas, Blok M, Kota Tua, Senayan — kamu mungkin tahu nama-namanya, tetapi belum tahu hubungan antarwilayah. Mana pusat, mana jalur yang dekat, mana yang macet, mana yang cocok untuk tujuan tertentu. Belajar AI juga begitu. Banyak istilah terdengar keren: Machine Learning, Deep

Learning, LLM, RAG, Agent, Data Science, Reinforcement Learning, Computer Vision, NLP. Tanpa peta, semua terasa seperti tumpukan kata.



Peta kota konsep AI

AI adalah payung besar. Di dalamnya ada banyak pendekatan. Beberapa pendekatan berbasis aturan dan pencarian. Sebagian belajar dari data. Sebagian menghasilkan konten baru. Sebagian belajar dari aksi dan reward. Data Science beririsan dengan AI karena membantu memahami data, melakukan eksperimen, dan mengambil keputusan berbasis bukti.

Catatan hijau — kata kunci: Peta Mental

Peta mental bukan untuk menghafal definisi kaku. Peta mental membantu memilih alat yang tepat untuk masalah yang tepat.

Contoh:

Masalah	Pendekatan yang mungkin	Alasan awal
Validasi umur minimal pengguna	Rule-based	Aturannya jelas.
Prediksi stok warung minggu depan	Supervised ML/regression	Ada data historis dan target angka.
Mengelompokkan pelanggan tanpa label	Unsupervised learning	Tidak ada label, ingin mencari pola.
Mengenali foto penyakit daun	Deep learning/CV	Input gambar kompleks.
Membuat draft caption promosi	Generative AI	Output berupa teks baru.
Agent mencari rute terbaik di simulasi	Reinforcement Learning	Ada aksi berurutan dan reward.

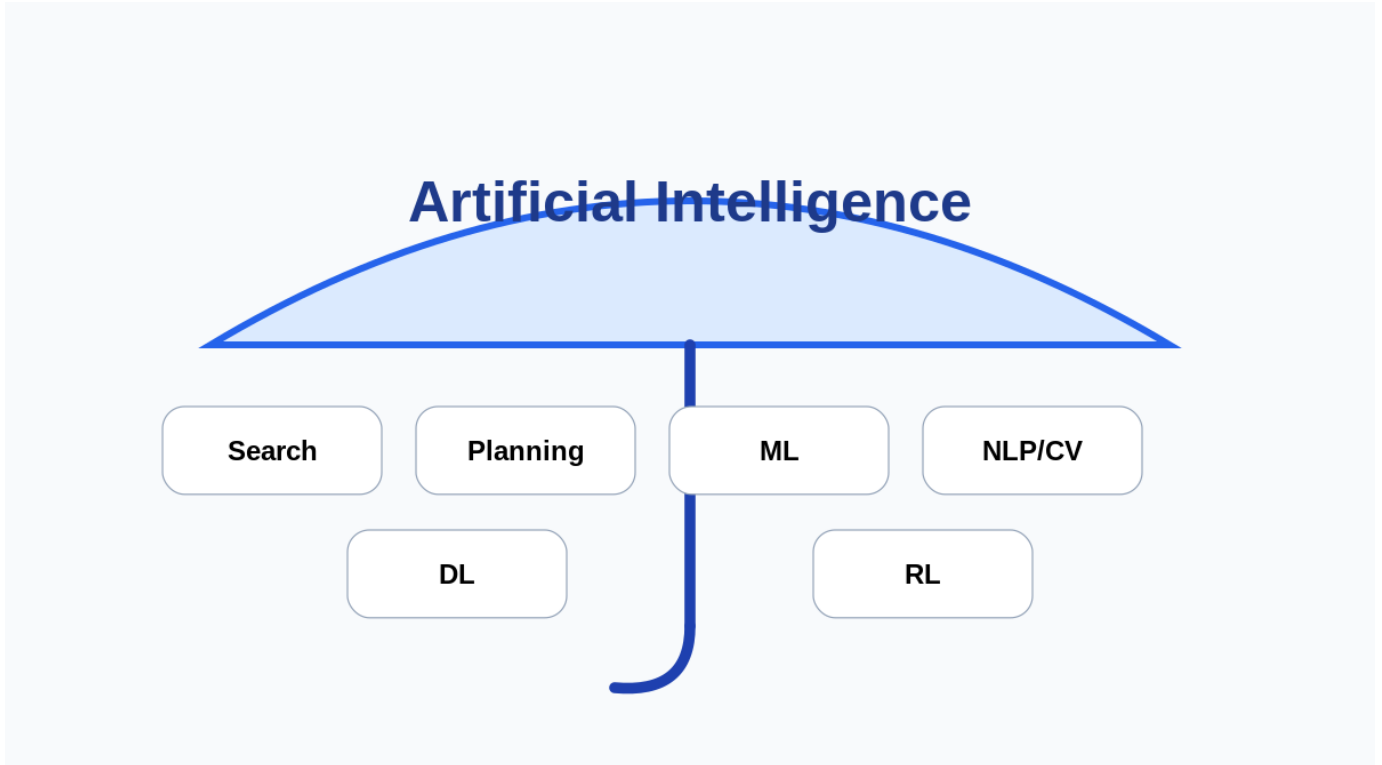
Inti subbab: peta AI membuat kita tidak salah kaprah. Tidak semua masalah butuh deep learning. Tidak semua chatbot adalah “cerdas” dalam arti manusia. Tidak semua analisis data adalah AI.

Tes cepat subbab 1

1. Mengapa pemula perlu peta AI sebelum belajar algoritma?
2. Berikan satu contoh masalah yang cukup diselesaikan dengan aturan sederhana.
3. Apa bahaya memakai pendekatan paling canggih tanpa memahami masalah?

Subbab 2 — AI sebagai payung besar

AI adalah bidang besar yang berusaha membuat sistem mampu melakukan tugas yang biasanya diasosiasikan dengan kecerdasan: menalar, mencari solusi, belajar, memahami pola, berbahasa, melihat, merencanakan, atau mengambil keputusan [R1][R2]. Karena itu, AI tidak sama dengan satu algoritma tertentu.



AI sebagai payung besar

Di bawah payung AI, kita bisa menemukan:

- Search dan planning: mencari jalur, jadwal, atau urutan tindakan.
- Knowledge representation: menyimpan pengetahuan dalam aturan, graf, atau logika.
- Machine Learning: belajar pola dari data.
- Deep Learning: ML berbasis neural network berlapis.
- Natural Language Processing: memproses bahasa manusia.
- Computer Vision: memproses gambar/video.
- Robotics: menghubungkan persepsi, keputusan, dan aksi fisik.
- Reinforcement Learning: belajar dari aksi dan reward.

Catatan kuning — kata kunci: Payung

AI adalah payung. Machine Learning adalah salah satu keluarga penting di bawahnya, tetapi bukan satu-satunya isi AI.

Poin penting: AI lebih luas daripada ML. Sistem catur klasik bisa memakai pencarian dan heuristik tanpa belajar dari data besar. Sistem perencanaan rute bisa memakai algoritma graph search. Expert system bisa memakai aturan dari pakar. Semuanya dapat berada dalam keluarga AI bila sistem melakukan penalaran/pemilihan aksi untuk mencapai tujuan.

Namun, AI modern sering identik dengan ML karena data digital sangat banyak dan metode pembelajaran dari data berkembang pesat. Buku ini tetap menjaga peta lengkap: kita menghargai rule-based/search, lalu masuk ML, DL, generative AI, dan RL secara bertahap.

Formulasi payung AI

AI dapat mencakup sistem berbasis aturan, pencarian, probabilitas, pembelajaran, optimisasi, dan generasi. Cara praktis membaca AI:

AI system = perception + reasoning/learning + action + evaluation

Tidak semua sistem punya keempatnya secara kuat. OCR punya persepsi. Sistem pakar punya reasoning aturan. Model ML punya learning. Agent punya action loop.

Contoh klasifikasi

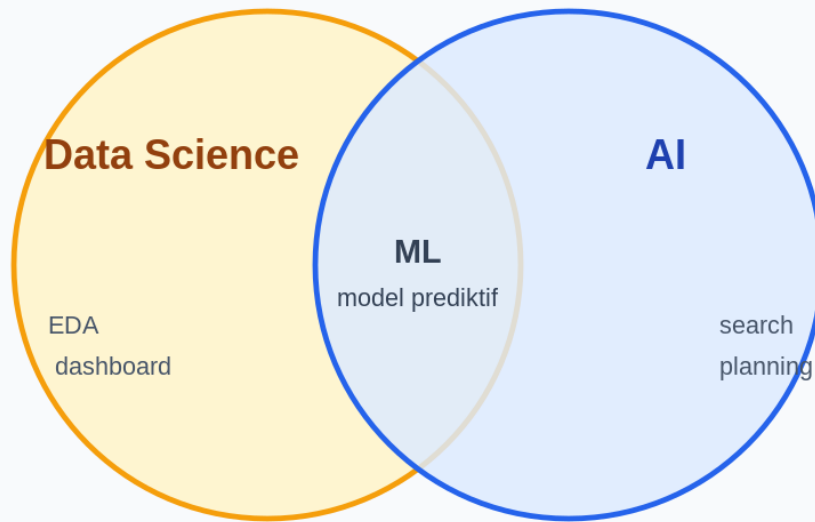
Sistem	AI?	ML?	Alasan
Kalkulator biasa	tidak	tidak	aturan aritmetika tetap
Sistem pakar diagnosis berbasis aturan	ya	tidak wajib	menalar dari aturan
Model prediksi stok dari data	ya	ya	belajar pola historis
Chatbot LLM	ya	ya/DL	model deep learning generatif

Tes cepat subbab 2

1. Mengapa AI lebih luas daripada Machine Learning?
2. Sebutkan dua contoh AI yang tidak harus belajar dari data besar.
3. Apa yang membuat ML sangat dominan dalam AI modern?

Subbab 3 — Data Science: bukan AI, tetapi sahabat dekat AI

Data Science adalah disiplin untuk mengumpulkan, membersihkan, menganalisis, memvisualisasikan, memodelkan, dan mengomunikasikan data agar keputusan lebih baik. Data Science dapat memakai statistik, visualisasi, eksperimen, database, domain knowledge, dan kadang Machine Learning [R3].



Irisan Data Science dan AI

Bayangkan pemilik jaringan warung ingin tahu mengapa penjualan minuman turun. Ia tidak langsung butuh AI. Ia mungkin perlu:

1. mengecek data kasir,
2. membersihkan nama produk yang tidak konsisten,
3. membuat grafik penjualan per hari,
4. membandingkan cuaca/libur/sekolah,
5. menulis rekomendasi bisnis.

Itu Data Science. Jika kemudian ia membuat model untuk memprediksi penjualan besok, barulah masuk wilayah ML.

Catatan ungu — kata kunci: Keputusan Berbasis Data

Data Science tidak selalu membuat model AI. Kadang grafik sederhana dan pertanyaan bisnis yang tepat lebih bernilai daripada model kompleks.

Perbedaan praktis:

Kegiatan	Data Science?	AI/ML?	Catatan
Membersihkan data transaksi	Ya	Belum tentu	Fondasi sebelum model.

Kegiatan	Data Science?	AI/ML?	Catatan
Membuat dashboard penjualan	Ya	Belum tentu	Visualisasi dan monitoring.
A/B testing harga promo	Ya	Belum tentu	Eksperimen bisnis.
Prediksi churn pelanggan	Ya	Ya	Biasanya supervised ML.
Chatbot jawab pertanyaan pelanggan	Bisa	Ya	NLP/generative AI.

Tes cepat subbab 3

1. Apa perbedaan Data Science dan Machine Learning?
2. Mengapa dashboard sederhana kadang lebih berguna daripada model AI?
3. Berikan contoh pekerjaan Data Science yang belum tentu AI.

Subbab 4 — Machine Learning: belajar pola dari data

Machine Learning adalah cara membuat sistem belajar pola dari data agar dapat membuat prediksi, klasifikasi, rekomendasi, atau keputusan pada data baru [R4]. ML tidak berarti komputer “mengerti” seperti manusia. ML berarti model menyesuaikan parameter berdasarkan contoh dan metrik tertentu.



Pipeline Machine Learning

Pipeline umum ML:

1. Problem framing: apa yang ingin diprediksi dan mengapa?
2. Data: dari mana data berasal dan apakah legal/etis dipakai?
3. Feature/representasi: informasi apa yang dipakai model?
4. Model: algoritma yang belajar pola.
5. Training: proses menyesuaikan parameter.
6. Evaluation: mengukur performa dengan data yang tidak dipakai belajar.
7. Deployment/monitoring: memakai model dan memantau penurunan kualitas.

Catatan hijau — kata kunci: Generalization

Tujuan ML bukan menghafal data lama, tetapi bekerja cukup baik pada data baru yang mirip dunia nyata.

Contoh bahasa manusia:

Data lama: penjualan Senin-Minggu, cuaca, tanggal gajian, libur sekolah
Target: jumlah es teh yang terjual besok
Model belajar: pola hubungan antara kondisi dan penjualan
Output: prediksi stok es teh besok

Kesalahan umum pemula adalah mengira ML hanya soal memilih algoritma. Padahal keberhasilan ML lebih sering ditentukan oleh framing masalah, kualitas data, evaluasi, dan pemahaman domain.

Persamaan inti ML

Training: cari f sehingga $f(x) \approx y$
Loss: $L(f(x), y)$
Tujuan: minimalkan rata-rata loss pada data baru

ML berbeda dari statistik klasik bukan karena statistik tidak belajar dari data, tetapi karena ML sangat menekankan prediksi, generalisasi, dan komputasi pada data besar.

Contoh hitung baseline

Target penjualan [30, 40, 50]. Baseline mean:

$$\text{mean} = (30+40+50)/3 = 40$$

$$\text{prediksi} = [40,40,40]$$

$$\text{MAE} = (10+0+10)/3 = 6,67$$

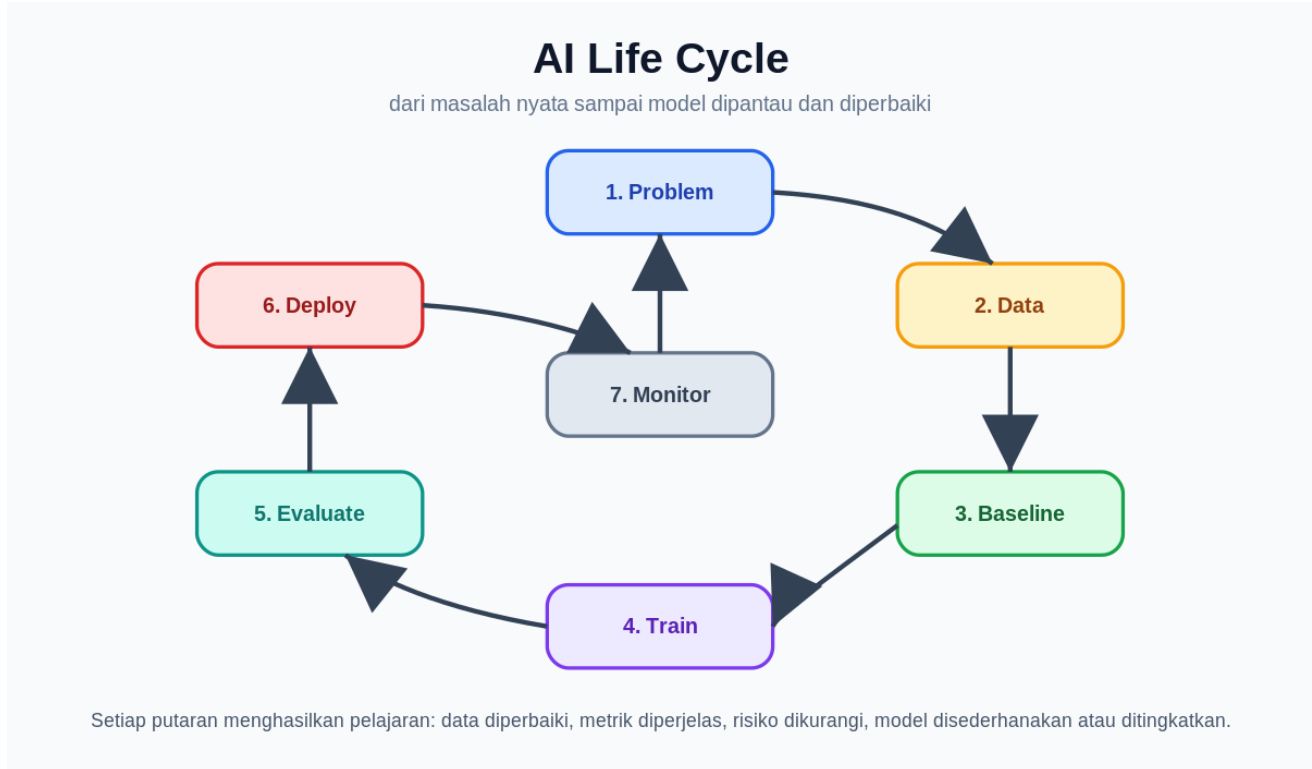
Model ML harus mengalahkan baseline ini agar berguna.

Tes cepat subbab 4

1. Apa arti “belajar” dalam Machine Learning?
2. Mengapa evaluasi harus memakai data yang tidak dipakai training?
3. Apa beda menghafal data dan generalization?

Subbab 4B — AI Life Cycle: dari ide sampai model dipantau

Setelah tahu peta AI, pembaca perlu memahami satu pola kerja yang akan muncul berulang di seluruh buku: AI Life Cycle. AI tidak dimulai dari memilih algoritma. AI dimulai dari masalah nyata, data yang dapat dipercaya, baseline sederhana, evaluasi jujur, deployment yang aman, lalu monitoring.



AI Life Cycle

Siklus hidup AI yang sehat:

1. Problem framing: tulis masalah, pengguna, dampak, dan metrik sukses.
2. Data understanding: cek sumber data, izin, kualitas, bias, dan missing value.
3. Baseline: buat solusi paling sederhana sebagai pembanding.
4. Training/experiment: latih model, catat konfigurasi, seed, versi data, dan hasil.
5. Evaluation: uji pada data yang tidak dipakai training dan cek error penting.
6. Deployment: gunakan model secara bertahap dengan kontrol manusia bila berdampak tinggi.
7. Monitoring: pantau performa, drift, keluhan pengguna, dan risiko setelah model dipakai.

Catatan biru — kata kunci: Life Cycle

Model AI bukan barang sekali jadi. Model adalah sistem hidup yang perlu dicek, diperbaiki, dan kadang dihentikan bila tidak aman.

Contoh sederhana: model prediksi stok es teh di warung. Masalahnya bukan “pakai algoritma X”, tetapi “berapa stok agar tidak sering habis dan tidak banyak terbuang”. Data perlu dicek: apakah catatan penjualan konsisten? Apakah hari libur tercatat? Baseline bisa berupa rata-rata penjualan 7 hari terakhir. Model baru hanya layak dipakai bila mengalahkan baseline dan tidak membuat keputusan berisiko.

Tahap	Pertanyaan praktis	Output
Problem	Keputusan apa yang dibantu?	problem statement
Data	Data legal, bersih, representatif?	data card sederhana
Baseline	Solusi sederhana seberapa baik?	angka pembandingan
Training	Eksperimen apa yang dicoba?	model + catatan run
Evaluation	Salahnya di mana?	metrik + error analysis
Deployment	Siapa mengawasi?	rencana rilis aman
Monitoring	Apakah performa turun?	dashboard + tindakan koreksi

Lifecycle dengan titik keputusan

problem → data → baseline → model → validasi → deploy → monitor → perbaiki

Setiap panah punya risiko. Dari problem ke data: target bisa salah. Dari data ke baseline: data bisa bias. Dari model ke validasi: leakage bisa muncul. Dari deploy ke monitor: distribusi dunia nyata bisa berubah.

Latihan audit lifecycle

Untuk kasus “prediksi siswa butuh bantuan”, isi:

problem: ...
data: ...
label: ...
baseline: ...
metrik utama: ...
risiko: ...
monitoring: ...

Tes cepat subbab 4B

1. Mengapa AI Life Cycle tidak dimulai dari memilih algoritma?
2. Apa fungsi baseline dalam proyek AI?
3. Mengapa monitoring tetap diperlukan setelah model dipakai?

Subbab 5 — Tiga keluarga awal ML: supervised, unsupervised, reinforcement

Peta ML sering dimulai dari tiga keluarga besar: supervised learning, unsupervised learning, dan reinforcement learning [R4][R8]. Pembagian ini tidak sempurna, tetapi sangat membantu pemula.



Tiga keluarga ML

Supervised learning memakai contoh input dan jawaban benar. Contoh: data rumah beserta harga jualnya; pesan beserta label spam/bukan spam; foto daun beserta label penyakit.

Unsupervised learning tidak diberi jawaban benar. Model mencari pola, kelompok, struktur, atau representasi. Contoh: mengelompokkan pelanggan berdasarkan pola belanja tanpa tahu segmen sebelumnya.

Reinforcement learning belajar dari aksi, keadaan, dan reward. Agent mencoba tindakan, menerima akibat, lalu memperbaiki strategi.

Catatan biru — kata kunci: Label

Pertanyaan awal yang bagus: “Apakah kita punya jawaban benar untuk contoh data?” Jika ya, kemungkinan supervised. Jika tidak, mungkin unsupervised. Jika masalahnya aksi berurutan dan reward, pertimbangkan RL.

Keluarga	Pertanyaan utama	Contoh output
Supervised	“Jawaban benar dari contoh ini apa?”	label atau angka prediksi
Unsupervised	“Pola tersembunyi apa di data ini?”	cluster, embedding, anomaly
Reinforcement	“Aksi mana yang memberi reward terbaik?”	policy/strategi

Perbedaan matematis singkat

Supervised: data = (x, y) , belajar $f(x) \rightarrow y$
Unsupervised: data = x , cari struktur z atau cluster c
RL: data muncul dari interaksi (state, action, reward)

Contoh pemilihan

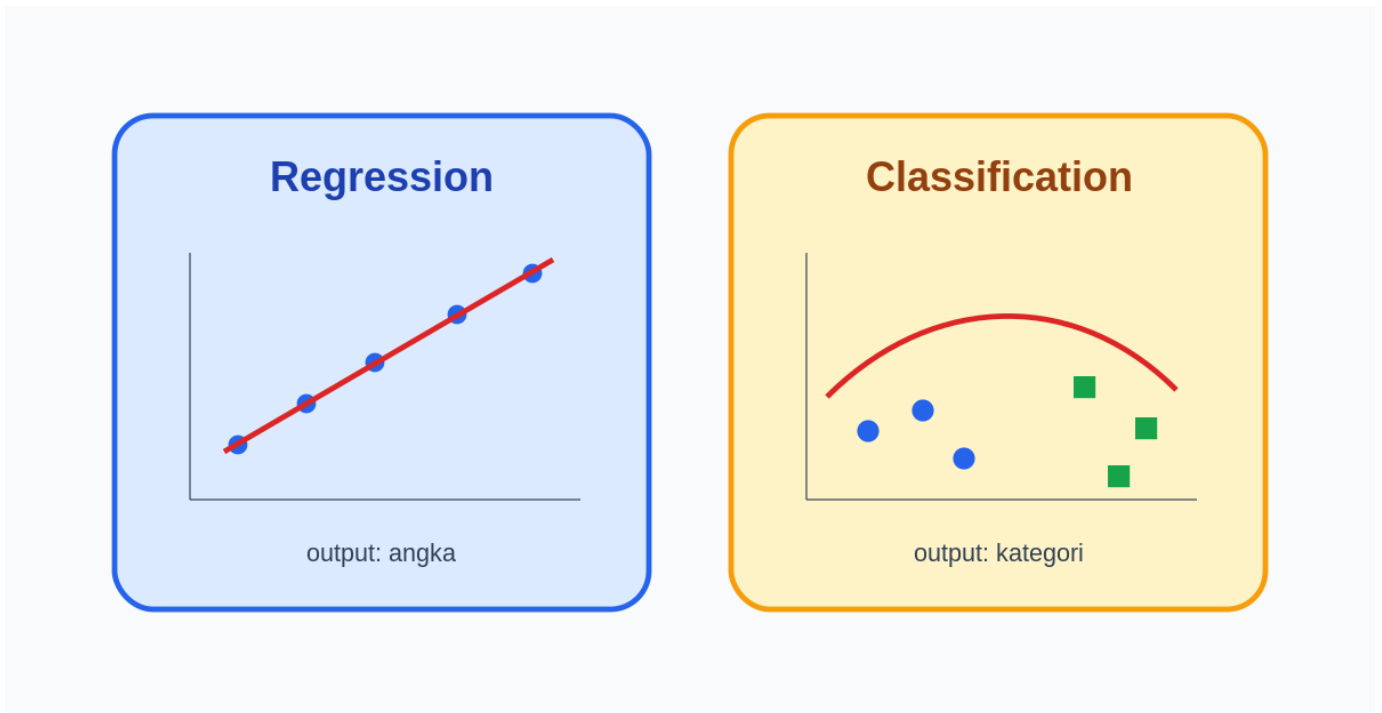
- Ada label “beli/tidak” → supervised classification.
- Tidak ada label, ingin segmentasi pelanggan → unsupervised clustering.
- Agent memilih promo berurutan dan mendapat reward penjualan → reinforcement learning atau bandit.

Tes cepat subbab 5

1. Apa ciri utama supervised learning?
2. Jika data tidak punya label, keluarga apa yang mungkin dipakai?
3. Mengapa RL berbeda dari supervised learning?

Subbab 6 — Supervised learning: regression dan classification

Supervised learning adalah keluarga ML yang paling mudah dijelaskan karena ada pasangan input → jawaban benar. Dua tugas paling umum adalah regression dan classification.



Regression dan classification

Regression memprediksi angka kontinu. Contoh:

- harga rumah,
- jumlah stok yang perlu dibeli,
- estimasi waktu pengiriman,
- prediksi konsumsi listrik.

Classification memilih kelas/label. Contoh:

- spam atau bukan spam,
- pelanggan churn atau tidak,
- daun sehat atau sakit,
- transaksi normal atau mencurigakan.

Catatan kuning — kata kunci: Target

Jika target berupa angka, pikirkan regression. Jika target berupa kategori, pikirkan classification.

Contoh tabel kecil:

Input	Target	Jenis tugas
luas rumah, lokasi, kamar	harga rumah	Regression
isi SMS, jumlah tautan	spam/bukan spam	Classification
umur mesin, jam operasi	biaya perawatan	Regression

Input	Target	Jenis tugas
foto produk	layak jual/tidak	Classification

Metrik juga berbeda. Regression sering memakai MAE/RMSE. Classification sering memakai accuracy, precision, recall, F1, dan confusion matrix. Nanti kita pelajari lebih rinci di Bab 06-07.

Contoh hitung klasifikasi

Aktual [1, 0, 1, 1], prediksi [1, 0, 0, 1]:

TP=2, TN=1, FN=1, FP=0
accuracy = $(TP+TN)/4 = 3/4 = 0,75$
recall = $TP/(TP+FN) = 2/3 = 0,667$

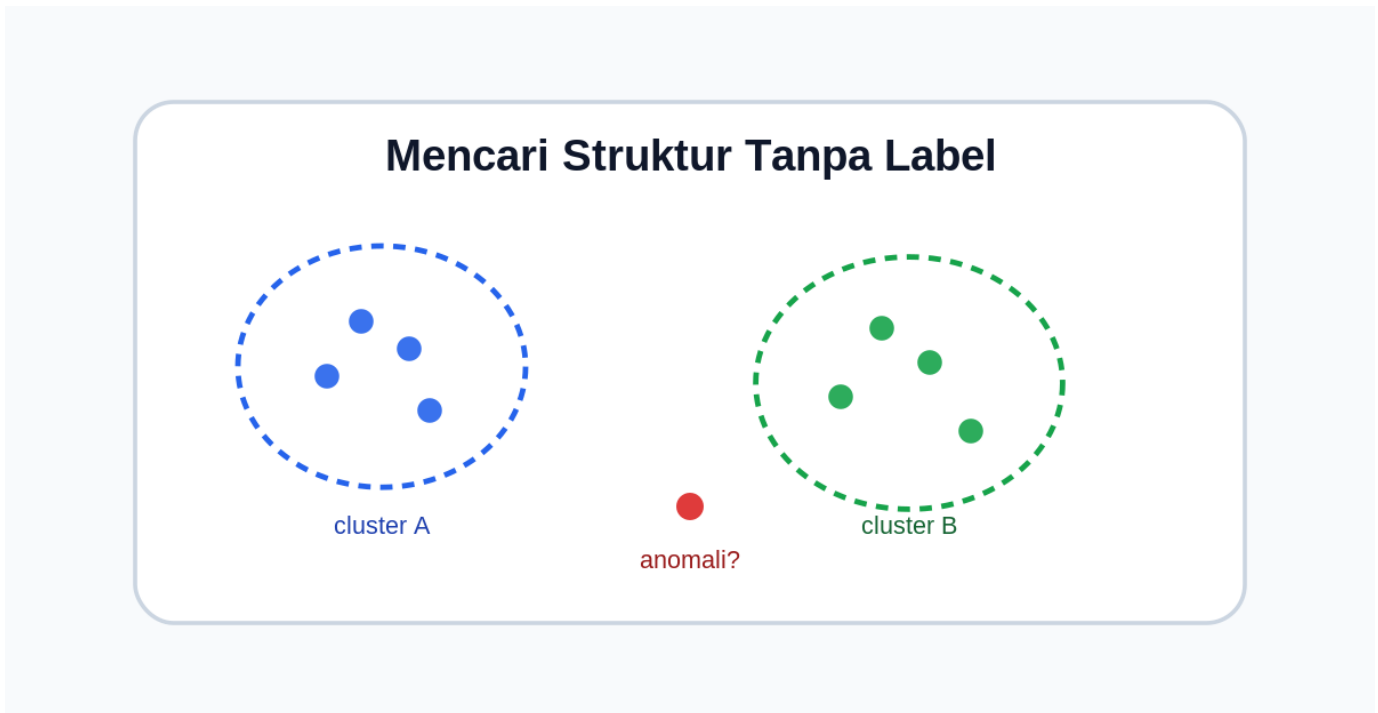
Accuracy saja tidak cukup. Jika kelas positif penting, recall rendah perlu perhatian.

Tes cepat subbab 6

1. Apa perbedaan regression dan classification?
2. Prediksi “jumlah penumpang besok” termasuk tugas apa?
3. Deteksi “transaksi fraud atau bukan” termasuk tugas apa?

Subbab 7 — Unsupervised learning: mencari struktur tanpa label

Unsupervised learning dipakai ketika kita punya data, tetapi tidak punya jawaban benar untuk setiap contoh. Tujuannya mencari struktur: kelompok, pola, dimensi penting, atau data yang aneh.



Contoh UMKM: toko punya data pelanggan berisi frekuensi belanja, rata-rata transaksi, jenis produk favorit, dan lokasi. Belum ada label “pelanggan hemat”, “pelanggan premium”, atau “pelanggan musiman”. Dengan clustering, kita bisa mencari kelompok alami. Hasil cluster bukan kebenaran mutlak; itu alat bantu analisis.

Catatan ungu — kata kunci: Struktur Tersembunyi

Unsupervised learning tidak memberi “jawaban akhir”. Ia memberi peta awal agar manusia bisa menafsirkan pola.

Keluarga unsupervised:

- Clustering: mencari kelompok data.
- Dimensionality reduction: merangkum banyak fitur menjadi dimensi lebih sedikit.
- Anomaly detection: mencari data yang tidak biasa.
- Association patterns: mencari barang/peristiwa yang sering muncul bersama.

Risiko: cluster bisa terlihat meyakinkan padahal tidak bermakna bisnis. Karena itu, hasil unsupervised harus dibaca bersama domain knowledge.

Contoh jarak untuk clustering

Dua pelanggan:

A = [belanja=10, diskon=2]
B = [belanja=12, diskon=3]
C = [belanja=2, diskon=9]

Jarak A-B:

$$\text{sqrt}((10-12)^2 + (2-3)^2) = \text{sqrt}(4+1)=2,24$$

Jarak A-C:

$$\text{sqrt}((10-2)^2 + (2-9)^2)=\text{sqrt}(64+49)=10,63$$

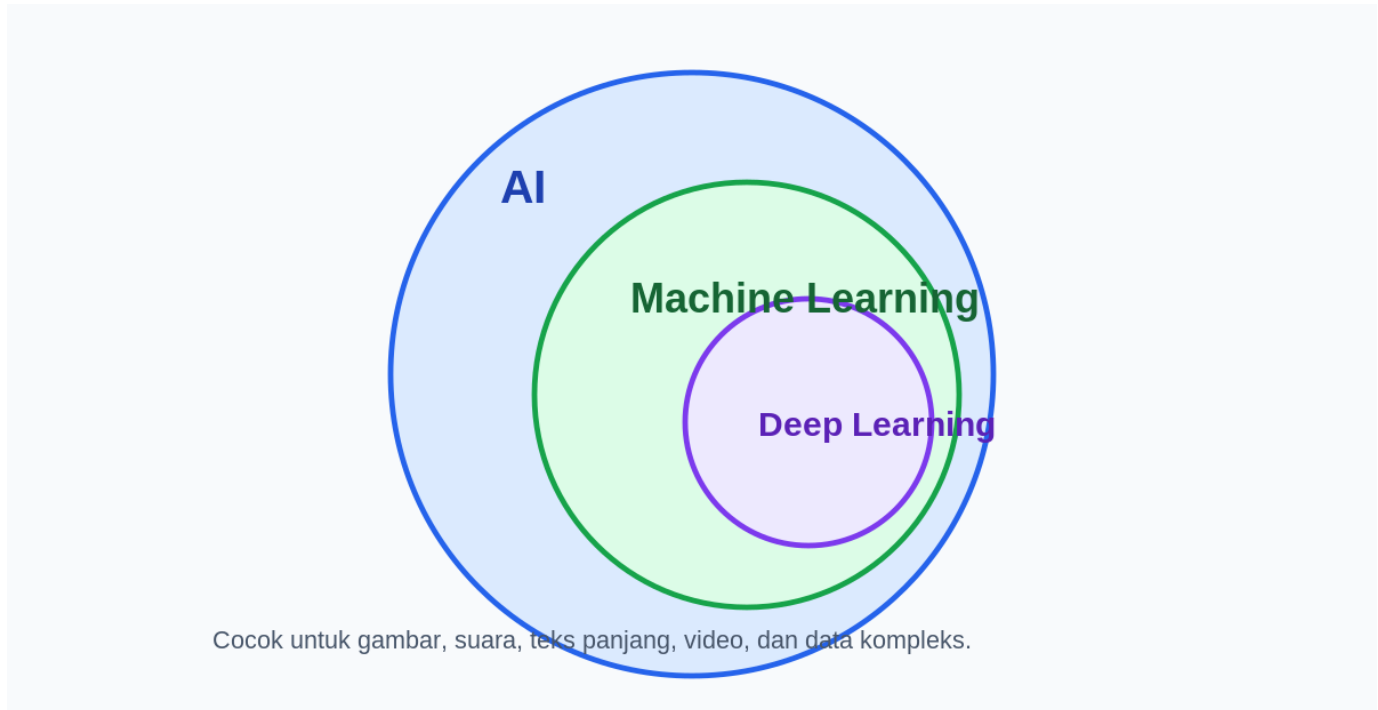
A lebih dekat ke B, sehingga mungkin satu cluster.

Tes cepat subbab 7

1. Mengapa unsupervised learning tidak memakai label?
2. Apa contoh clustering di bisnis lokal?
3. Mengapa hasil cluster perlu ditafsirkan oleh manusia?

Subbab 8 — Deep Learning: ketika data mentah terlalu kompleks

Deep Learning adalah Machine Learning berbasis neural network berlapis [R6]. Ia sangat berguna ketika input mentah kompleks: gambar, suara, teks panjang, video, atau kombinasi banyak modalitas. Deep learning belajar representasi, yaitu cara mengubah data mentah menjadi fitur internal yang berguna.



Posisi Deep Learning

Dalam ML klasik, manusia sering membuat fitur manual. Untuk prediksi harga rumah, kita bisa membuat fitur `luas`, `jumlah_kamar`, `jarak_ke_stasiun`. Untuk gambar, fitur manual jauh lebih sulit: tepi, tekstur, bentuk, komposisi, warna, objek. Deep learning membantu mempelajari fitur tersebut dari data.

Catatan hijau — kata kunci: Representasi

Deep learning kuat bukan karena “ajaib”, tetapi karena ia dapat belajar representasi bertingkat dari data kompleks.

Namun deep learning bukan jawaban untuk semua masalah. Jika data tabular kecil, model sederhana sering lebih cepat, murah, dan mudah dijelaskan. Prinsip profesional: mulai dari baseline sederhana, baru naik ke model lebih kompleks jika data, metrik, dan kebutuhan membenarkan.

Representasi berlapis

Satu layer neural network:

$$h = \text{activation}(Wx + b)$$

Banyak layer:

$$\begin{aligned} h_1 &= \text{activation}(W_1x + b_1) \\ h_2 &= \text{activation}(W_2h_1 + b_2) \\ \hat{y} &= \text{output}(W_3h_2 + b_3) \end{aligned}$$

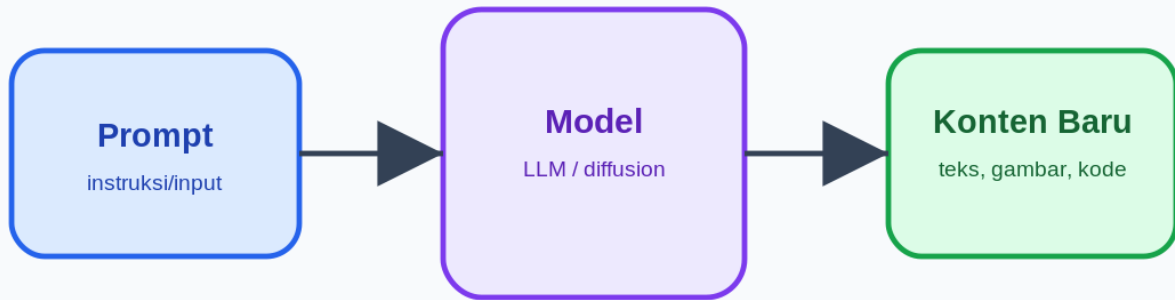
Deep learning kuat karena representasi dipelajari bertahap. Namun ia butuh data, komputasi, dan evaluasi yang hati-hati.

Tes cepat subbab 8

1. Kapan deep learning biasanya berguna?
2. Apa arti representasi dalam konteks deep learning?
3. Mengapa model sederhana tetap penting?

Subbab 9 — Generative AI dan LLM: pembuat konten, bukan oracle

Generative AI menghasilkan konten baru: teks, gambar, kode, audio, video, atau desain. Large Language Model (LLM) adalah model bahasa besar yang dilatih untuk memprediksi dan menghasilkan urutan teks. LLM dapat menjadi dasar chatbot, coding assistant, ringkasan dokumen, dan sistem tanya jawab [R7].



Perlu verifikasi: output bisa terdengar yakin tetapi salah.

Generative AI dalam peta AI

Generative AI berurusan dengan Deep Learning karena model generatif modern biasanya memakai neural network besar. Ia juga berurusan dengan NLP ketika input/output-nya bahasa. Tetapi generative AI bukan sinonim dari seluruh AI.

Catatan merah — kata kunci: Bukan Oracle

LLM dapat membantu berpikir dan menulis, tetapi tidak otomatis benar. Output penting harus dicek dengan sumber dan penalaran manusia.

Gunakan generative AI untuk:

- membuat draft awal,
- menjelaskan ulang konsep,
- membantu brainstorming,
- membuat template kode,
- merangkum dokumen yang sumbernya jelas.

Jangan gunakan tanpa validasi untuk:

- diagnosis medis final,
- keputusan hukum/keuangan penting,
- klaim ilmiah tanpa sumber,
- data pribadi sensitif,

- keputusan yang merugikan orang tanpa review.

LLM sebagai prediksi token

$$P(\text{token}_t \mid \text{token}_1, \dots, \text{token}_{\{t-1\}})$$

Model memilih token berikutnya berdasarkan distribusi. Karena itu, LLM bisa fasih tetapi salah. Kefasihan bukan bukti kebenaran.

Contoh probabilitas token

Konteks “Ibukota Indonesia adalah ...”:

Jakarta: 0,70
Nusantara: 0,20
Bandung: 0,02

Jawaban tergantung konteks waktu, data, dan instruksi. Untuk fakta penting, butuh verifikasi.

Tes cepat subbab 9

1. Apa bedanya Generative AI dan classification?
2. Mengapa LLM bukan oracle?
3. Berikan satu contoh penggunaan generative AI yang perlu validasi ketat.

Subbab 10 — Reinforcement Learning: peta agent, aksi, reward

Reinforcement Learning (RL) dipakai ketika masalah berbentuk keputusan berurutan. Agent melihat state, memilih action, mendapat reward, lalu belajar policy: strategi memilih aksi [R8].



Posisi Reinforcement Learning

Contoh game gridworld: agent mulai dari satu kotak, ingin mencapai tujuan, menghindari rintangan, dan memaksimalkan reward. Contoh bisnis: sistem simulasi mencoba strategi diskon beberapa minggu dan melihat dampaknya pada margin. Contoh operasi: agent menjadwalkan pengiriman agar hemat waktu.

Catatan biru — kata kunci: Keputusan Berurutan

RL cocok ketika aksi hari ini memengaruhi kondisi dan pilihan besok. Jika hanya memprediksi satu label dari satu input, biasanya bukan RL.

Perbandingan cepat:

Pertanyaan	Supervised ML	RL
Data utama	contoh + jawaban benar	state, action, reward
Output	prediksi label/angka	policy/strategi aksi
Tantangan	generalization	exploration vs exploitation
Risiko	bias data, leakage	reward hacking, aksi berbahaya

RL sangat menarik, tetapi tidak selalu praktis untuk pemula. Sering kali kita perlu simulasi yang aman sebelum mencoba di dunia nyata. Karena itu, Bab 12 memakai gridworld terlebih dahulu.

Return dengan discount factor

$$G_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots$$

Jika reward [3 , 4 , 5] dan $\gamma=0,9$:

$$G = 3 + 0,9 \times 4 + 0,81 \times 5 = 3 + 3,6 + 4,05 = 10,65$$

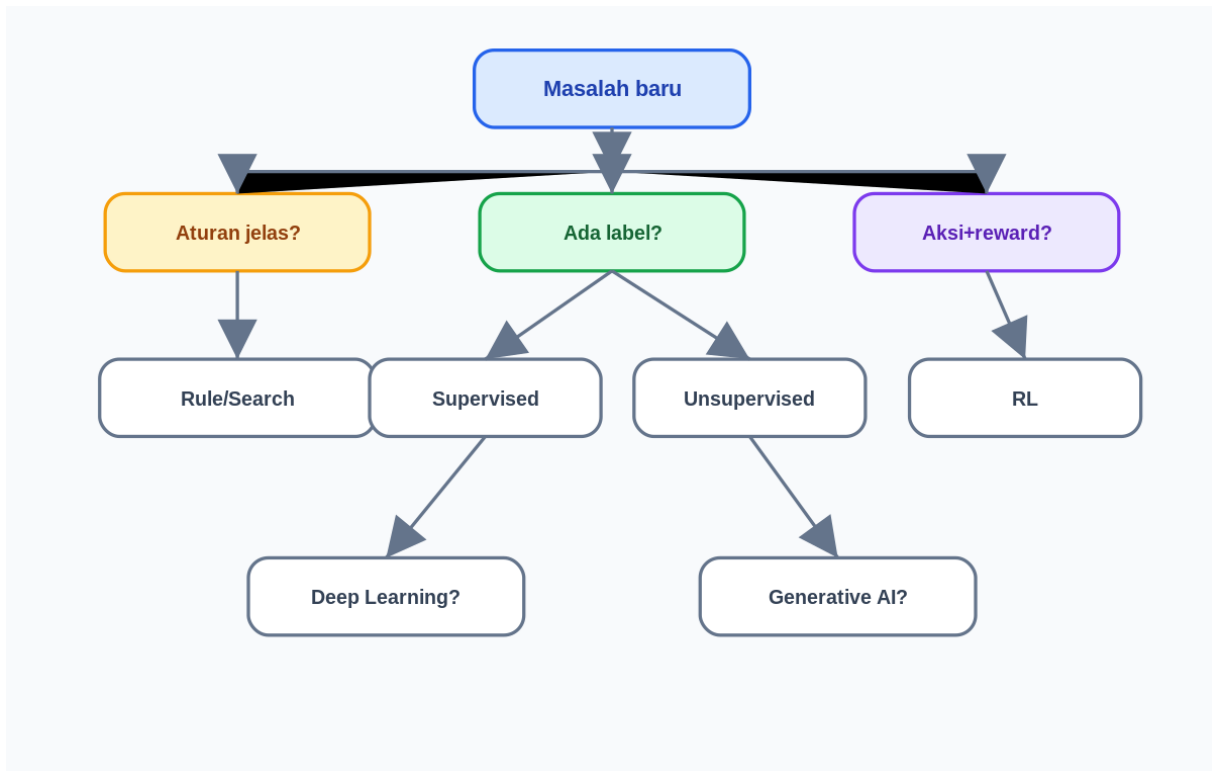
RL cocok jika keputusan berurutan dan aksi sekarang memengaruhi kondisi berikutnya.

Tes cepat subbab 10

1. Apa ciri utama masalah RL?
2. Apa itu policy dalam bahasa sederhana?
3. Mengapa RL sebaiknya diuji di simulasi dulu?

Subbab 11 — Cara memilih pendekatan: pohon keputusan sederhana

Sekarang kita gabungkan peta menjadi alat praktis. Saat bertemu masalah baru, jangan langsung bertanya “model apa yang keren?”. Tanyakan urutan berikut.



Pohon keputusan memilih pendekatan

Catatan kuning — kata kunci: Mulai Sederhana

Pendekatan terbaik bukan yang paling populer, tetapi yang paling cukup, terukur, aman, dan bisa dirawat.

Pohon keputusan praktis:

1. Apakah aturannya jelas dan stabil? Jika ya, mulai dari rule-based.
2. Apakah tujuannya memahami data? Jika ya, mulai dari Data Science/EDA.
3. Apakah ada label/jawaban benar? Jika ya, supervised learning.
4. Apakah tidak ada label tetapi ingin mencari struktur? Jika ya, unsupervised learning.
5. Apakah input kompleks seperti gambar/suara/teks panjang? Pertimbangkan deep learning.
6. Apakah output harus berupa konten baru? Pertimbangkan generative AI.
7. Apakah masalahnya aksi berurutan dan reward? Pertimbangkan RL.
8. Apakah dampaknya tinggi? Tambahkan validasi, audit, dan review manusia.

Checklist pemilihan berbasis data

- Ada label y ? → supervised
- Tidak ada label, ingin pola? → unsupervised
- Ada aksi berurutan dan reward? → RL
- Butuh membuat konten? → generative model
- Butuh aturan transparan dan domain jelas? → symbolic/rule-based

Contoh terstruktur

Kasus: “menentukan pelanggan prioritas promo”. Jika punya histori pelanggan menerima promo dan membeli/tidak, gunakan supervised learning. Jika belum punya label, mulai dari segmentasi unsupervised. Jika promo diberikan berulang dan reward penjualan datang setelah aksi, pertimbangkan bandit/RL.

Tes cepat subbab 11

1. Mengapa kita sebaiknya mulai dari solusi sederhana?
2. Jika ada label “lulus/tidak lulus”, pendekatan awal apa yang cocok?
3. Jika masalahnya membuat teks promosi baru, keluarga apa yang relevan?

Subbab 12 — Praktikum: klasifikasi 30 kasus AI

Tugas akhir Bab 2 adalah mengklasifikasikan 30 kasus. Ini melatih peta mental sebelum kamu masuk kode dan matematika.

Template Klasifikasi 30 Kasus

No	Kasus	Label	Alasan	Keyakinan
1		supervised?		tinggi/sedang/rendah
2				
3				
...				

Template klasifikasi kasus AI

Gunakan label berikut:

- Rule-based/search/planning
- Data Science/analytics
- Supervised ML — regression
- Supervised ML — classification
- Unsupervised learning
- Deep Learning
- Generative AI/LLM
- Reinforcement Learning
- Tidak cukup informasi

Contoh awal:

Kasus	Label awal	Alasan
Menentukan diskon jika total belanja di atas Rp200.000	Rule-based	Aturan jelas.
Mengelompokkan pelanggan dari pola transaksi	Unsupervised	Tidak ada label, mencari segmen.
Membuat caption promosi otomatis	Generative AI/LLM	Output konten baru.
Agent game belajar menghindari banjir	RL	Aksi berurutan dan reward.

Catatan hijau — kata kunci: Alasan Lebih Penting dari Label

Dalam praktikum ini, jawaban bisa diperdebatkan. Yang dinilai bukan hanya label, tetapi alasan dan asumsi yang kamu tulis.

Tes cepat subbab 12

1. Buat diagram kecil hubungan AI, ML, DL, Generative AI, Data Science, dan RL.
2. Klasifikasikan 10 kasus pertama dari praktikum tanpa melihat kunci jawaban.
3. Pilih satu kasus yang ambigu dan jelaskan dua kemungkinan labelnya.
4. Jelaskan mengapa Data Science tidak selalu AI.
5. Jelaskan kapan RL lebih cocok daripada supervised learning.

Latihan hitung terstruktur Bab 2

1. Tentukan kategori pendekatan untuk 10 kasus: prediksi stok, segmentasi pelanggan, chatbot, sistem pakar aturan pajak, robot belajar navigasi, deteksi spam, rekomendasi produk, ringkasan dokumen, clustering berita, dan bandit promo.
2. Hitung MAE baseline mean untuk target $[30, 40, 50]$.
3. Hitung accuracy dan recall dari aktual $[1, 0, 1, 1]$, prediksi $[1, 0, 0, 1]$.
4. Hitung jarak pelanggan A $[10, 2]$ ke B $[12, 3]$ dan C $[2, 9]$.
5. Hitung return RL untuk reward $[3, 4, 5]$ dan $\gamma=0,9$.
6. Buat lifecycle mini untuk satu ide AI lokal: problem, data, label/reward, baseline, metrik, risiko, monitoring.

Ringkasan Bab 2

- AI adalah payung besar; ML adalah salah satu keluarga di bawahnya.
- Data Science beririsan dengan AI, tetapi tidak semua Data Science adalah AI.
- Machine Learning belajar pola dari data untuk generalization.
- Supervised learning memakai input dan jawaban benar.
- Regression memprediksi angka; classification memprediksi kategori.
- Unsupervised learning mencari struktur tanpa label.
- Deep Learning berguna untuk data kompleks dan belajar representasi bertingkat.
- Generative AI membuat konten baru, tetapi bukan oracle.
- Reinforcement Learning belajar strategi aksi dari reward.
- Pemilihan pendekatan harus dimulai dari masalah, data, risiko, dan metrik — bukan dari hype.

Referensi utama bab

[R1] OECD AI Principles dan updated AI system definition. Digunakan untuk posisi AI sebagai sistem berbasis tujuan, input, output, dan dampak. [R2] Russell, S., & Norvig, P. *Artificial Intelligence: A

Modern Approach*. Digunakan untuk peta AI, agent, search, dan planning. [R3] National Academies / NIST / literatur Data Science umum. Digunakan untuk konsep Data Science sebagai proses data-ke-keputusan. [R4] Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*. Digunakan sebagai rujukan pipeline ML dan keluarga supervised/unsupervised; tidak disalin. [R5] Hastie, Tibshirani, & Friedman. *The Elements of Statistical Learning*. Digunakan sebagai rujukan konseptual supervised/unsupervised dan generalization; tidak disalin. [R6] Goodfellow, Bengio, & Courville. *Deep Learning*. Digunakan untuk konsep deep learning dan representasi; tidak disalin. [R7] Stanford/IBM/OpenAI/Google public docs tentang generative AI dan LLM. Digunakan untuk konteks generative AI modern; klaim final perlu dicek ulang saat publikasi. [R8] Sutton, R. S., & Barto, A. G. *Reinforcement Learning: An Introduction*. Digunakan untuk konsep state, action, reward, policy, dan RL; tidak disalin.

Catatan validasi internal v0.3

Aspek	Status	Catatan
Hubungan AI/ML/DL/RL	Tervalidasi awal	Dijelaskan sebagai peta konseptual, bukan batas matematis kaku.
Data Science	Tervalidasi awal	Diposisikan sebagai disiplin data-ke-keputusan yang beririsan dengan AI.
Supervised/unsupervised/RL	Tervalidasi awal	Definisi disederhanakan tetapi selaras dengan literatur ML/RL.
Generative AI	Perlu update berkala	Bidang cepat berubah; referensi final perlu diperbarui sebelum v1.0.
Gambar	Orisinal	SVG dibuat sendiri untuk bab ini.
Praktikum	Siap v0.3	30 kasus + script klasifikasi sederhana.