



PROGRAM STUDI
TEKNIK INFORMATIKA – S1
FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO

MATA KULIAH
DATA MINING



[Technology vector created by sentavio - www.freepik.com](https://www.freepik.com/vectors/technology)

DATA MINING

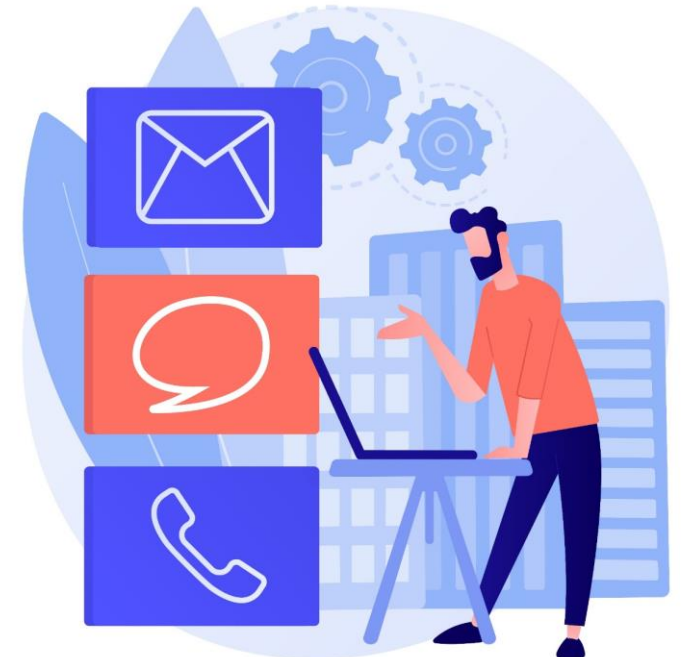
“Klastering dengan AHC (*Agglomerative Hierarchical Clustering*)”

TIM PENGAMPU DOSEN DATA MINING

2023

Kontak Dosen

- Junta Zeniarja, M.Kom
- Email: junta@dsn.dinus.ac.id
- Youtube : <https://www.youtube.com/JuntaZeniarja>
- Scholar : <http://bit.do/JuntaScholar>



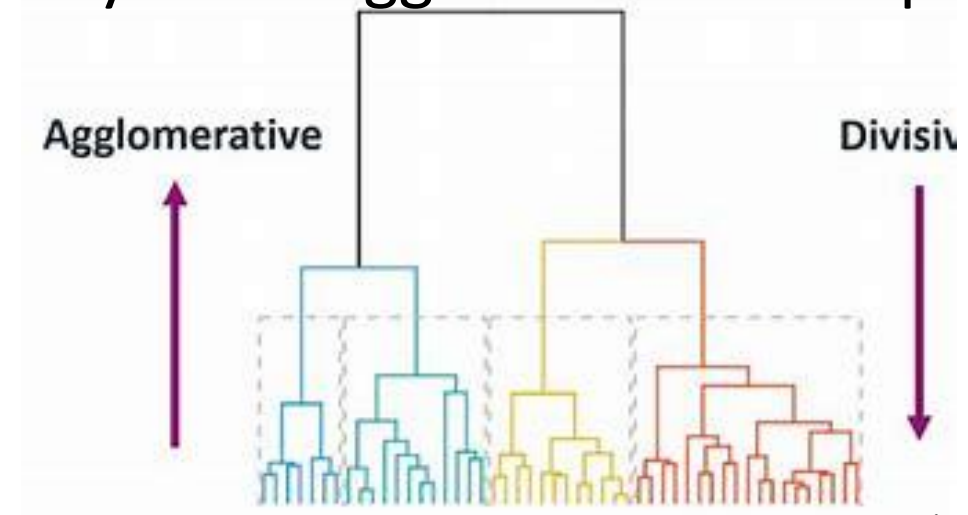


Klastering dengan AHC (Agglomerative Hierarchical Clustering)

Pertemuan 11

Algoritma AHC (*Agglomerative 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 Divisive (Top-Down). (Pada bagian ini akan dibatasi hanya menggunakan konsep Agglomerative (Bottom-Up).



Algoritma AHC (*Agglomerative Hierarchical Clustering*)

1. Hitung matrik jarak antar data (menggunakan Euclidean atau Manhattan Distance).
2. Gabungkan dua kelompok terdekat menjadi satu kelompok data:
 Single linkage (Jarak Terdekat).
 Complete Linkage (Jarak Terjauh).
 Average Linkage (Jarak Rata-rata)
3. Perbarui matrik jarak antar data untuk merepresentasikan antara kelompok baru dengan kelompok yang masih tersisa.
4. Ulangi langkah 2-3 sampai hanya satu kelompok yang tersisa.

Latihan Soal

- Kelompokkan dataset di bawah ini dengan menggunakan metode AHC (Single Linkage) dengan Manhattan Distance.
- Dataset Skor Kedisiplinan Siswa

Siswa	DTW	DTT	DMT	DDB
1	2	4	4	3
2	3	4	3	5
3	4	3	2	5
4	1	5	4	2
5	3	2	1	3

Kriteria Kedisiplinan		
1	Disiplin terhadap Waktu (DTW)	(1-5)
2	Disiplin terhadap tata terib (DTT)	(1-5)
3	Disiplin mengerjakan tugas (DMT)	(1-5)
4	Disiplin dalam berpakaian(DDB)	(1-5)

Langkah 1

- Hitung matrik jarak antar data.
- Menghitung jarak dua data secara berpasang-pasangan dengan menggunakan Manhattan Distance.

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

Siswa	DTW	DTT	DMT	DDB
1	2	4	4	3
2	3	4	3	5
3	4	3	2	5
4	1	5	4	2
5	3	2	1	3

$$D_{man}(D_1, D_1) = \sum_{j=1}^4 |x_j - y_j| = |2 - 2| + |4 - 4| + |4 - 4| + |3 - 3| = 0$$

$$D_{man}(D_1, D_2) = |2 - 3| + |4 - 4| + |4 - 3| + |3 - 5| = 4$$

$$D_{man}(D_1, D_3) = |2 - 4| + |4 - 3| + |4 - 2| + |3 - 5| = 7$$

$$D_{man}(D_1, D_4) = |2 - 1| + |4 - 5| + |4 - 4| + |3 - 2| = 3$$

$$D_{man}(D_1, D_5) = |2 - 3| + |4 - 2| + |4 - 1| + |3 - 3| = 6$$

Langkah 1 [2]

- Hitung matrik jarak antar data.
- Menghitung jarak dua data secara berpasang-pasangan dengan menggunakan Manhattan Distance.

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

Siswa	DTW	DTT	DMT	DDB
1	2	4	4	3
2	3	4	3	5
3	4	3	2	5
4	1	5	4	2
5	3	2	1	3

$$D_{man}(D_2, D_3) = |3 - 4| + |4 - 3| + |3 - 2| + |5 - 5| = 3$$

$$D_{man}(D_2, D_4) = |3 - 1| + |4 - 5| + |3 - 4| + |5 - 2| = 7$$

$$D_{man}(D_2, D_5) = |3 - 3| + |4 - 2| + |3 - 1| + |5 - 3| = 6$$

$$D_{man}(D_3, D_4) = |4 - 1| + |3 - 5| + |2 - 4| + |5 - 2| = 10$$

$$D_{man}(D_3, D_5) = |4 - 3| + |3 - 2| + |2 - 1| + |5 - 3| = 5$$

$$D_{man}(D_4, D_5) = |1 - 3| + |5 - 2| + |4 - 1| + |2 - 3| = 9$$

Langkah 2

- Gabungkan dua kelompok terdekat menjadi satu.
- Menggunakan Single Linkage pilih jarak dari dua kelompok yang terdekat.
 $\min(D_{man}) = \min(d_{23}) = 3$
- Terpilih kelompok 2 dan 3, sehingga kedua kelompok digabungkan.
- Menghitung jarak terdekat antar kelompok (2 dan 3) dengan kelompok yang tersisa, yaitu 1, 4, 5

$$d_{(23)1} = \min\{d_{21}, d_{31}\} = \min\{4, 7\} = 4$$

$$d_{(23)4} = \min\{d_{24}, d_{34}\} = \min\{7, 10\} = 7$$

$$d_{(23)5} = \min\{d_{25}, d_{35}\} = \min\{6, 5\} = 5$$

Dman	1	2	3	4	5
1	0	4	7	3	6
2	4	0	3	7	6
3	7	3	0	10	5
4	3	7	10	0	9
5	6	6	5	9	0

Langkah 3

- Perbarui matrik jarak antar data.
- Hasil matriks jarak penggabungan kelompok 2 dengan kelompok 3.

Dman	1	(23)	4	5
1	0	4	3	6
(23)	4	0	7	5
4	3	7	0	9
5	6	5	9	0

Langkah 4

- Gabungkan dua kelompok terdekat menjadi satu.

Dman	1	(23)	4	5
1	0	4	3	6
(23)	4	0	7	5
4	3	7	0	9
5	6	5	9	0

- Pilih jarak dua kelompok terdekat.

$$\min(D_{man}) = \min(d_{14}) = 3$$

- Terpilih kelompok 1 dan 4, sehingga kedua kelompok digabungkan.
- Menghitung jarak terdekat antar kelompok (1 dan 4) dengan kelompok yang tersisa, yaitu {23} dan 5

$$d_{(14)23} = \min\{d_{12}, d_{13}, d_{42}, d_{43}\} = \min\{4, 7, 7, 10\} = 4$$

$$d_{(14)5} = \min\{d_{15}, d_{45}\} = \min\{6, 9\} = 6$$

Langkah 5

- Perbarui matrik jarak antar data.
- Hasil matriks jarak penggabungan kelompok 1 dengan kelompok 4.

Dman	(1,4)	(2,3)	5
(1,4)	0	4	6
(2,3)	4	0	5
5	6	5	0

Langkah 6

- Gabungkan dua kelompok terdekat menjadi satu.

Dman	(1,4)	(2,3)	5
(1,4)	0	4	6
(2,3)	4	0	5
5	6	5	0

- Pilih jarak dua kelompok terdekat.

$$\min(D_{man}) = \min(d_{1234}) = 4$$

- Terpilih kelompok 1 dan 4, sehingga kedua kelompok digabungkan.
- Menghitung jarak terdekat antar kelompok (1 dan 4) dengan kelompok yang tersisa, yaitu {23} dan 5

$$d_{(1234)(5)} = \min\{d_{15}, d_{25}, d_{35}, d_{45}\} = \min\{6, 6, 5, 9\} = 5$$

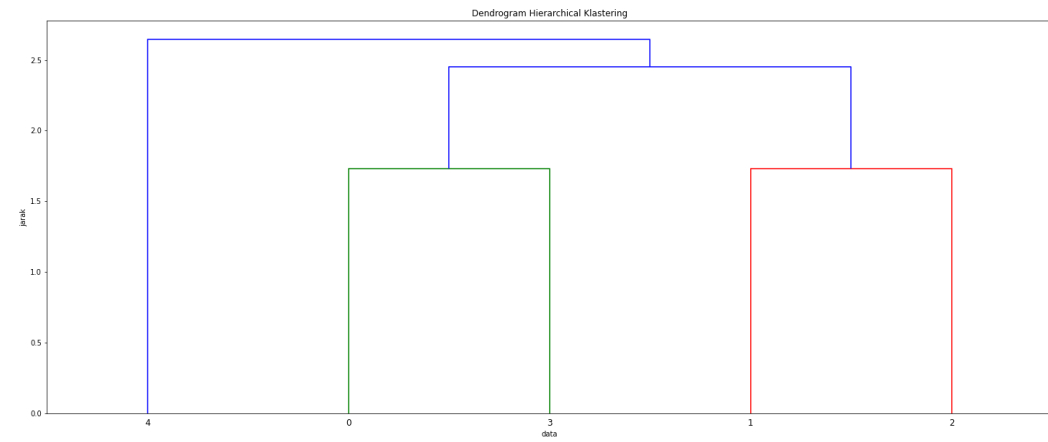
Langkah 7

- Perbarui matrik jarak antar data.
- Hasil matriks jarak penggabungan kelompok {23} dengan kelompok {14}.

Dman	(1234)	5
(1234)	0	5
5	5	0

- Kelompok (1234) dan 5 digabung menjadi kelompok tunggal dari lima data, yaitu kelompok (12345) dengan jarak terdekat dengan jarak terdekat 5.

- Hasil Klastering dengan AHC





Implementasi Python

(Agglomerative Hierarchical Clustering)

Pertemuan 11

Implementasi Python

```
# import library
from matplotlib import pyplot as plt
from scipy.cluster.hierarchy import dendrogram, linkage
import numpy as np

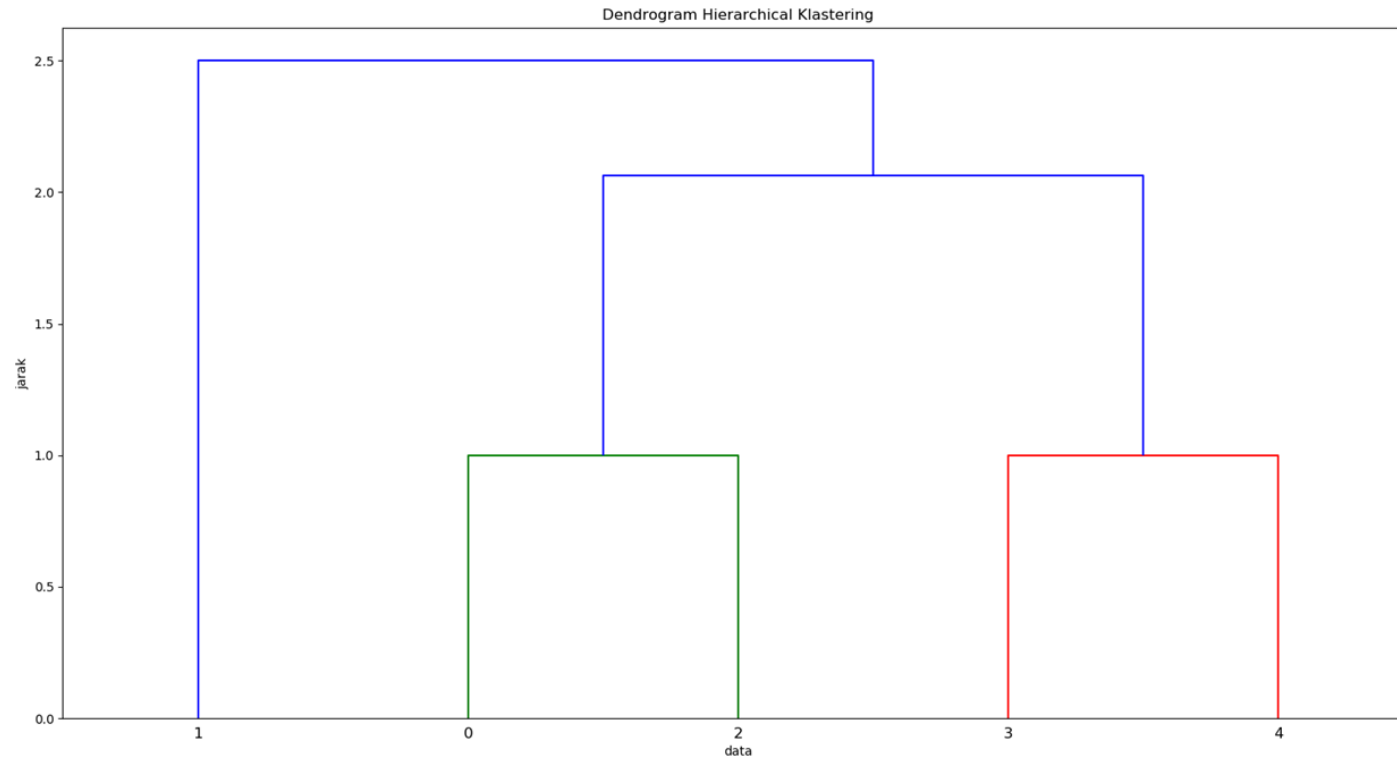
#buat array 2D ke dalam variabel x. Data ini yang akan kita klaster.
x = np.array([[1,2],[2.5,4.5],[2,2],[4,1.5],[4,2.5]])

#lakukan hierachical klastering
Z = linkage(x)

#plot dendogram
plt.figure(figsize=(25, 10))
plt.title('Dendrogram Hierarchical Klastering')
plt.xlabel('data')
plt.ylabel('jarak')
dendrogram(Z)
plt.show()
```

- Langkah pertama import library yang diperlukan.
- Kemudian buat array 2D kedalam variable x, data ini yang akan diklaster.
- Lakukan hierarchical klastering.
- Visualisasikan plot Dendrogram.

Implementasi Python



- Hasil visualisasi kluster seperti gambar dendrogram disamping.

Latihan Soal (Kuis)

- Kelompokkan dataset di bawah ini dengan menggunakan metode AHC :
 1. Complete Linkage
 2. Average Linkage
- Menggunakan Manhattan Distance

Dataset Skor Kedisiplinan Siswa

Siswa	DTW	DTT	DMT	DDB
1	2	4	4	3
2	3	4	3	5
3	4	3	2	5
4	1	5	4	2
5	3	2	1	3

Kriteria Kedisiplinan		
1	Disiplin terhadap Waktu (DTW)	(1-5)
2	Disiplin terhadap tata terib (DTT)	(1-5)
3	Disiplin mengerjakan tugas (DMT)	(1-5)
4	Disiplin dalam berpakaian(DDB)	(1-5)

Referensi

1. Kusrini, Taufiq Emha, Algoritma Data Mining, *Penerbit Andi*, 2009.
2. Ian H. Witten, Frank Eibe, Mark A. Hall, Data mining: Practical Machine Learning Tools and Techniques 4th Edition, *Elsevier*, 2017.
3. Budi Santosa, Ardian Umam, Data Mining dan Big Data Analytics, Penebar Media Pustaka, 2018.
4. Yaya Heryadi, Teguh Wahyono, Machine Learning: Konsep dan Implementasi, Penerbit Gava Media, 2020.
5. Sumber gambar: www.freepik.com.



THANKS

ANY QUESTIONS?

