



# CLUSTERING

Nur Fitriyah Ayu M.Cs.



# Pokok Bahasan

## APA ISI MATERI INI

Pengertian *Clustering*

Implementasi *Clustering*

Kategori Algoritma *Clustering*

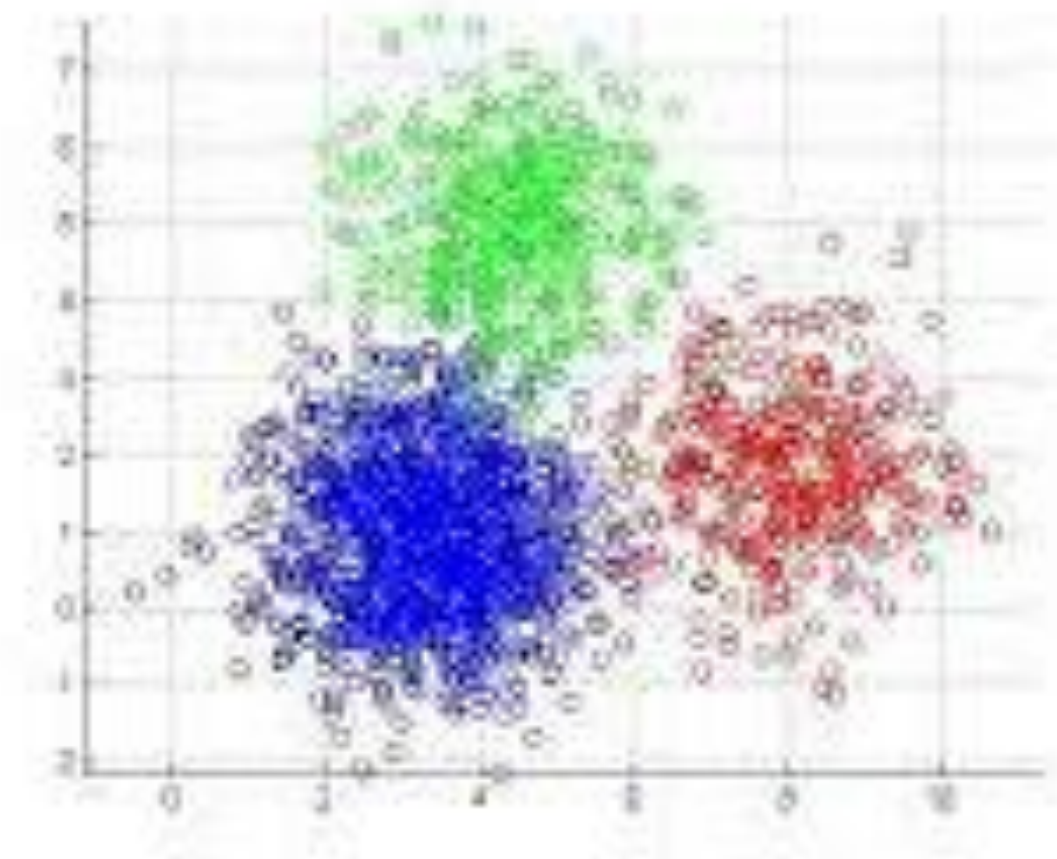
Algoritma *Clustering*

*Distance Space* (Rumus Jarak)

Studi Kasus

# Apa itu *Clustering*

- Proses mengelompokkan data (pola, entitas, kejadian, unit, hasil observasi) menjadi beberapa kategori/*cluster* berdasarkan ciri-ciri atau kesamaan tertentu.
- Menemukan *cluster* yang berkualitas -> **memaksimalkan kesamaan antar anggota satu *cluster* dan meminimalkan kesamaan antar *cluster*.**
- Tidak memiliki label kelas di awal (*unsupervised learning*).



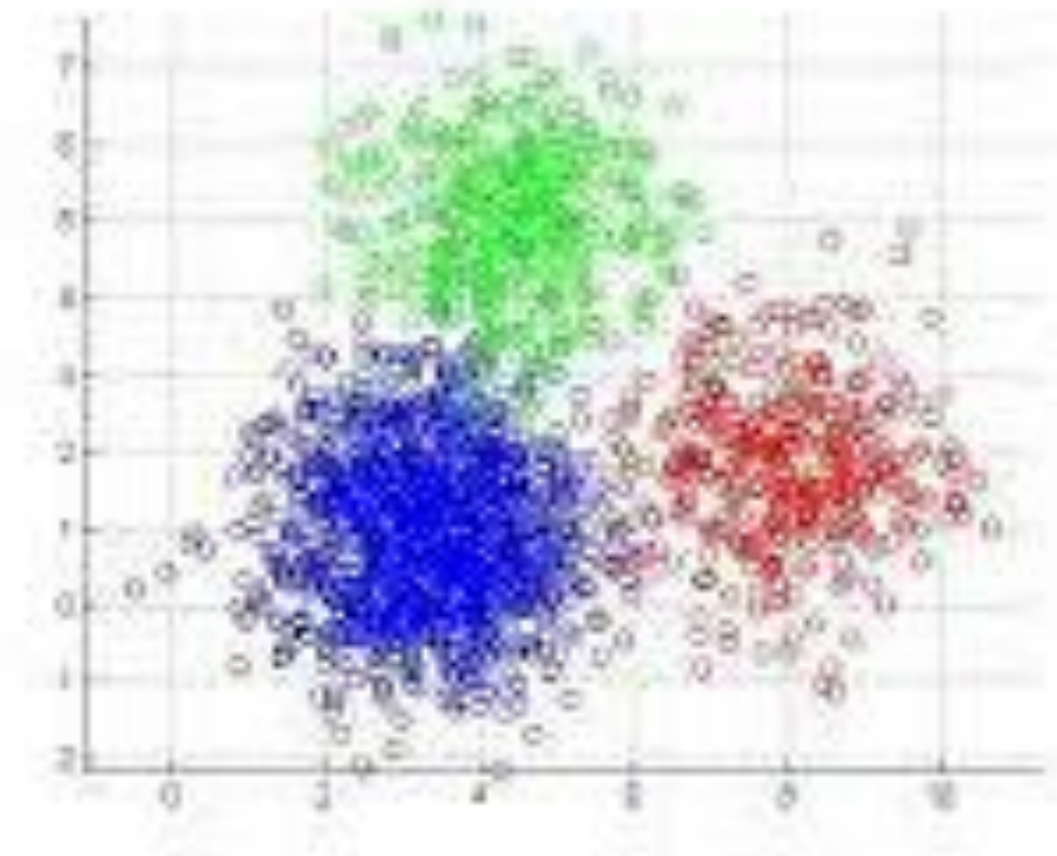


# Implementasi *Clustering*

1. Segmentasi wilayah (kasus covid)
2. Memahami perilaku pembeli
3. Mengetahui peluang produk baru
4. Segmentasi pelanggan

# Kategori Algoritma *Clustering*

- **Metode Partisi**
  - ✓ Menentukan n partisi kemudian setiap data diuji untuk dimasukkan pada salah satu partisi.
  - ✓ Satu data hanya masuk ke dalam 1 partisi.
  - ✓ Contoh : *K-Means*, *K-Medoids* dan DBSCAN



# Kategori Algoritma *Clustering*

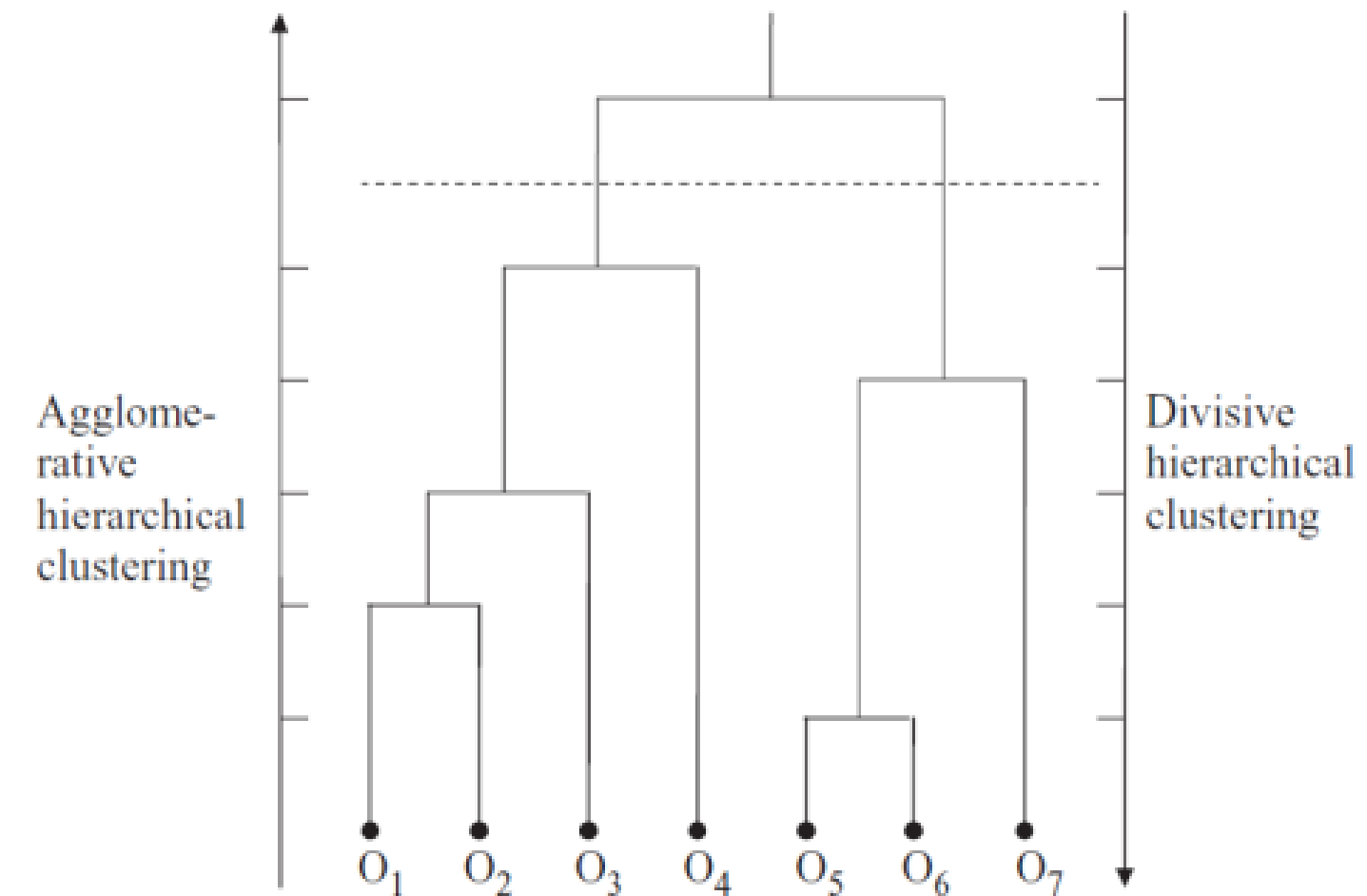
- **Metode Hierarki**

- ✓ Suatu data dapat memiliki *cluster* lebih dari satu (bersarang).

- ✓ *Bottom-up* dan *Top-down*.

memecah cluster besar menjadi cluster yang lebih kecil

menggabungkan *cluster* yang lebih kecil menjadi *cluster* besar

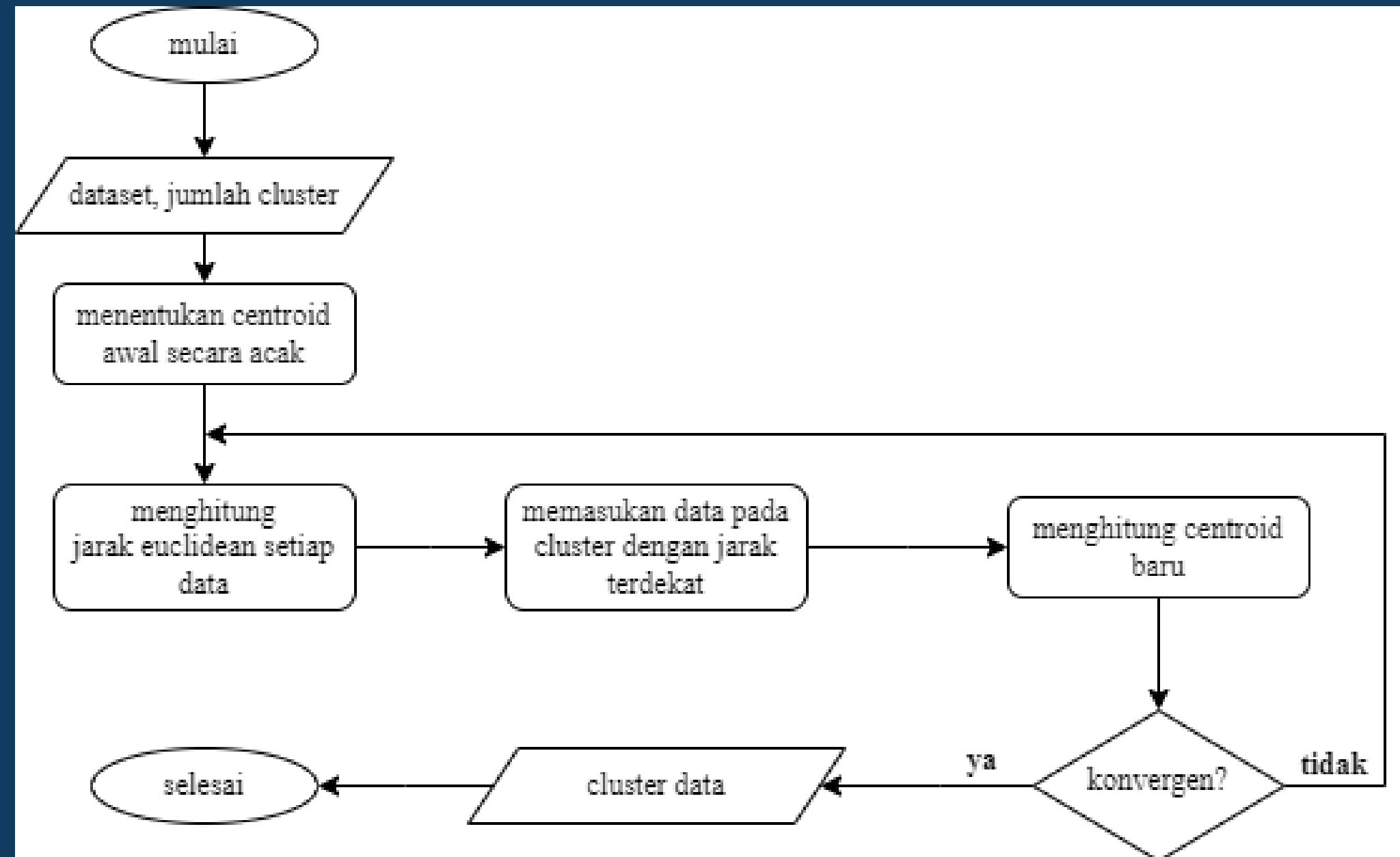


# Algoritma *K-Means*

- Termasuk di dalam algoritma dalam teknik non hierarki (partisi).
- Membagi data ke dalam beberapa *n cluster*.
- Data yang berkarakteristik sama dimasukkan ke dalam satu kelompok sementara data yang berkarakteristik berbeda dimasukkan dalam kelompok yang lain.



# Algoritma *K-Means*





# Rumus Jarak

- **Manhattan**
- **Minkowski**
- **Euclidean Distance**
  - ✓ Menghasilkan jarak terpendek dari dua titik yang diperhitungkan.
  - ✓ Bernilai mutlak -> nilai selalu positif.

$$d(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2}$$

$d(i, j)$  = jarak data ke i ke pusat *cluster* (*centroid*)

$X_{ki}$  = data ke i pada atribut data ke k

$X_{kj}$  = titik pusat ke j pada atribut ke k

# Contoh Kasus

*Clustering* data nasabah sebagai berikut:

No	Jumlah Rumah	Jumlah Mobil
A	1	3
B	3	3
C	4	3
D	5	3
E	1	2
F	4	2
G	1	1
H	2	1

Dimana:  
Titik pusat *cluster* B dan F

# TUGAS

Lakukan *clustering* apabila diketahui data sebagai berikut:

Mahasiswa	BD	IMK	Alpro	Sister	Web
A	75	71	71	81	72
B	80	72	75	73	72
C	73	80	74	81	72
D	74	82	80	73	88
E	78	78	77	87	74
F	71	71	66	78	86
G	73	71	71	85	71
H	73	72	66	73	76
I	75	72	71	80	73
J	76	71	71	71	73
K	72	67	71	71	71
L	76	73	74	74	74

Titik pusat cluster:  
C1 adalah data ketiga  
C2 adalah data keenam