

Statistika Industri II

TIP – FTP – UB

# **Analisis Klaster** *(Cluster Analysis)*

---

# Pendahuluan

- Analisis yang menghasilkan sejumlah klaster (kelompok).
- Analisis klaster diawali dengan pemahaman bahwa sejumlah data tertentu memiliki kemiripan antaranggotanya.
- Tujuan: menempatkan sekelompok objek ke dalam grup-grup berdasarkan atas kesamaan yang dimiliki objek dari berbagai karakteristik.

# Asumsi

- Sampel yang diambil harus mewakili populasi
- Multikolinearitas, yaitu kemungkinan adanya korelasi antar-objek.

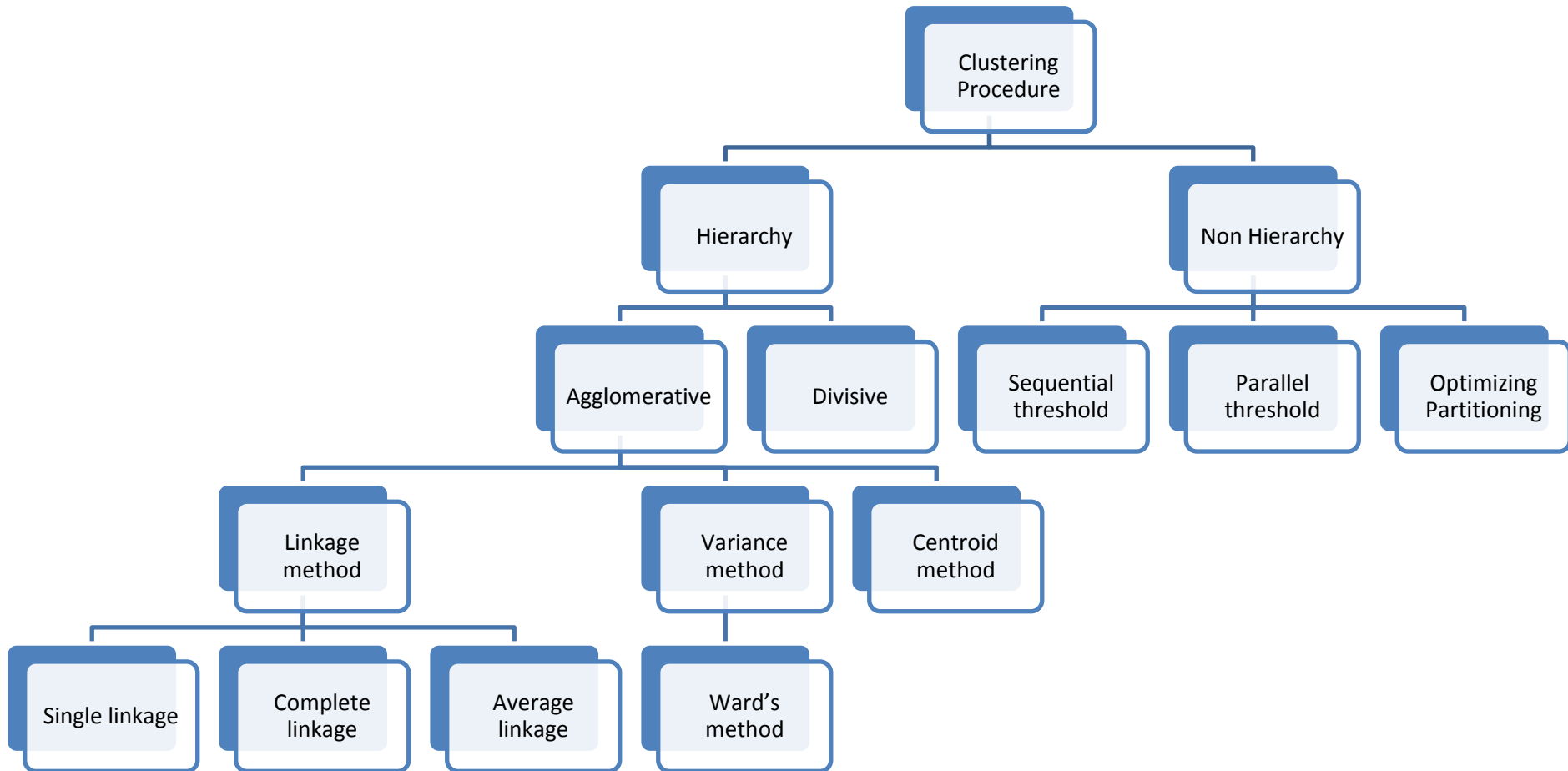
# Proses Analisis Kluster

- Menetapkan ukuran jarak antar-data
- Melakukan proses standarisasi data jika diperlukan (data nominal dan ordinal)
- Melakukan proses clustering
  - Hirarki atau non-hirarki (dianjurkan dipakai bersamaan)
- Melakukan penamaan klaster-klaster yang terbentuk
- Melakukan validasi dan profiling klaster
  - Bagi dua grup acak dan uji
  - Melihat hasil klaster yang jelas berbeda
  - Melakukan analisis dengan data tp metode berbeda

# Pemilihan Ukuran Kesamaan

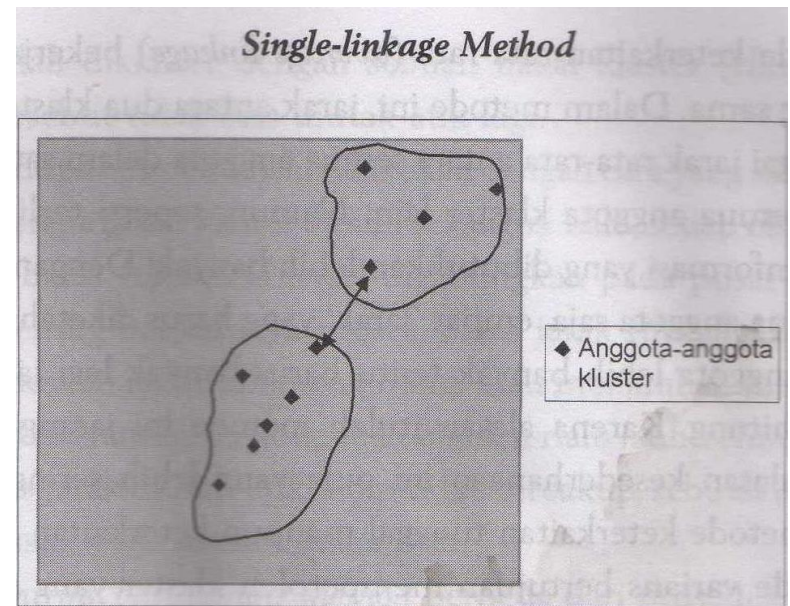
- Jarak euclidean  $d_{ij} = \sqrt{\sum_{k=1}^n (v_{ik} - v_{jk})^2}$ 
  - *Squared euclidean distance*
- Jarak cityblock
  - Jarak antara objek yang berpasangan dimutlakkan kemudian dijumlahkan.
- Jarak chebychev → selisih nilai mutlak tertinggi antar objek

# Prosedur Pengklasteran

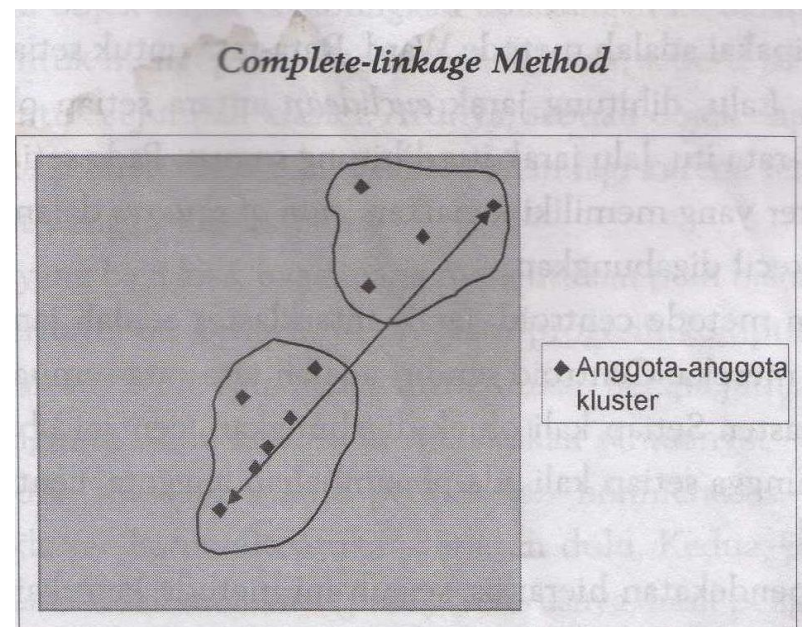


# Hierarki kluster

- Single linkage → mengelompokkan dua objek yang mempunyai jarak terdekat terlebih dahulu.

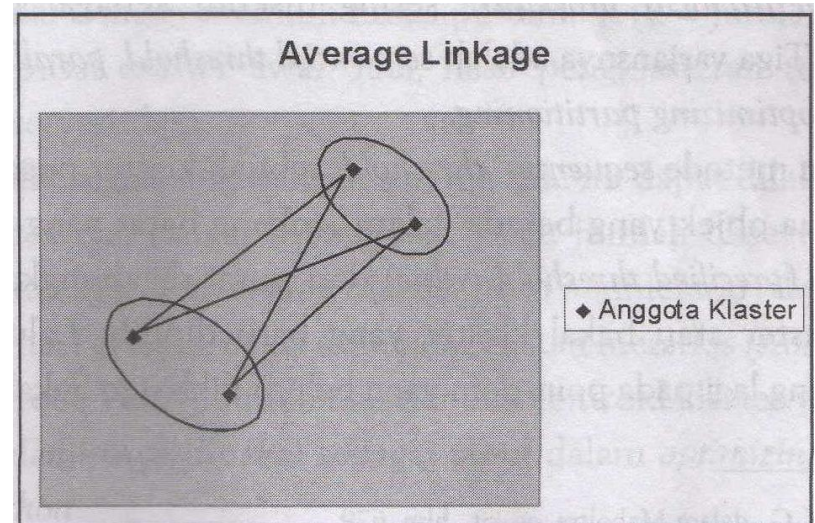


- Complete linkage → mengelompokkan dua objek yang mempunyai jarak terjauh terlebih dahulu, kemudian dilanjutkan untuk jarak antar-variabel yang makin dekat.





- Average linkage → mengelompokkan objek berdasar jarak rata-rata yang didapat dengan melakukan rata-rata semua jarak antar-objek terlebih dahulu.



- Ward's method
  - Memiliki varians internal klaster sekecil mungkin
  - Hitung jarak euclidean dan rata-rata, kemudian sum of squares di antara dua cluster yang memiliki kenaikan terkecil digabungkan
- Centroid method
  - Centroid adalah rata-rata jarak yang ada pada sebuah klaster.
  - Jarak antarklaster adalah jarak antara centroid klaster-klaster.

<b>Responden</b>	<b>v<sub>1</sub></b>	<b>v<sub>2</sub></b>	<b>v<sub>3</sub></b>	<b>v<sub>4</sub></b>	<b>v<sub>5</sub></b>	<b>v<sub>6</sub></b>
1	6	5	6	3	3	4
2	2	3	2	4	5	4
3	6	3	6	4	2	3
4	3	7	4	5	2	7
5	2	4	2	2	7	4
6	6	4	6	3	3	4
7	5	3	6	3	3	4
8	7	2	7	4	2	4
9	2	7	2	3	7	3
10	3	5	3	6	4	6
11	2	5	2	3	5	3
12	5	4	5	4	2	4
13	2	3	2	5	4	4
14	4	6	4	6	3	6
15	6	5	4	2	1	4
16	3	5	4	6	5	7
17	4	4	7	2	2	5
18	3	7	2	6	4	3
19	4	6	3	6	3	6
20	3	4	3	4	7	3

### Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster	Cluster 2		Cluster	Cluster 2	
	1			1		
1	14	19	.500	0	0	9
2	1	6	1.000	0	0	7
3	2	13	2.000	0	0	17
4	10	16	3.500	0	0	13
5	7	12	5.500	0	0	7
6	3	8	7.500	0	0	15
7	1	7	10.250	2	5	11
8	5	20	13.750	0	0	10
9	4	14	17.250	0	1	13
10	5	11	21.083	8	0	12
11	1	17	26.633	7	0	14
12	5	9	32.550	10	0	17
13	4	10	39.450	9	4	16
14	1	15	47.150	11	0	15
15	1	3	57.275	14	6	19
16	4	18	70.708	13	0	18
17	2	5	85.292	3	12	18
18	2	4	150.792	17	16	19
19	1	2	288.550	15	18	0

dimension  
0

Jumlah  
klaster

19  
18  
17  
16  
15  
14  
13  
12  
11  
10  
9  
8  
7  
6  
5  
4  
3  
2  
1

# Dendrogram using Ward Linkage

## Rescaled Distance Cluster Combine

