



# Course Contents

1

Agglomerative Hierarchical Clustering (AHC)

2

Studi Kasus

3

Latihan dan Diskusi

4

Progress Final Project

# » Hierarchical Clustering

- Hierarchical Clustering adalah metode analisis kelompok yang berusaha untuk membangun sebuah hirarki kelompok data.
- Strategi pengelompokannya umumnya ada 2 jenis yaitu Agglomerative (Bottom-Up) dan Devisive (Top-Down). (Pada bagian ini akan dibatasi hanya menggunakan konsep Agglomerative).
- Algoritma Agglomerative Hierarchical Clustering :
  1. Hitung Matrik Jarak antar data.
  2. Ulangi langkah 3 dan 4 hingga hanya satu kelompok yang tersisa.
  3. Gabungkan dua kelompok terdekat berdasarkan parameter kedekatan yang ditentukan.
  4. Perbarui Matrik Jarak antar data untuk merepresentasikan kedekatan diantara kelompok baru dan kelompok yang masih tersisa.
  5. Selesai.



## Rumus Umum

- Membentuk Matrik Jarak, misal dengan Manhattan Distance :

$$D_{man}(x, y) = \sum_{j=1}^d |x_j - y_j|$$

atau menggunakan Euclidian Distance :

$$D(x_2, x_1) = \sqrt{\sum_{j=1}^d |x_{2j} - x_{1j}|^2}$$

- Beberapa metode Pengelompokan Secara Hierarki Aglomeratif :

- Single Linkage (Jarak Terdekat)

$$d_{uv} = \min\{d_{uv}\}, d_{uv} \in D$$

- Complete Linkage (Jarak Terjauh)

$$d_{uv} = \max\{d_{uv}\}, d_{uv} \in D$$

- Average Linkage (Jarak Rata-Rata)

$$d_{uv} = \text{average}\{d_{uv}\}, d_{uv} \in D$$



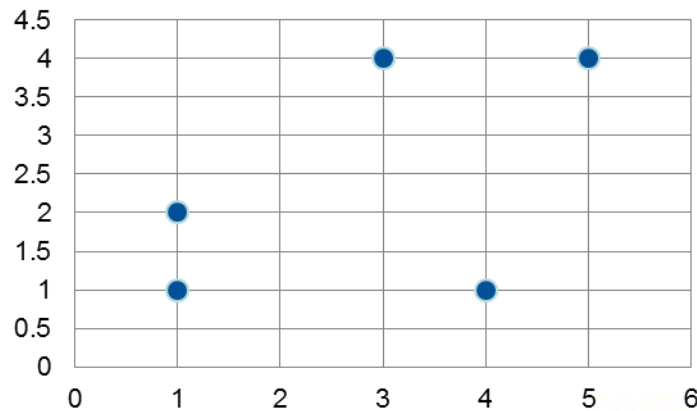
# Contoh Studi Kasus

- Perhatikan dataset berikut :

Data	Fitur x	Fitur y
1	1	1
2	4	1
3	1	2
4	3	4
5	5	4

Kelompokkan dataset tersebut dengan menggunakan metode AHC (Single Linkage, Complete Linkage dan Average Linkage) menggunakan jarak Manhattan !

## Visualisasi Data





- Menghitung Jarak Pada Semua Pasangan dua data :

Data	Fitur x	Fitur y
1	1	1
2	4	1
3	1	2
4	3	4
5	5	4

$$D_{man}(Data_1, Data_1) = \sum_{j=1}^2 |x_j - y_j| = |1-1| + |1-1| = 0$$

$$D_{man}(Data_1, Data_2) = |1-4| + |1-1| = 3$$

$$D_{man}(Data_1, Data_3) = |1-1| + |1-2| = 1$$

$$D_{man}(Data_1, Data_4) = |1-3| + |1-4| = 2 + 3 = 5$$

$$D_{man}(Data_3, Data_4) = |1-3| + |2-4| = 2 + 2 = 4$$

$$D_{man}(Data_1, Data_5) = |1-5| + |1-4| = 4 + 3 = 7$$

$$D_{man}(Data_3, Data_5) = |1-5| + |2-4| = 4 + 2 = 6$$

$$D_{man}(Data_2, Data_3) = |4-1| + |1-2| = 3 + 1 = 4$$

$$D_{man}(Data_4, Data_5) = |3-5| + |4-4| = 2 + 0 = 2$$

$$D_{man}(Data_2, Data_4) = |4-3| + |1-4| = 1 + 3 = 4$$

$$D_{man}(Data_2, Data_5) = |4-5| + |1-4| = 1 + 3 = 4$$

- Hasil Matrik Jarak :

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

# Single Linkage

- Menggunakan Metode Single Linkage :

Dengan memperlakukan data sebagai kelompok, selanjutnya kita pilih jarak dua kelompok yang terkecil.

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

$$\min(D_{man}) = \min(d_{13}) = 1$$

terpilih kelompok 1 dan 3, sehingga kedua kelompok ini digabungkan. (Melanjutkan pengelompokan).

- Menghitung jarak antar kelompok (1 dan 3) dengan kelompok lain yang tersisa, yaitu 2, 4 dan 5.

$$d_{(13)2} = \min\{d_{12}, d_{32}\} = \min\{3, 4\} = 3$$

$$d_{(13)4} = \min\{d_{14}, d_{34}\} = \min\{5, 4\} = 4$$

$$d_{(13)5} = \min\{d_{15}, d_{35}\} = \min\{7, 6\} = 6$$



- Menghitung jarak antar kelompok (1 dan 3) dengan kelompok lain yang tersisa, yaitu 2, 4 dan 5.

$$d_{(13)2} = \min\{d_{12}, d_{32}\} = \min\{3, 4\} = 3 \quad \text{red dot}$$

$$d_{(13)4} = \min\{d_{14}, d_{34}\} = \min\{5, 4\} = 4 \quad \text{green dot}$$

$$d_{(13)5} = \min\{d_{15}, d_{35}\} = \min\{7, 6\} = 6 \quad \text{blue dot}$$

- Dengan menghapus baris-baris dan kolom-kolom matrik jarak yang bersesuaian dengan kelompok 1 dan 3, serta menambahkan baris dan kolom untuk kelompok (13).

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0



Dman	(13)	2	4	5
(13)	0	red 3	green 4	blue 6
2	3	0	4	4
4	4	4	0	2
5	6	4	2	0

- Selanjutnya dipilih jarak dua kelompok yang terkecil.

$$\min(D_{man}) = \min(d_{45}) = 2$$



- Dengan menghapus baris-baris dan kolom-kolom matrik jarak yang bersesuaian dengan kelompok 1 dan 3, serta menambahkan baris dan kolom untuk kelompok (13).

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

→

Dman	(13)	2	4	5
(13)	0	3	4	6
2	3	0	4	4
4	4	4	0	2
5	6	4	2	0

- Selanjutnya dipilih jarak dua kelompok yang terkecil.

$$\min(D_{man}) = \min(d_{45}) = 2$$

- Menghitung jarak antar kelompok (4 dan 5) dengan kelompok lain yang tersisa, yaitu (13) dan 2.

$$d_{(45)(13)} = \min\{d_{41}, d_{43}, d_{51}, d_{53}\} = \min\{5, 4, 7, 6\} = 4$$

$$d_{(45)2} = \min\{d_{42}, d_{52}\} = \min\{4, 4\} = 4$$

- Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok 4 dan 5, serta menambahkan baris dan kolom untuk kelompok (45)






- Menghitung jarak antar kelompok (4 dan 5) dengan kelompok lain yang tersisa, yaitu (13) dan 2.

$$d_{(45)(13)} = \min\{d_{41}, d_{43}, d_{51}, d_{53}\} = \min\{5, 4, 7, 6\} = 4 \quad \bullet$$

$$d_{(45)2} = \min\{d_{42}, d_{52}\} = \min\{4, 4\} = 4 \quad \bullet$$

- Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok 4 dan 5, serta menambahkan baris dan kolom untuk kelompok (45)

Dman	(13)	2	4	5
(13)	0	3	4	6
2	3	0	4	4
4	4	4	0	2
5	6	4	2	0



Dman	(45)	(13)	2
(45)	0	4	4
(13)	4	0	3
2	4	3	0

- Selanjutnya dipilih jarak dua kelompok yang terkecil.

$$\min(D_{man}) = \min(d_{(13)2}) = 3$$

terpilih kelompok (13) dan 2, sehingga kedua kelompok ini digabungkan. (Melanjutkan pengelompokan).

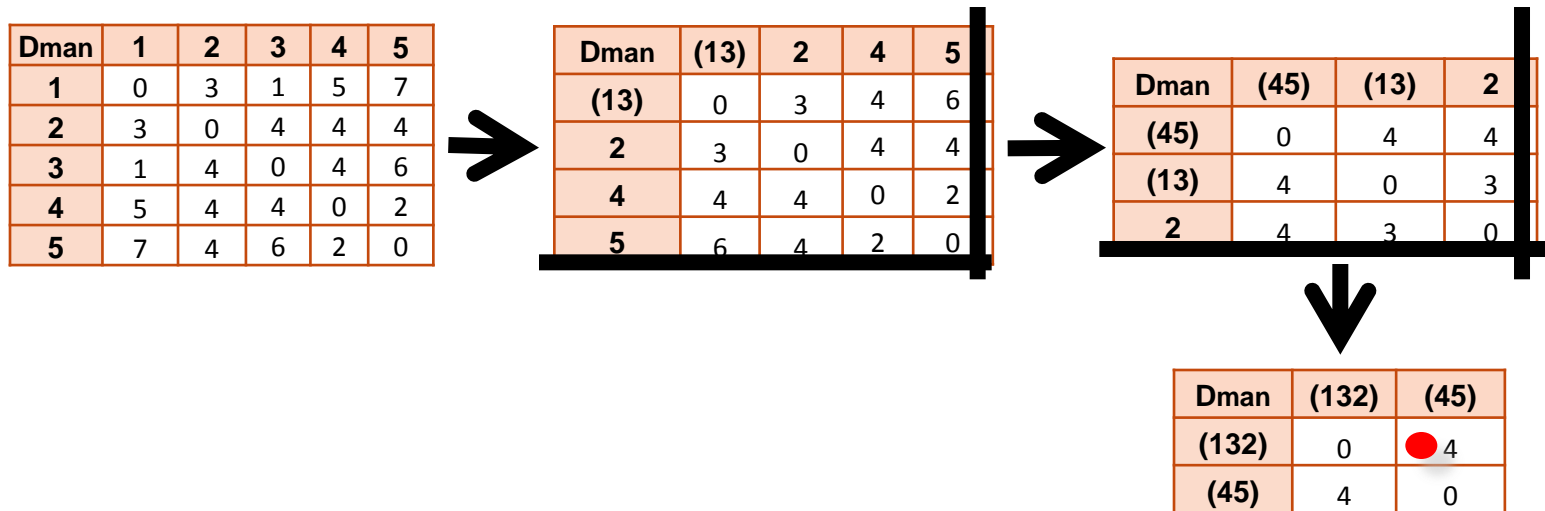
- Menghitung jarak antar kelompok ((13) dan 2) dengan kelompok lain yang tersisa, yaitu (45).



- Menghitung jarak antar kelompok ((13) dan 2) dengan kelompok lain yang tersisa, yaitu (45).

$$d_{(132)(45)} = \min\{d_{14}, d_{15}, d_{34}, d_{35}, d_{24}, d_{25}\} = \min\{5, 7, 4, 6, 4, 4\} = 4 \quad \bullet$$

- Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok (13) dan 2, serta menambahkan baris dan kolom untuk kelompok (123)



- Jadi kelompok (132) dan (45) digabung untuk menjadi kelompok tunggal dari lima data, yaitu kelompok (13245) dengan jarak terdekat 4.



- Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok (13) dan 2, serta menambahkan baris dan kolom untuk kelompok (132)

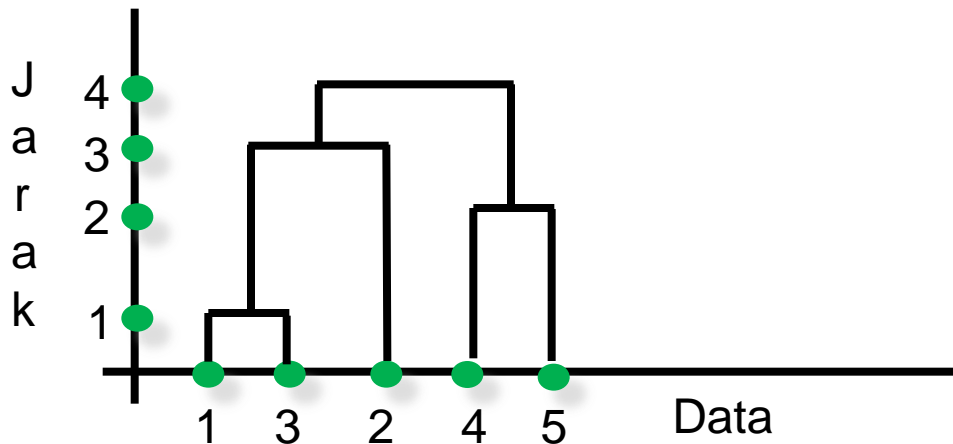
Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

Dman	(13)	2	4	5
(13)	0	3	4	6
2	3	0	4	4
4	4	4	0	2
5	6	4	2	0

Dman	(45)	(13)	2
(45)	0	4	4
(13)	4	0	3
2	4	3	0

Dman	(132)	(45)
(132)	0	4
(45)	4	0

- Jadi kelompok (132) dan (45) digabung untuk menjadi kelompok tunggal dari lima data, yaitu kelompok (13245) dengan jarak terdekat 4. Berikut Dendrogram Hasil Metode Single Linkage :



# Complete Linkage

- Menggunakan Metode Complete Linkage :

Dengan memperlakukan data sebagai kelompok, selanjutnya kita pilih jarak dua kelompok yang terkecil.

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

$$\min(D_{man}) = \min(d_{13}) = 1$$

terpilih kelompok 1 dan 3, sehingga kedua kelompok ini digabungkan. (Melanjutkan pengelompokan).

- Menghitung jarak antar kelompok (1 dan 3) dengan kelompok lain yang tersisa, yaitu 2, 4 dan 5.

$$d_{(13)2} = \max\{d_{12}, d_{32}\} = \max\{3, 4\} = 4$$

$$d_{(13)4} = \max\{d_{14}, d_{34}\} = \max\{5, 4\} = 5$$

$$d_{(13)5} = \max\{d_{15}, d_{35}\} = \max\{7, 6\} = 7$$



- Menghitung jarak antar kelompok (1 dan 3) dengan kelompok lain yang tersisa, yaitu 2, 4 dan 5.

$$d_{(13)2} = \max\{d_{12}, d_{32}\} = \max\{3, 4\} = 4 \quad \text{red dot}$$

$$d_{(13)4} = \max\{d_{14}, d_{34}\} = \max\{5, 4\} = 5 \quad \text{green dot}$$

$$d_{(13)5} = \max\{d_{15}, d_{35}\} = \max\{7, 6\} = 7 \quad \text{blue dot}$$

- Dengan menghapus baris-baris dan kolom-kolom matrik jarak yang bersesuaian dengan kelompok 1 dan 3, serta menambahkan baris dan kolom untuk kelompok (13).

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0



Dman	(13)	2	4	5
(13)	0	4	5	7
2	4	0	4	4
4	5	4	0	2
5	7	4	2	0

- Selanjutnya dipilih jarak dua kelompok yang terkecil.

$$\min(D_{man}) = \min(d_{45}) = 2$$



- Dengan menghapus baris-baris dan kolom-kolom matrik jarak yang bersesuaian dengan kelompok 1 dan 3, serta menambahkan baris dan kolom untuk kelompok (13).

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

→

Dman	(13)	2	4	5
(13)	0	4	5	7
2	4	0	4	4
4	5	4	0	2
5	7	4	2	0

- Selanjutnya dipilih jarak dua kelompok yang terkecil.

$$\min(D_{man}) = \min(d_{45}) = 2$$

- Menghitung jarak antar kelompok (4 dan 5) dengan kelompok lain yang tersisa, yaitu (13) dan 2.

$$d_{(45)(13)} = \max\{d_{41}, d_{43}, d_{51}, d_{53}\} = \max\{5, 4, 7, 6\} = 7$$

$$d_{(45)2} = \max\{d_{42}, d_{52}\} = \max\{4, 4\} = 4$$

- Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok 4 dan 5, serta menambahkan baris dan kolom untuk kelompok (45)




- Menghitung jarak antar kelompok (4 dan 5) dengan kelompok lain yang tersisa, yaitu (13) dan 2.

$$d_{(45)(13)} = \max\{d_{41}, d_{43}, d_{51}, d_{53}\} = \max\{5, 4, 7, 6\} = 7 \quad \bullet$$

$$d_{(45)2} = \max\{d_{42}, d_{52}\} = \max\{4, 4\} = 4 \quad \bullet$$

- Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok 4 dan 5, serta menambahkan baris dan kolom untuk kelompok (45)

Dman	(13)	2	4	5
(13)	0	4	5	7
2	4	0	4	4
4	5	4	0	2
5	7	4	2	0



Dman	(45)	(13)	2
(45)	0	7	4
(13)	7	0	4
2	4	4	0

- Selanjutnya dipilih jarak dua kelompok yang terkecil.

$$\min(D_{man}) = \min(d_{(45)2}) = 4$$

terpilih kelompok (45) dan 2, sehingga kedua kelompok ini digabungkan. (Melanjutkan pengelompokan).

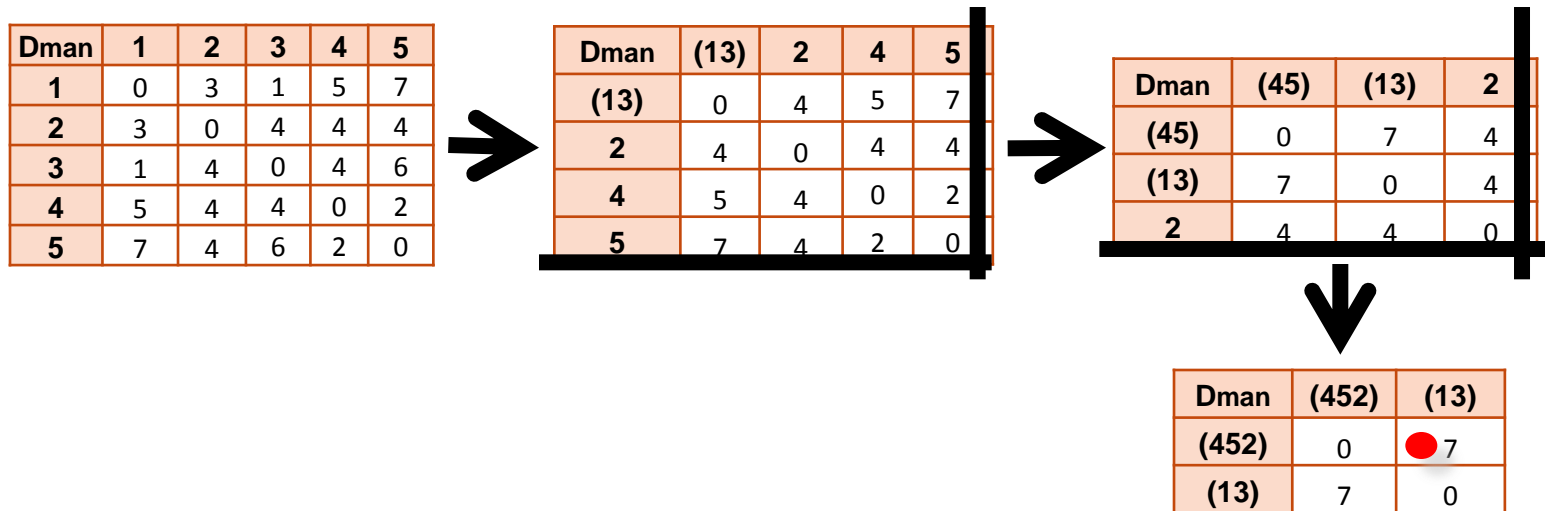
- Menghitung jarak antar kelompok ((45) dan 2) dengan kelompok lain yang tersisa, yaitu (13).



- Menghitung jarak antar kelompok ((45) dan 2) dengan kelompok lain yang tersisa, yaitu (13).

$$d_{(452)(13)} = \max\{d_{41}, d_{43}, d_{51}, d_{53}, d_{21}, d_{23}\} = \max\{5, 4, 7, 6, 3, 4\} = 7 \quad \bullet$$

- Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok (45) dan 2, serta menambahkan baris dan kolom untuk kelompok (452)



- Jadi kelompok (452) dan (13) digabung untuk menjadi kelompok tunggal dari lima data, yaitu kelompok (13452) dengan jarak terdekat 7.





- Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok (45) dan 2, serta menambahkan baris dan kolom untuk kelompok (452)

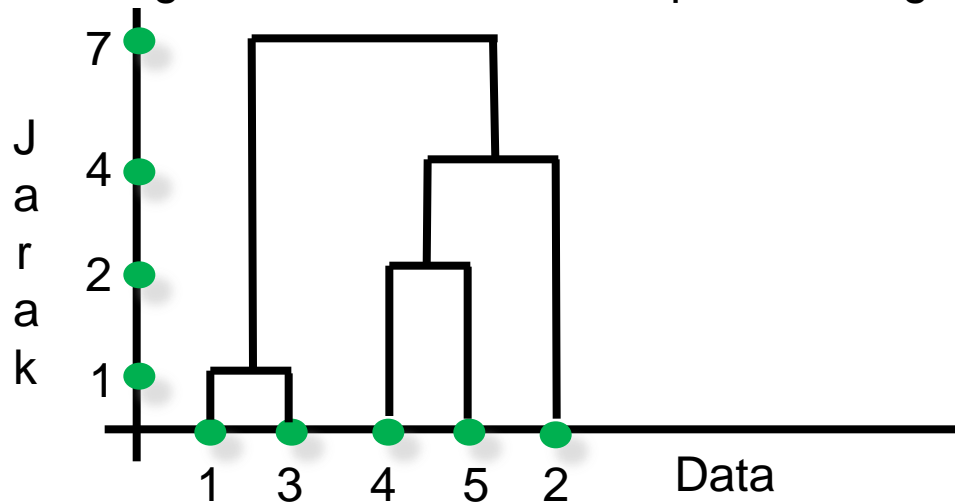
Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

Dman	(13)	2	4	5
(13)	0	4	5	7
2	4	0	4	4
4	5	4	0	2
5	7	4	2	0

Dman	(45)	(13)	2
(45)	0	7	4
(13)	7	0	4
2	4	4	0

Dman	(452)	(13)
(452)	0	7
(13)	7	0

- Jadi kelompok (452) dan (13) digabung untuk menjadi kelompok tunggal dari lima data, yaitu kelompok (13452) dengan jarak terdekat 7. Berikut Dendrogram Hasil Metode Complete Linkage :



# Average Linkage

- Menggunakan Metode Average Linkage :

Dengan memperlakukan data sebagai kelompok, selanjutnya kita pilih jarak dua kelompok yang terkecil.

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

$$\min(D_{man}) = \min(d_{13}) = 1$$

terpilih kelompok 1 dan 3, sehingga kedua kelompok ini digabungkan. (Melanjutkan pengelompokan).

- Menghitung jarak antar kelompok (1 dan 3) dengan kelompok lain yang tersisa, yaitu 2, 4 dan 5.

$$d_{(13)2} = \text{average}\{d_{12}, d_{32}\} = \text{average}\{3, 4\} = (3 + 4) / 2 = 3.5$$

$$d_{(13)4} = \text{average}\{d_{14}, d_{34}\} = \text{average}\{5, 4\} = (5 + 4) / 2 = 4.5$$

$$d_{(13)5} = \text{average}\{d_{15}, d_{35}\} = \text{average}\{7, 6\} = (7 + 6) / 2 = 6.5$$



- Menghitung jarak antar kelompok (1 dan 3) dengan kelompok lain yang tersisa, yaitu 2, 4 dan 5.

$$d_{(13)2} = \text{average}\{d_{12}, d_{32}\} = \text{average}\{3, 4\} = (3 + 4) / 2 = 3.5 \quad \bullet$$

$$d_{(13)4} = \text{average}\{d_{14}, d_{34}\} = \text{average}\{5, 4\} = (5 + 4) / 2 = 4.5 \quad \bullet$$

$$d_{(13)5} = \text{average}\{d_{15}, d_{35}\} = \text{average}\{7, 6\} = (7 + 6) / 2 = 6.5 \quad \bullet$$

- Dengan menghapus baris-baris dan kolom-kolom matrik jarak yang bersesuaian dengan kelompok 1 dan 3, serta menambahkan baris dan kolom untuk kelompok (13).

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0



Dman	(13)	2	4	5
(13)	0	3.5	4.5	6.5
2	3.5	0	4	4
4	4.5	4	0	2
5	6.5	4	2	0

- Selanjutnya dipilih jarak dua kelompok yang terkecil.

$$\min(D_{man}) = \min(d_{45}) = 2$$



- Dengan menghapus baris-baris dan kolom-kolom matrik jarak yang bersesuaian dengan kelompok 1 dan 3, serta menambahkan baris dan kolom untuk kelompok (13).

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

→

Dman	(13)	2	4	5
(13)	0	3.5	4.5	6.5
2	3.5	0	4	4
4	4.5	4	0	2
5	6.5	4	2	0

- Selanjutnya dipilih jarak dua kelompok yang terkecil.

$$\min(D_{man}) = \min(d_{45}) = 2$$

- Menghitung jarak antar kelompok (4 dan 5) dengan kelompok lain yang tersisa, yaitu (13) dan 2.

$$d_{(45)(13)} = \text{average}\{d_{41}, d_{43}, d_{51}, d_{53}\} = \text{average}\{5, 4, 7, 6\} = (5 + 4 + 7 + 6) / 4 = 5.25$$

$$d_{(45)2} = \text{average}\{d_{42}, d_{52}\} = \text{average}\{4, 4\} = (4 + 4) / 2 = 4$$

- Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok 4 dan 5, serta menambahkan baris dan kolom untuk kelompok (45)



- Menghitung jarak antar kelompok (4 dan 5) dengan kelompok lain yang tersisa, yaitu (13) dan 2.

- Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok 4 dan 5, serta menambahkan baris dan kolom untuk kelompok (45)

Dman	(13)	2	4	5
(13)	0	3.5	4.5	6.5
2	3.5	0	4	4
4	4.5	4	0	2
5	6.5	4	2	0



Dman	(45)	(13)	2
(45)	0	5.25	4
(13)	5.25	0	3.5
2	4	3.5	0

- Selanjutnya dipilih jarak dua kelompok yang terkecil.

$$\min(D_{man}) = \min(d_{(13)2}) = 3.5$$

terpilih kelompok (13) dan 2, sehingga kedua kelompok ini digabungkan. (Melanjutkan pengelompokan).

- Menghitung jarak antar kelompok ((13) dan 2) dengan kelompok lain yang tersisa, yaitu (45).



- Menghitung jarak antar kelompok ((13) dan 2) dengan kelompok lain yang tersisa, yaitu (45).

$$d_{(132)(45)} = \text{average}\{d_{14}, d_{15}, d_{34}, d_{35}, d_{24}, d_{25}\} = \text{average}\{5, 7, 4, 6, 4, 4\} = (5 + 7 + 4 + 6 + 4 + 4) / 6 = 5$$

- Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok (45) dan 2, serta menambahkan baris dan kolom untuk kelompok (452)

Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0



Dman	(13)	2	4	5
(13)	0	3.5	4.5	6.5
2	3.5	0	4	4
4	4.5	4	0	2
5	6.5	4	2	0



Dman	(45)	(13)	2
(45)	0	5.25	4
(13)	5.25	0	3.5
2	4	3.5	0



Dman	(132)	(45)
(132)	0	5
(45)	5	0

- Jadi kelompok (132) dan (45) digabung untuk menjadi kelompok tunggal dari lima data, yaitu kelompok (13245) dengan jarak terdekat 5.



- Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok (45) dan 2, serta menambahkan baris dan kolom untuk kelompok (452)

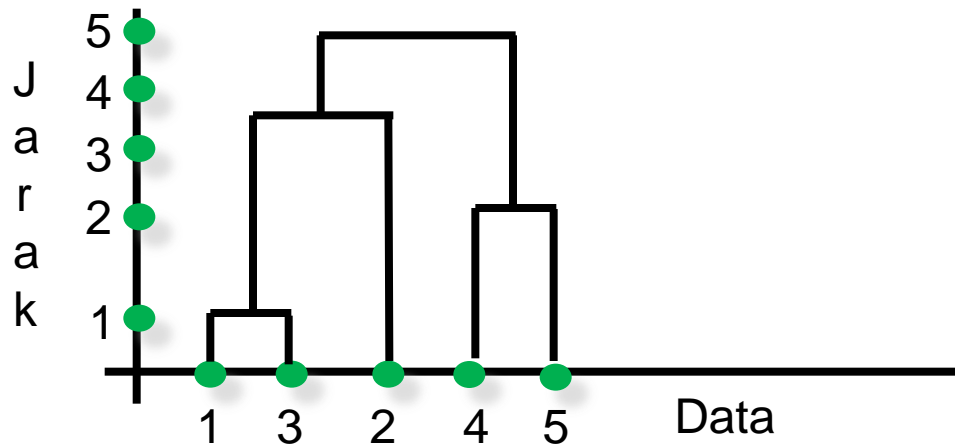
Dman	1	2	3	4	5
1	0	3	1	5	7
2	3	0	4	4	4
3	1	4	0	4	6
4	5	4	4	0	2
5	7	4	6	2	0

Dman	(13)	2	4	5
(13)	0	3.5	4.5	6.5
2	3.5	0	4	4
4	4.5	4	0	2
5	6.5	4	2	0

Dman	(45)	(13)	2
(45)	0	5.25	4
(13)	5.25	0	3.5
2	4	3.5	0

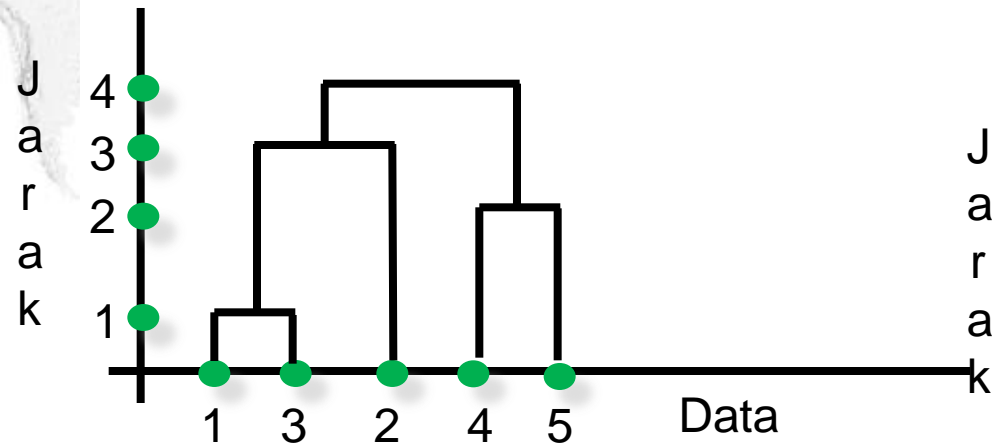
Dman	(132)	(45)
(132)	0	5
(45)	5	0

- Jadi kelompok (132) dan (45) digabung untuk menjadi kelompok tunggal dari lima data, yaitu kelompok (13245) dengan jarak terdekat 5. Berikut Dendrogram Hasil Metode Average Linkage :

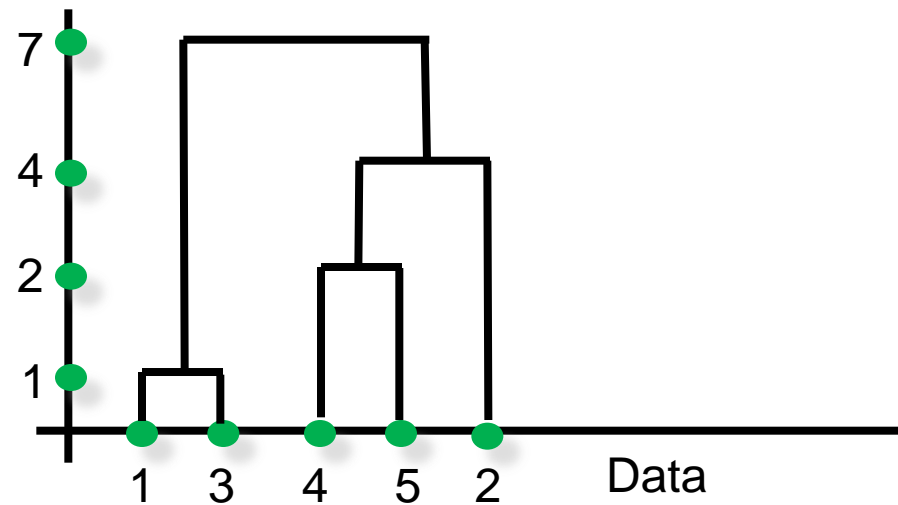




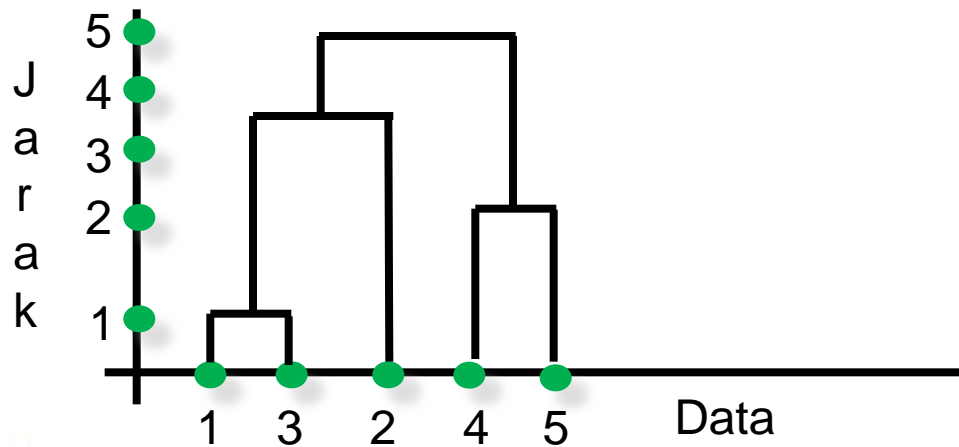
## Single-Linkage



## Complete-Linkage



## Average-Linkage

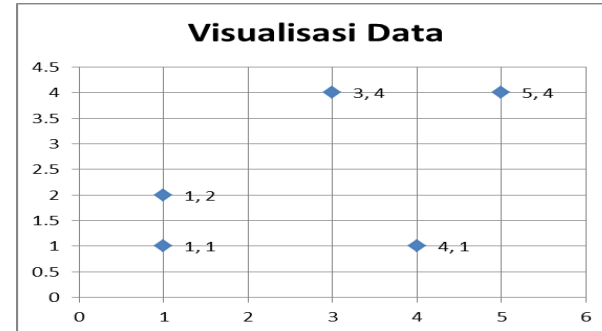




# Latihan

1. Perhatikan dataset 1 berikut :

Data	Fitur x	Fitur y
1	1	1
2	4	1
3	1	2
4	3	4
5	5	4



Kelompokkan dataset tersebut dengan menggunakan metode AHC (Single Linkage, Complete Linkage dan Average Linkage) menggunakan jarak Euclidian dan Visualisasikan Dendogramnya !

2. Perhatikan dataset 2 berikut :

Data	Fitur x	Fitur y
1	1	1
2	4	1
3	6	1
4	1	2
5	2	3
6	5	3
7	2	5
8	3	5
9	2	6
10	3	8

Kelompokkan dataset tersebut dengan menggunakan metode AHC Average Linkage menggunakan jarak Euclidian dan Visualisasikan Dendogramnya !

