

CPMK 07.1 · SEMESTER 5 & 7

Supervised, Unsupervised & Semi-supervised

PERTEMUAN

03

IF261220 3 SKS · Program Studi Informatika, Fakultas Ilmu Komputer

SEBELUM MULAI

Recap Pertemuan 2

Tiga hal yang sudah kita bangun minggu lalu, sebagai pijakan hari ini.

Kutukan Dimensi

Makin banyak fitur, makin jarang data mengisi ruangnya, model jadi mudah overfitting.

Selection vs Extraction

Selection memilih subset fitur asli; extraction mentransformasi menjadi fitur baru.

PCA & LDA

PCA mencari arah variasi terbesar; LDA mencari arah pemisah antar-kelas terbaik.

AGENDA PERTEMUAN 3

01

Recap singkat tiga paradigma

02

Kelebihan & kekurangan tiap paradigma

03

Visualisasi & tabel perbandingan data

04

Empat kategori studi kasus unsupervised

05

Tiga studi kasus mendalam

06

Memilih paradigma & kaitan final project

Mengingat Kembali Tiga Paradigma



Supervised

Belajar dari data berlabel, masukan dan keluaran sudah diketahui.



Unsupervised

Belajar dari data tanpa label, mencari struktur tersembunyi.



Semi-supervised

Menggabungkan sedikit data berlabel dengan banyak data tanpa label.

Pertanyaan hari ini bukan lagi “apa itu?”, melainkan “kapan masing-masing layak dipakai, dan apa yang harus dikorbankan?”

Setiap paradigma punya trade-off tersendiri antara kebutuhan data, biaya persiapan, dan kemudahan evaluasi.

Supervised Learning

Kelebihan

Belajar langsung dari contoh berlabel untuk memprediksi label baru.

- Akurasi tinggi bila data label berkualitas dan cukup banyak.
- Evaluasi objektif: akurasi, precision, recall, F1-score.
- Tersedia banyak algoritma yang matang dan teruji.
- Hasil mudah dijelaskan ke pemangku kepentingan non-teknis.

Contoh: mengklasifikasikan email sebagai spam atau bukan spam.

Kekurangan

Bergantung penuh pada ketersediaan dan kualitas label.

- Anotasi data berlabel mahal dan memakan waktu.
- Rentan bias bila label tidak merepresentasikan populasi.
- Sulit menangkap pola yang tidak tercakup label yang ada.
- Performa dapat menurun tajam saat distribusi data berubah.

Contoh: model gagal mengenali pola spam baru yang belum pernah dilabeli.

Unsupervised Learning

Kelebihan

Menemukan struktur tersembunyi tanpa memerlukan data berlabel.

- Tidak memerlukan proses anotasi yang mahal.
- Mampu mengungkap pola atau kelompok yang tak terduga.
- Skalabel untuk data dalam jumlah sangat besar.
- Cocok sebagai langkah eksplorasi sebelum pemodelan lanjutan.

Contoh: mengelompokkan pelanggan berdasarkan pola belanja.

Kekurangan

Hasil lebih sulit diukur dan dimaknai secara objektif.

- Tidak ada metrik evaluasi tunggal yang disepakati bersama.
- Interpretasi kluster membutuhkan keahlian domain.
- Sensitif terhadap parameter, misalnya jumlah kluster.
- Hasil bisa berbeda signifikan antar-inisialisasi algoritma.

Contoh: dua analis bisa menafsirkan jumlah kluster berbeda dari data yang sama.

Semi-Supervised Learning

Menggabungkan sedikit data berlabel dengan data tanpa label yang melimpah.

Banyak persoalan nyata punya sedikit data berlabel namun memiliki banyak data mentah. Semi-supervised learning memanfaatkan keduanya.

KELEBIHAN

- Mengurangi kebutuhan pelabelan yang mahal.
- Sering menghasilkan akurasi lebih baik dibanding unsupervised murni.
- Efektif saat label langka tapi data mentah melimpah.

KEKURANGAN

- Performa bergantung pada asumsi struktur data yang dipakai.
- Label yang salah dapat menyesatkan seluruh proses pelatihan.
- Implementasi lebih kompleks (self-training, co-training).

Contoh: mengklasifikasikan ratusan ribu artikel berita padahal hanya ratusan yang diberi label kategori.

Bagaimana Data Terlihat pada Tiap Paradigma



Tabel Perbandingan Tiga Paradigma

Aspek	Supervised	Unsupervised	Semi-supervised
Kebutuhan Label	Wajib, seluruh data	Tidak diperlukan	Sebagian kecil data
Biaya Persiapan Data	Tinggi (anotasi manual)	Rendah	Sedang
Tujuan Utama	Memprediksi label baru	Menemukan struktur/pola	Prediksi + manfaatkan data tanpa label
Evaluasi Hasil	Jelas: akurasi, F1-score	Sulit, sering subjektif	Kombinasi keduanya
Contoh Algoritma	Decision Tree, Neural Network	K-Means, SOM	Self-training, Label Propagation

Studi Kasus Unsupervised Learning

Bagaimana pola tersembunyi dalam data tanpa label dimanfaatkan untuk memecahkan persoalan nyata.

02

Empat Kategori Studi Kasus Unsupervised Learning

01

Clustering

Mengelompokkan data berdasarkan kemiripan tanpa label yang ditentukan.

K-Means, Hierarchical, DBSCAN

02

Deteksi Anomali

Menemukan data yang menyimpang jauh dari pola umum mayoritas.

Isolation Forest, One-Class SVM

03

Asosiasi

Menemukan pola kombinasi kemunculan bersama antar-item.

Apriori, FP-Growth

04

Reduksi Dimensi & Discovery

Menyederhanakan data untuk menemukan struktur tersembunyi.

PCA, t-SNE, Topic Modeling

Kategori keempat telah dibahas pada Pertemuan 2, kini dilihat sebagai alat penemuan struktur.

Segmentasi Pelanggan

K-Means mengelompokkan pelanggan tanpa kategori yang ditentukan sebelumnya.

Sebuah bisnis ritel ingin menyesuaikan strategi pemasaran untuk kelompok pelanggan berbeda, tanpa kategori yang sudah ditentukan sebelumnya.

PENDEKATAN

Data

Frekuensi belanja, nilai transaksi rata-rata, kategori favorit.

Standardisasi

Samakan skala fitur, tentukan k lewat elbow method.

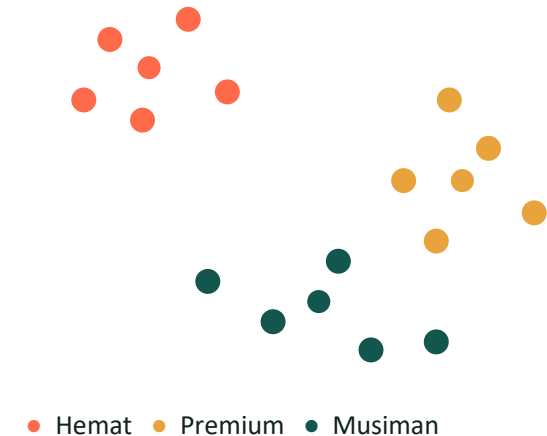
K-Means

Tiap pelanggan otomatis masuk ke kluster terdekat.

Interpretasi

Tiap kluster diberi nama sesuai karakteristiknya.

HASIL KLASLER



Hasil: ditemukan tiga kluster: Hemat, Premium, dan Musiman, yang masing-masing mendapat strategi pemasaran berbeda.

Keamanan & Deteksi Fraud

Model belajar pola 'normal', lalu menandai titik yang menyimpang jauh.

Institusi keuangan ingin mendeteksi transaksi mencurigakan tanpa memiliki cukup contoh penipuan berlabel untuk dilatih.

PENDEKATAN

Data

Pola transaksi: waktu, jumlah, lokasi, frekuensi akses.

Pelatihan

Model mempelajari bentuk 'normal' dari mayoritas data.

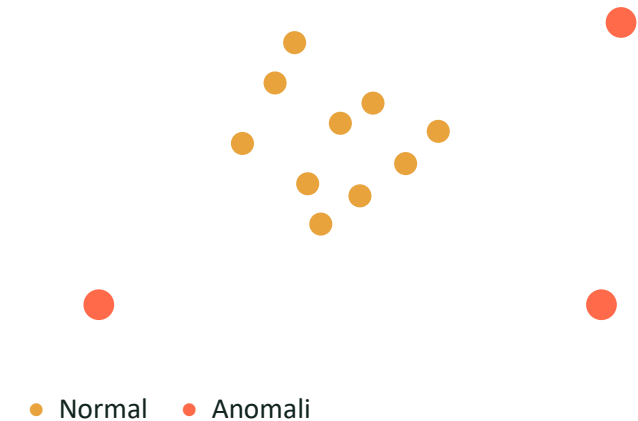
Penilaian Jarak

Titik yang jauh dari pola normal diberi skor anomali tinggi.

Tinjauan Manual

Titik berskor tinggi ditandai untuk diperiksa manusia.

TITIK TERTANDAI



Hasil: kurang dari 1% transaksi ditandai untuk dicek manual, membantu tim keamanan fokus pada kasus paling berisiko.

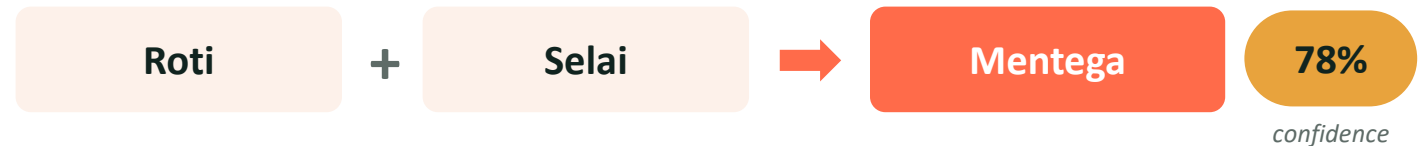
Market Basket Analysis

Apriori menemukan pola kombinasi produk yang sering dibeli bersamaan.

Peritel ingin memahami kombinasi produk yang sering dibeli bersamaan, untuk penataan rak dan strategi promosi bundling.

PENDEKATAN

- Data: riwayat transaksi kasir (isi tiap keranjang belanja).
- Algoritma Apriori/FP-Growth menghitung pola kemunculan bersama.
- Diukur dengan support, confidence, dan lift.



Jika membeli Roti dan Selai, kemungkinan besar juga membeli Mentega.

Hasil: aturan seperti ini dipakai untuk menata rak berdekatan dan merancang promo bundling produk.

Kategori, Algoritma, dan Contoh Kasus

Kategori	Algoritma Umum	Contoh Kasus
Clustering	K-Means, Hierarchical, DBSCAN	Segmentasi pelanggan, kompresi citra
Deteksi Anomali	Isolation Forest, One-Class SVM	Deteksi fraud, keamanan jaringan
Asosiasi	Apriori, FP-Growth	Market basket analysis, rekomendasi produk
Reduksi Dimensi & Discovery	PCA, t-SNE, Topic Modeling	Visualisasi data, pengelompokan dokumen

Memilih Paradigma yang Tepat

01	Apakah Anda memiliki data berlabel?	Banyak → supervised. Sedikit → semi-supervised. Tidak ada → unsupervised.
02	Apa tujuan utama analisis Anda?	Memprediksi label baru → supervised. Menemukan struktur tersembunyi → unsupervised.
03	Seberapa mahal proses pelabelan data?	Mahal dan lambat → pertimbangkan semi-supervised atau unsupervised terlebih dahulu.
04	Apakah performa perlu diukur secara objektif?	Ya, dengan metrik baku → supervised umumnya paling mudah dievaluasi.

Dalam praktik, banyak proyek dimulai dari eksplorasi unsupervised sebelum memutuskan pendekatan akhir.

PRAKTIK

Memetakan Data Anda ke Paradigma

Situasi data yang dimiliki menentukan paradigma sekaligus jenis studi kasus yang relevan.

Punya dataset berlabel lengkap

Pendekatan supervised dapat langsung diterapkan; fokus pada pemilihan algoritma dan evaluasi.

Punya data tapi tanpa label

Coba clustering atau deteksi anomali untuk eksplorasi awal, sesuai kategori di Bagian 2.

Ingin menemukan pola kombinasi

Analisis asosiasi (market basket) relevan jika data berupa transaksi atau kejadian bersama.

Belum yakin bentuk datanya

Mulai dengan eksplorasi unsupervised sebelum menentukan pendekatan akhir.

Paradigma bukan pilihan permanen, banyak proyek berpindah paradigma seiring data yang terkumpul bertambah.

Rangkuman dan Persiapan

SUDAH KITA PAHAMI

- Kelebihan dan kekurangan tiap paradigma pembelajaran
- Empat kategori studi kasus unsupervised learning
- Tiga studi kasus mendalam: clustering, anomali, asosiasi

PERBEDAAN MENDASAR

- Supervised: perlu label penuh, evaluasi paling jelas
- Unsupervised: tanpa label, hasil perlu ditafsirkan
- Semi-supervised: jalan tengah paling realistis di industri

PERSIAPAN PERTEMUAN BERIKUTNYA

- Latihan Soal (10%): program Python sederhana + review jurnal
- Pastikan lingkungan Python sudah siap sebelum sesi
- Pilih satu jurnal ML terkait topik final project Anda