

CPMK 09.1

Unsupervised Learning: Kegunaan & Jenis Algoritma

PERTEMUAN

06

IF261220 3 SKS · Program Studi Informatika, Fakultas Ilmu Komputer

UPN "Veteran" Jawa Timur

SEBELUM MULAI

Recap Pertemuan 5

Dua hal yang sudah kita bangun pada pertemuan lalu.

Algoritma Supervised & Semi-supervised

Decision Tree (5 tingkat) dan Self-training — lengkap dengan studi kasus dan latihan.

Studi Kasus & Latihan Soal

Klasifikasi, regresi, semi-supervised, serta pengantar Latihan Soal Minggu 5 (bobot 10%).

AGENDA PERTEMUAN 6

- | | | | |
|-----------|---|-----------|---------------------------------------|
| 01 | Kenapa Unsupervised Learning? | 02 | Tiga kategori algoritma unsupervised |
| 03 | K-Means: cara kerja & ilustrasi | 04 | Menentukan k & cuplikan kode Python |
| 05 | Hierarchical, DBSCAN, Apriori, Isolation Forest | 06 | Diskusi kelas & persiapan Pertemuan 7 |

Kenapa Unsupervised Learning Penting?

Di dunia nyata, data berlabel itu mahal dan langka — di sinilah unsupervised learning berperan.



Data Tanpa Label Jauh Lebih Banyak

Sebagian besar data di dunia nyata (log transaksi, teks, gambar) tidak punya label siap pakai.



Eksplorasi Sebelum Supervised

Sering dipakai sebagai langkah awal untuk memahami data sebelum melabeli atau membangun model supervised.



Menemukan Pola Tersembunyi

Model bisa menemukan struktur atau kelompok yang bahkan tidak terpikirkan sebelumnya oleh manusia.



Efisiensi Biaya

Mengurangi kebutuhan tenaga dan waktu untuk memberi label pada jutaan baris data secara manual.

Tiga Kategori Algoritma Unsupervised

Reduksi dimensi (PCA/LDA) sudah dibahas di Pertemuan 2 — fokus hari ini pada tiga kategori berikut.

Clustering

Mengelompokkan data yang mirip tanpa label sebelumnya.

ALGORITMA UMUM

K-Means, Hierarchical, DBSCAN

Association Rule

Menemukan aturan asosiasi antar-item dalam data transaksi.

ALGORITMA UMUM

Apriori, FP-Growth

Anomaly Detection

Mendeteksi data yang menyimpang jauh dari pola umum.

ALGORITMA UMUM

Isolation Forest, One-Class SVM

Fokus hari ini: Clustering, dengan K-Means sebagai algoritma yang paling mendasar untuk dipahami lebih dulu.

K-Means

Algoritma clustering paling populer — sederhana, cepat, dan mudah divisualisasikan.

K-Means mengelompokkan data ke dalam k kelompok dengan cara mengulang dua langkah utama hingga posisi centroid stabil.

CARA KERJA

1. Inisialisasi

Pilih k titik acak sebagai centroid awal (mis. $k=3$).

2. Assignment

Setiap titik data digabungkan ke centroid terdekat.

3. Update Centroid

Posisi tiap centroid dipindah ke rata-rata anggotanya.

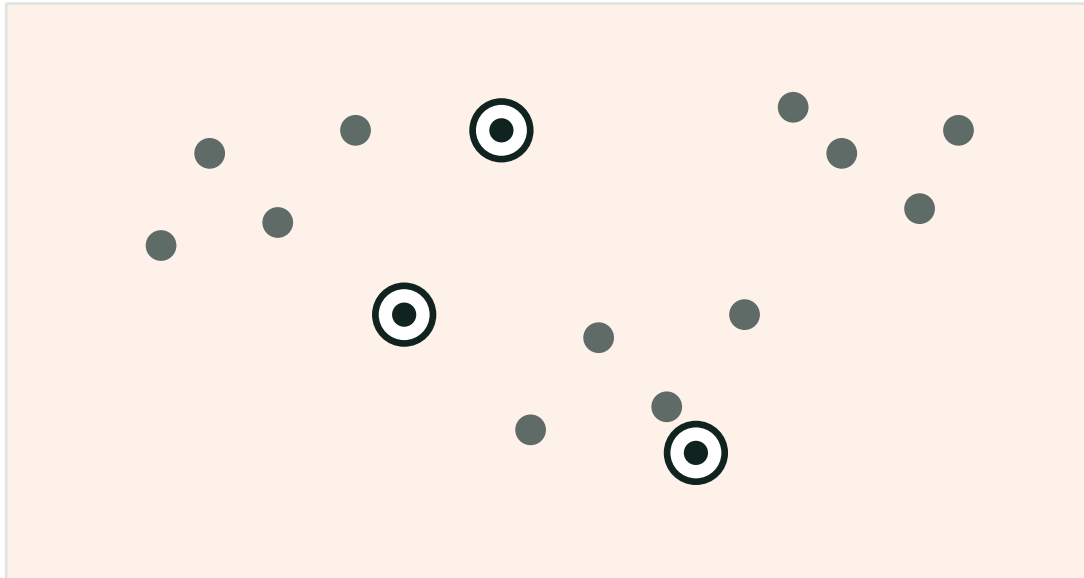
4. Ulangi

Ulangi langkah 2–3 sampai posisi centroid tidak berubah lagi.

K-Means: Sebelum dan Sesudah Iterasi

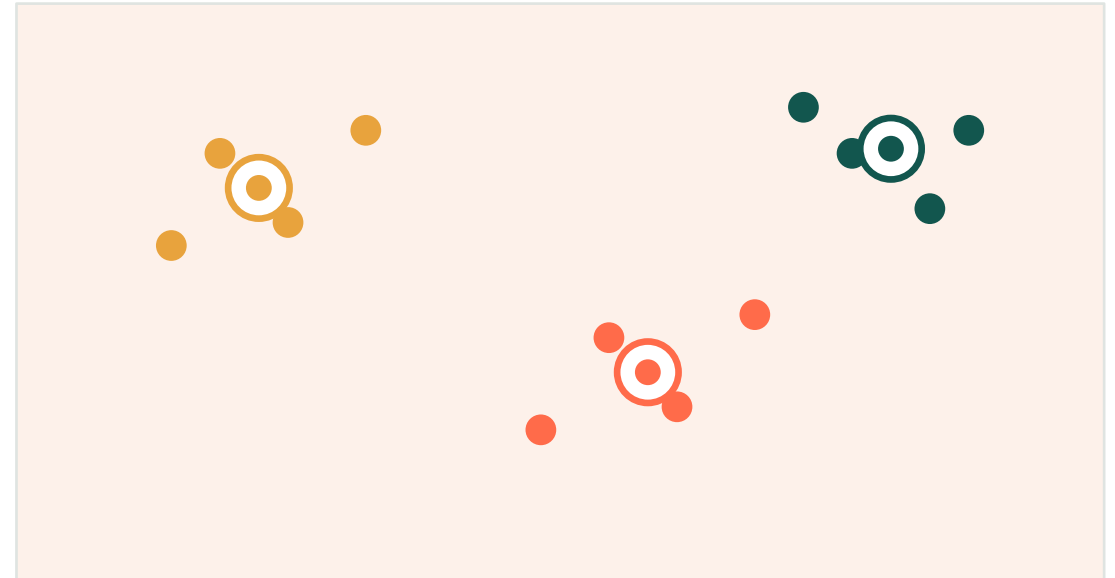
Titik data yang sama, dikelompokkan ulang setelah centroid berpindah beberapa kali.

ITERASI 1 · CENTROID ACAK



● Data (belum dikelompokkan) ◎ Centroid awal (acak)

SETELAH KONVERGEN



● Klaster 1 ● Klaster 2 ● Klaster 3 ◎ Centroid akhir

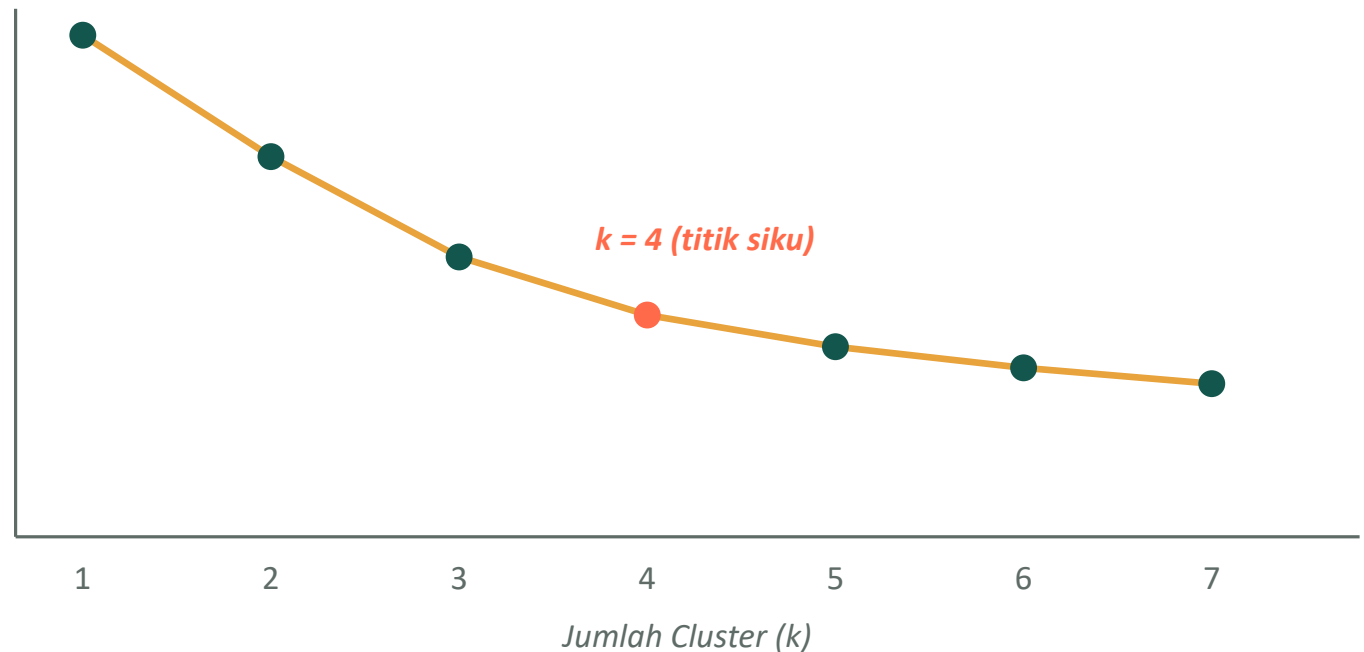
Menentukan Nilai k

Salah satu tantangan terbesar K-Means: berapa banyak cluster yang tepat?

Metode Elbow membantu memilih k dengan melihat titik di mana penurunan inertia (jarak dalam cluster) mulai melambat.

Inertia (Within-Cluster Sum of Squares)

GRAFIK ELBOW: INERTIA vs JUMLAH CLUSTER (k)



Intuisi: setelah k=4, menambah cluster tidak banyak mengurangi inertia — pertanda 4 adalah jumlah yang cukup.

Cuplikan Kode: K-Means dengan Python

CPMK 09.1 mulai bergeser ke praktik — begini bentuknya menggunakan library scikit-learn.

```
from sklearn.cluster import KMeans

# 1. Siapkan model dengan k=3 cluster
model = KMeans(n_clusters=3, random_state=42)

# 2. Latih model (assignment + update berulang)
model.fit(data)

# 3. Ambil hasil clustering
labels = model.labels_
centroids = model.cluster_centers_
```

Poin penting: kelima baris kode ini merepresentasikan seluruh langkah K-Means yang dijelaskan sebelumnya — library menangani iterasinya secara otomatis.

Dua Algoritma Clustering Lainnya

K-Means bukan satu-satunya pilihan — dua alternatif ini populer untuk kasus tertentu.

Algoritma	Cara Kerja Singkat	Kapan Dipakai
Hierarchical Clustering	Menyusun data dalam struktur pohon (dendrogram) dari yang paling mirip hingga membentuk satu klaster besar.	Saat jumlah cluster belum diketahui & data tidak terlalu besar.
DBSCAN	Mengelompokkan berdasarkan kepadatan titik; titik yang terisolasi dianggap noise, bukan dipaksa masuk cluster.	Saat bentuk cluster tidak bulat atau ada banyak outlier.
Catatan: berbeda dari K-Means, DBSCAN tidak memerlukan penentuan k di awal — jumlah cluster ditemukan sendiri oleh algoritma.		

Apriori

Algoritma klasik di balik rekomendasi
"pelanggan yang membeli ini juga
membeli itu."

Mengingat kembali studi kasus Market Basket Analysis di Pertemuan 3 — begini algoritma di baliknya bekerja.

CARA KERJA

1. Hitung Frekuensi Item

Cari item atau kombinasi item yang sering muncul bersamaan dalam transaksi.

2. Bentuk Aturan

Susun aturan "jika beli A, maka beli B" dari kombinasi yang sering muncul.

3. Ukur Confidence

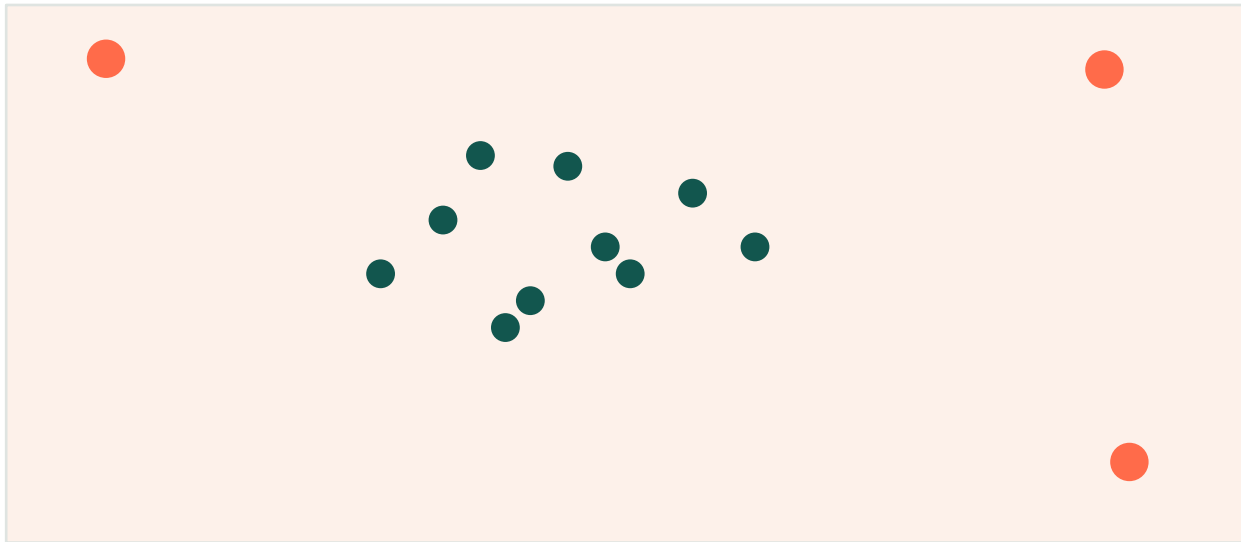
Hitung seberapa sering aturan itu benar-benar terjadi (mis. 78% dari pembeli roti juga beli selai).

Contoh: "Roti + Selai → Mentega" dengan confidence 78% — artinya 78% pembeli roti dan selai juga membeli mentega.

Anomaly Detection: Isolation Forest

Dari studi kasus deteksi fraud di Pertemuan 3 — begini cara algoritmanya menemukan data yang "aneh."

ILUSTRASI: DATA NORMAL vs ANOMALI



● Data normal ● Anomali (terisolasi)

CARA KERJA

- Data dipecah acak berulang kali menggunakan garis pemisah.
- Titik anomali lebih cepat "terisolasi" sendirian karena letaknya jauh dari kerumunan.
- Titik normal butuh lebih banyak pemisahan karena berada di tengah kerumunan data.

Algoritma, Kegunaan, dan Kapan Dipakai

Kategori	Algoritma	Kapan Dipakai
Clustering	K-Means, Hierarchical, DBSCAN	Segmentasi pelanggan, pengelompokan dokumen
Association Rule	Apriori, FP-Growth	Rekomendasi produk, analisis keranjang belanja
Anomaly Detection	Isolation Forest, One-Class SVM	Deteksi fraud, pemantauan keamanan jaringan

TANPA BOBOT NILAI · DISKUSI KELAS

Mari Berdiskusi

01

Di aplikasi atau platform apa saja kalian pernah melihat hasil clustering bekerja diam-diam di belakang layar?

02

Menurut kalian, kapan lebih baik memakai K-Means dibanding DBSCAN, atau sebaliknya?

03

Data apa di kampus (nilai, kehadiran, aktivitas UKM) yang menarik dianalisis dengan clustering?

Rangkuman dan Persiapan

SUDAH KITA PAHAMI

- Kegunaan unsupervised learning di dunia nyata
- Tiga kategori algoritma & cara kerja K-Means
- Cuplikan kode Python dengan scikit-learn

ALGORITMA PELENGKAP

- Hierarchical Clustering & DBSCAN
- Apriori untuk association rule
- Isolation Forest untuk anomaly detection

PERSIAPAN PERTEMUAN 07

- Penerapan algoritma unsupervised pada dataset nyata
- Praktik langsung menggunakan Python
- Siapkan laptop dengan Python & scikit-learn terpasang