



ANALISIS GEROMBOL

Dhea Dewanti & Nur Khamidah

LATAR BELAKANG

Dalam beberapa hal, tujuan suatu penelitian adalah untuk **mengelompokkan individu-individu berdasarkan banyak peubah penciri**

Analisis gerombol merupakan metode yang dapat menggabungkan beberapa individu ke dalam kelompok-kelompok berdasarkan sifat kemiripan atau sifat ketidakmiripan antar objek, sehingga objek dalam kelompok lebih mirip dibandingkan dengan objek antar kelompok

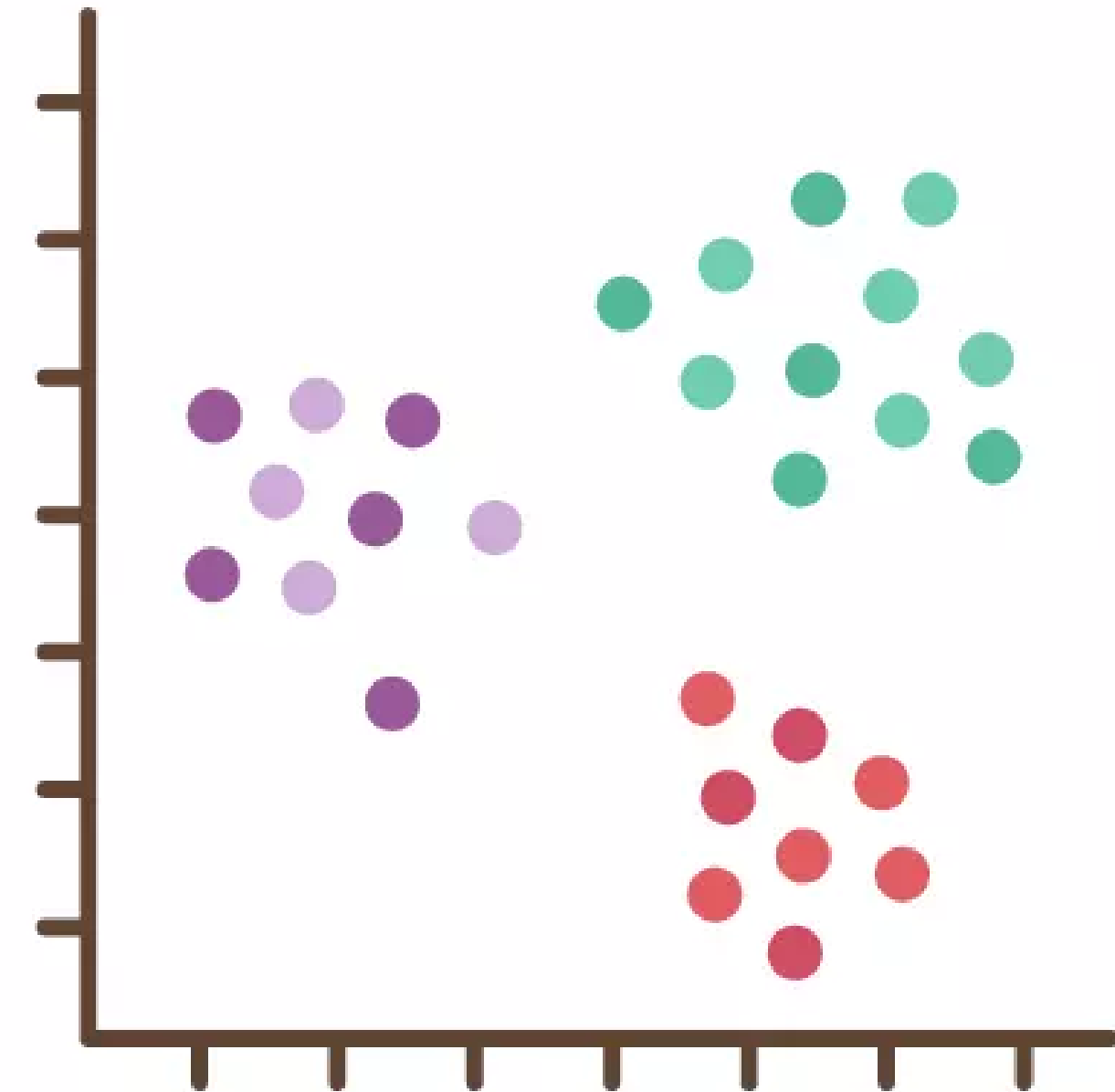
Individu-individu dalam satu kelompok memiliki kemiripan atau keragaman yang kecil dibandingkan individu-individu di kelompok yang berbeda



APA ITU ANALISIS GEROMBOL?

Analisis gerombol (Clustering)

adalah pengelompokan objek data ke dalam kelompok dengan kemiripan menurut ukuran jarak yang ditentukan.



STRUKTUR DATA AMATAN

Individu	Peubah				
	X1	X2	X3	...	Xp
1	x11	x12	x13		x1p
2	x21	x22	x23		x2p
3	x31	x32	x33		x3p
4	x41	x42	x43		x4p
5	x51	x52	x53		x5p
...
...
n	xn1	xn2	xn3		xnp

n: banyaknya sampel atau amatan

p: banyaknya peubah

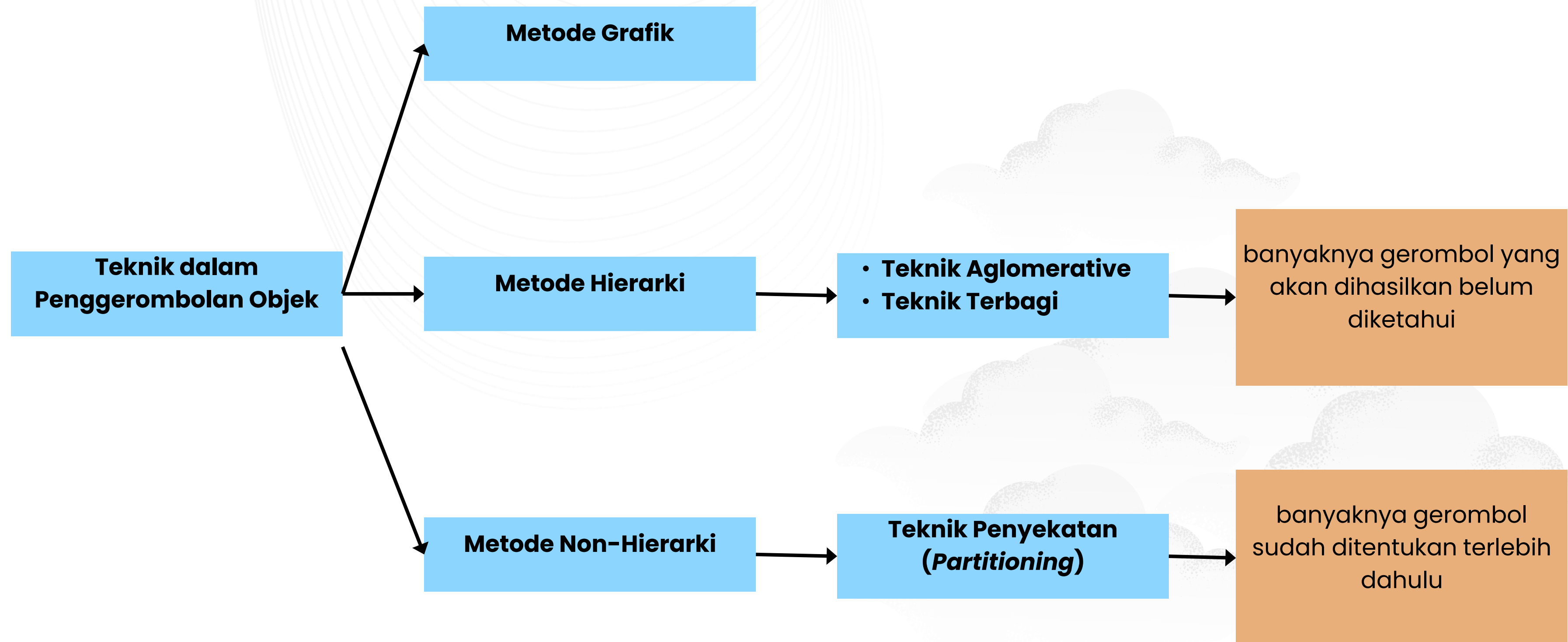
Struktur data sama dengan AKU

UKURAN JARAK

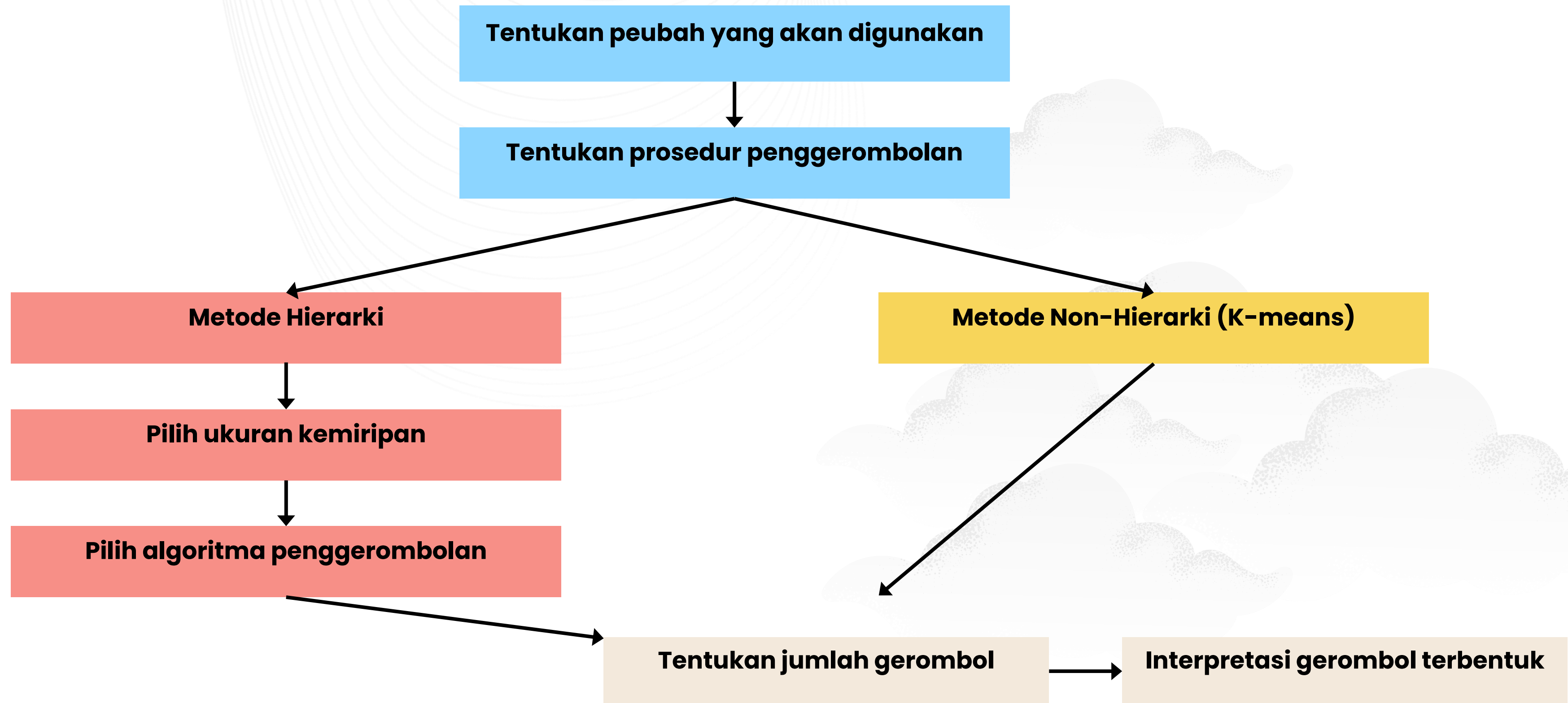
Jarak	Formula
<i>Jarak Euclidean</i>	$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y})}$ $= \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$
<i>Jarak Minkowski / Jarak city-block / Jarak Manhattan</i>	$d(\mathbf{x}, \mathbf{y}) = \left[\sum_{i=1}^p x_i - y_i ^k \right]^{\frac{1}{k}}$
<i>Jarak Mahalanobis</i>	$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})' \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y})}$

- **Jarak euclidean dan manhattan** baik digunakan jika peubah memiliki satuan yang sama
- **Jarak mahalanobis** dapat digunakan untuk kondisi peubah memiliki satuan yang sama maupun tidak

BEBERAPA TEKNIK PENGGEROMBOLAN

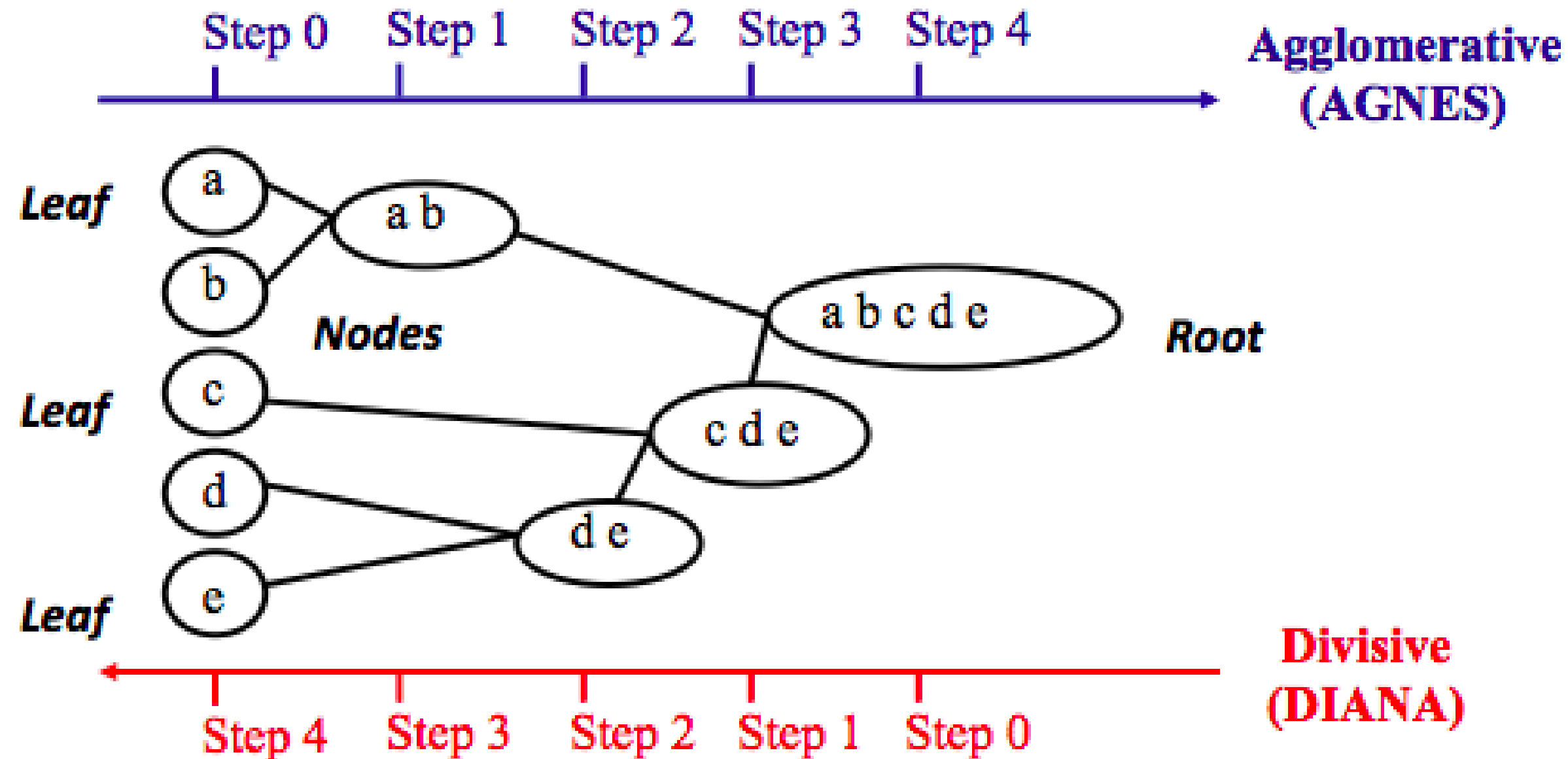


TAHAPAN PROSES ANALISIS GEROMBOL



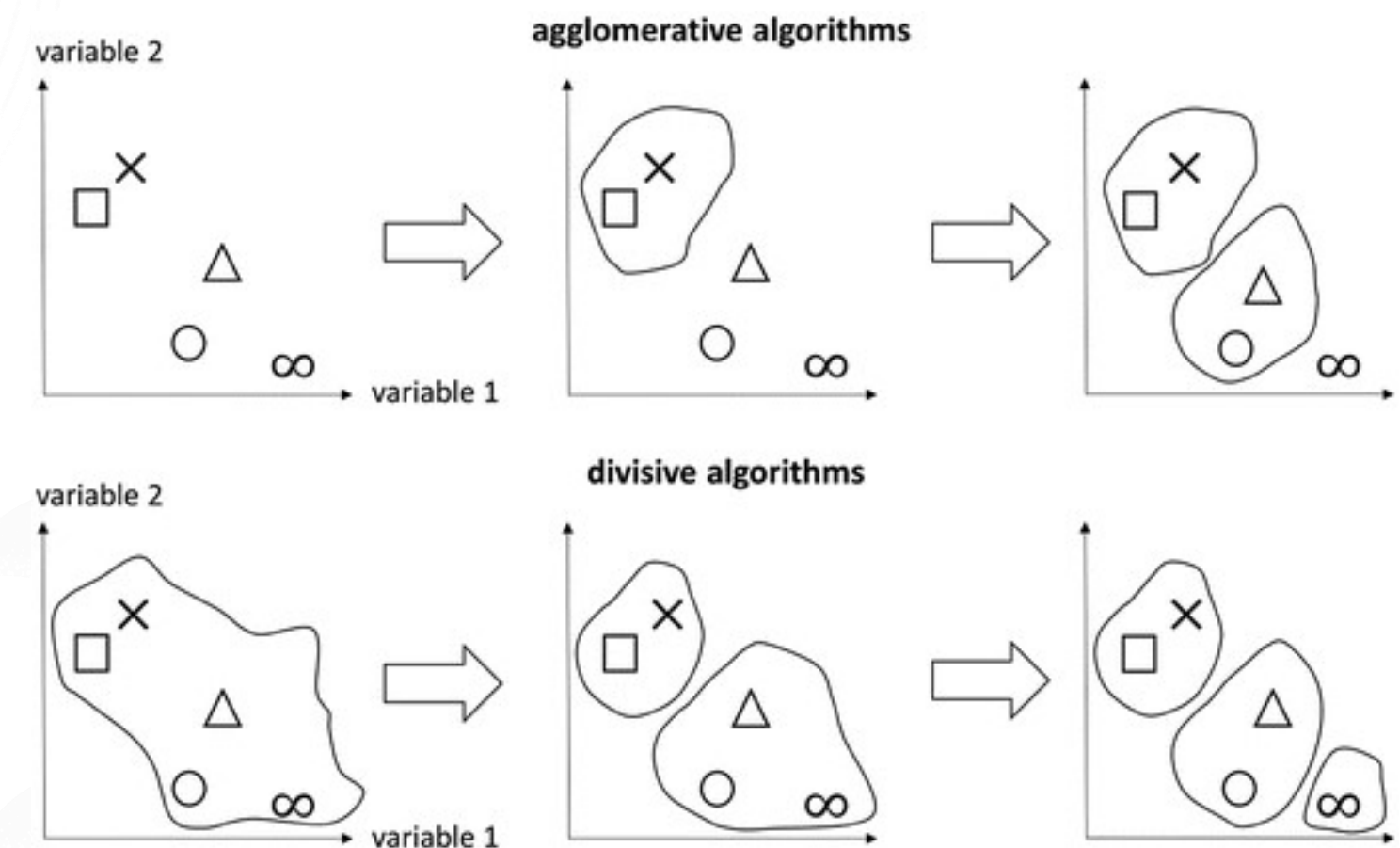
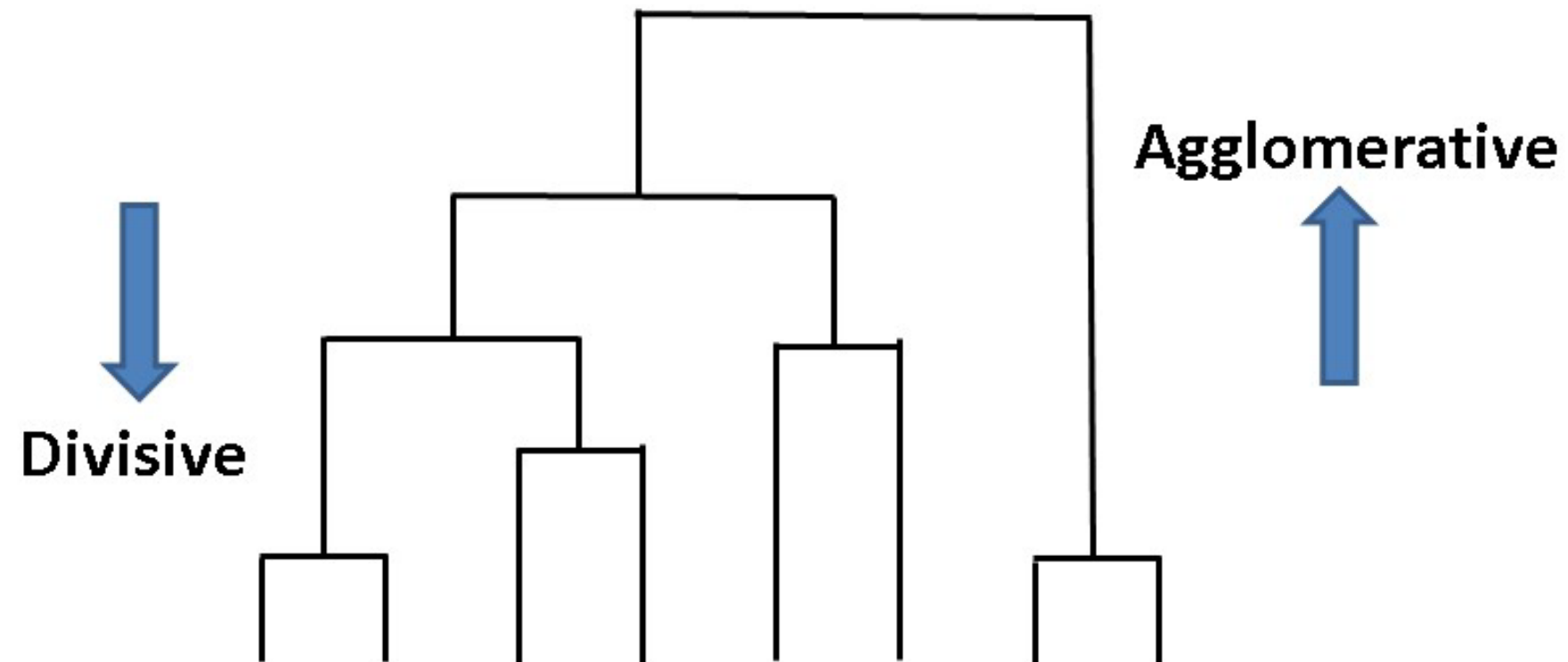
METODE HIERARKI – PENDEKATAN

Secara umum, hierarchical clustering dibagi menjadi dua jenis, yaitu **agglomerative** dan **divisive**. Kedua metode ini dibedakan berdasarkan pendekatan dalam melakukan pengelompokan data hingga membentuk dendrogram, menggunakan **bottom-up** atau **top-down manner**.



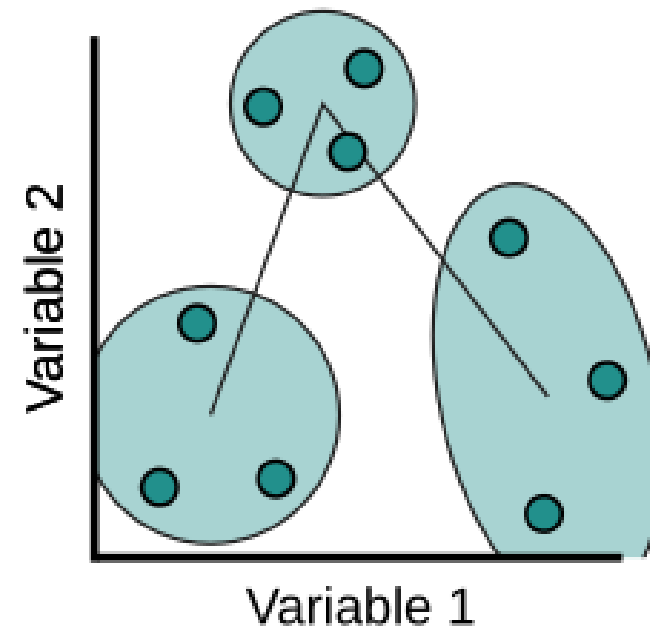
METODE HIERARKI – PENDEKATAN

Secara umum, hierarchical clustering dibagi menjadi dua jenis, yaitu **agglomerative** dan **divisive**. Kedua metode ini dibedakan berdasarkan pendekatan dalam melakukan pengelompokan data hingga membentuk dendrogram, menggunakan **bottom-up** atau **top-down manner**.

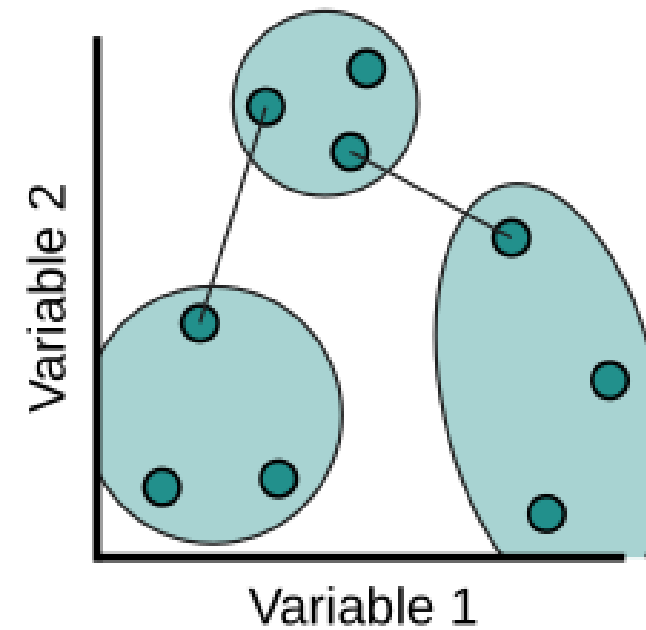


METODE HIERARKI – UKURAN KEMIRIPAN (LINKAGE)

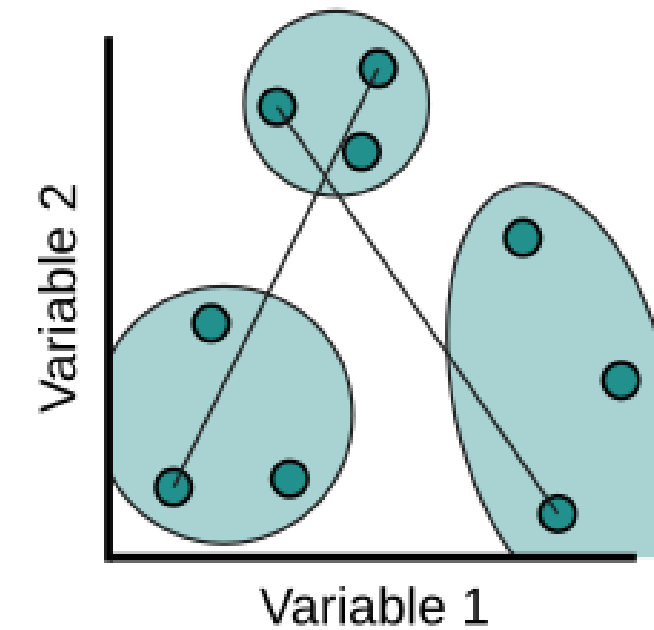
Centroid linkage



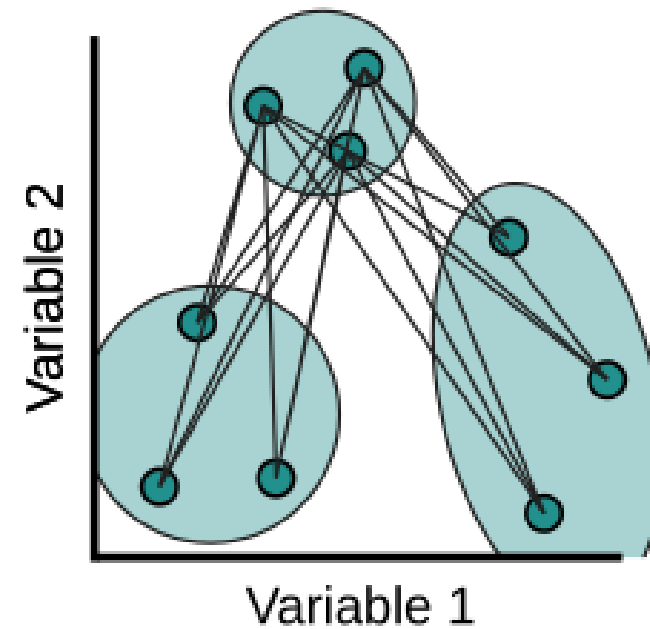
Single linkage



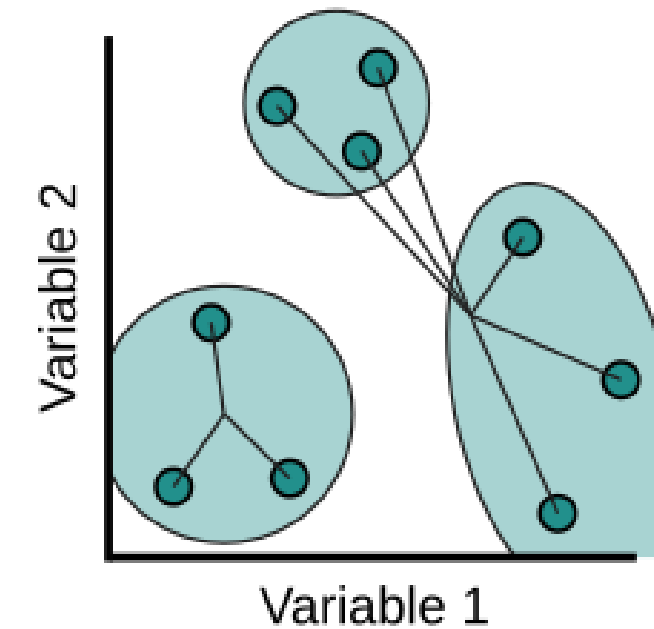
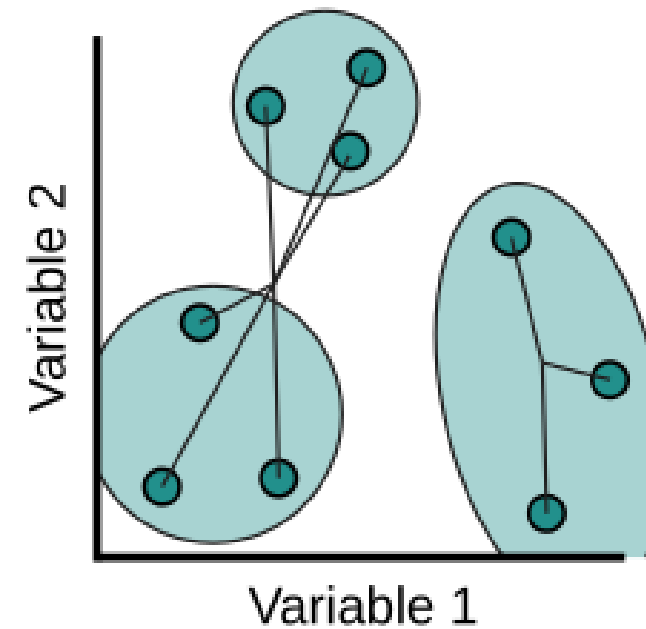
Complete linkage



Average linkage



Ward's method



METODE HIERARKI – ILUSTRASI

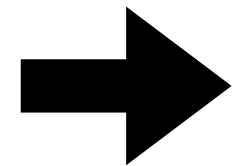
Diketahui korelasi diantara 5 toko penyedia bahan makanan pokok di kota Bogor sebagai berikut:

	A	B	C	D	E
A	1				
B	0,58	1			
C	0,51	0,6	1		
D	0,39	0,39	0,44	1	
E	0,46	0,32	0,43	0,52	1

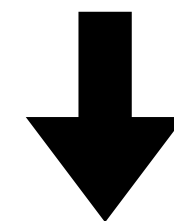
METODE HIERARKI – ILUSTRASI (SINGLE LINKAGE)

Diketahui korelasi diantara 5 toko penyedia bahan makanan pokok di kota Bogor sebagai berikut:

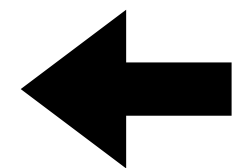
	A	B	C	D	E
A	1				
B	0,58	1			
C	0,51	0,6	1		
D	0,39	0,39	0,44	1	
E	0,46	0,32	0,43	0,52	1



	A	BE	C	D
A	1			
BE	0,46	1		
C	0,51	0,43	1	
D	0,39	0,39	0,44	1



	AD	BE	C
AD	1		
BE	0,39	1	
C	0,44	0,43	1

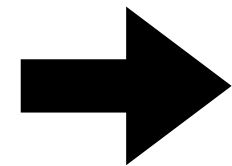


	ADBE	C
ADBE	1	
C	0,43	1

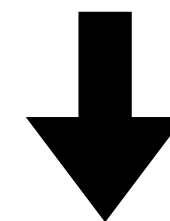
METODE HIERARKI – ILUSTRASI (COMPLETE LINKAGE)

Diketahui korelasi diantara 5 toko penyedia bahan makanan pokok di kota Bogor sebagai berikut:

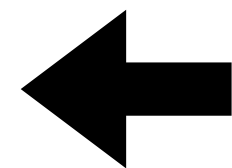
	A	B	C	D	E
A	1				
B	0,58	1			
C	0,51	0,6	1		
D	0,39	0,39	0,44	1	
E	0,46	0,32	0,43	0,52	1



	A	BE	C	D
A	1			
BE	0,58	1		
C	0,51	0,60	1	
D	0,39	0,52	0,44	1



	AD	BE	C
AD	1		
BE	0,58	1	
C	0,51	0,60	1



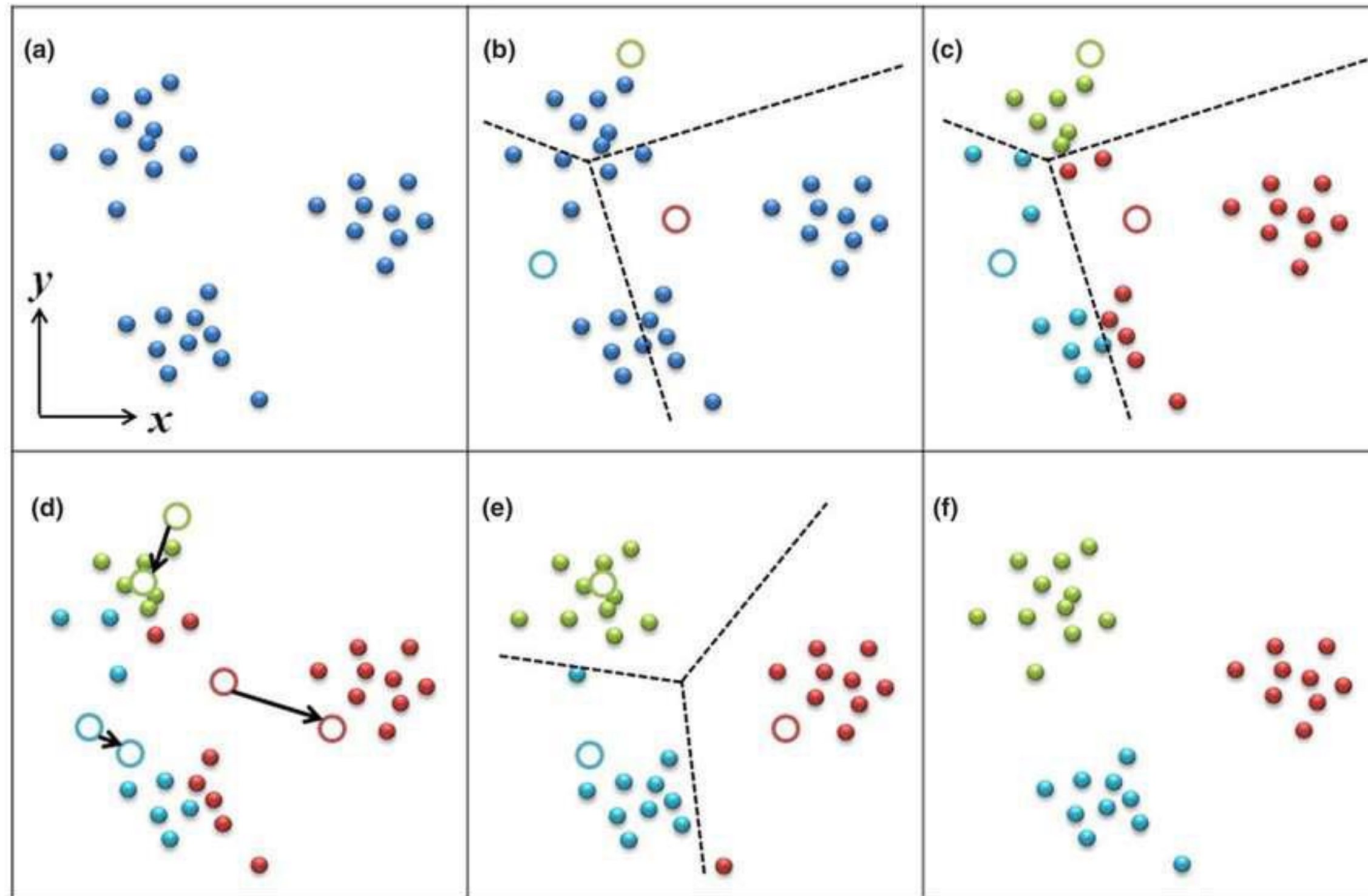
	ADC	BE
ADC	1	
BE	0,60	1

METODE NON-HIERARKI – K-MEANS

Algoritma:

- Tentukan besarnya k , yaitu banyaknya gerombol, dan tentukan juga centroid di tiap gerombol.
- Hitung jarak antara setiap objek dengan setiap centroid.
- Letakkan objek ke gerombol dengan jarak terdekat
- Hitung kembali rata-rata (centroid) untuk gerombol yang baru terbentuk.
- Ulangi langkah 2 sampai 4 hingga tidak ada lagi pemindahan objek antar gerombol.

METODE NON-HIERARKI – K-MEANS



STUDI KASUS

CustomerID	Genre	Age	Annual Income	Spending Score
1	Male	19	15	39
2	Male	21	15	81
3	Female	20	16	6
4	Female	23	16	77
5	Female	31	17	40
6	Female	22	17	76
7	Female	35	18	6

Data:

https://raw.githubusercontent.com/nurkhamidah/dat/main/mall_customers.csv

Seorang pemilik Mall ingin mengelompokan customer di Mall yang ia miliki, sehingga tim marketing bisa mengembangkan strategi yang tepat untuk customer yang tepat pula. Data yang dimiliki oleh Mall tersebut adalah **Customer ID**, umur pelanggan (**Age**), pendapatan tahunan dalam ribu dollar (**Annual Income**), dan **Spending Score**. Spending score merupakan nilai yang diberikan oleh Mall kepada customer berdasarkan perilaku customer (waktu kunjungan, jenis barang yang dibeli, dan banyaknya uang yang dihabiskan dalam belanja) yang memiliki rentang nilai 1-100. Semakin besar nilai Spending Score berarti customer semakin loyal pada Mall tersebut dan semakin besar pula uang belanja yang digunakan.



THANKYOU

Any Question, Just Ask don't be shy