**Algoritma Hierarchical Clustering**

by [informatikalogi](https://informatikalogi.com/author/cskaditya/) · Published July 6, 2017 · Updated July 6, 2017

Daftar Isi “**Algoritma Hierarchical Clustering**“

[1. Contoh Soal Perhitungan](https://informatikalogi.com/algoritma-hierarchical-clustering/#1)  
[2. Metode Single Linkage](https://informatikalogi.com/algoritma-hierarchical-clustering/#2)  
[3. Metode Complete Linkage](https://informatikalogi.com/algoritma-hierarchical-clustering/#3)  
[4. Metode Average Linkage](https://informatikalogi.com/algoritma-hierarchical-clustering/#4)  
[5. Latihan Soal](https://informatikalogi.com/algoritma-hierarchical-clustering/#5)

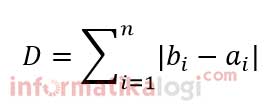
Hierarchical Clustering adalah metode analisis kelompok yang berusaha untuk membangun sebuah hirarki kelompok data.

Strategi pengelompokannya umumnya ada 2 jenis yaitu **Agglomerative (Bottom-Up)** dan **Devisive (Top-Down)**.

**Langkah Algoritma Agglomerative Hierarchical Clustering :**

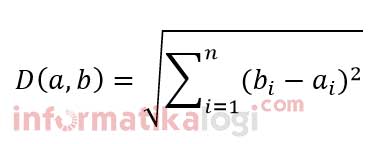
1. Hitung Matrik Jarak antar data.
2. Gabungkan dua kelompok terdekat berdasarkan parameter kedekatan yang ditentukan.
3. Perbarui Matrik Jarak antar data untuk merepresentasikan kedekatan diantara kelompok baru dan kelompok yang masih tersisa.
4. Ulangi langkah 2 dan 3 higga hanya satu kelompok yang tersisa.

Membentuk Matrik Jarak, misal dengan **Manhattan Distance** :



Persamaan Manhattan Distance

atau menggunakan **Euclidian Distance** :



Persamaan Euclidean Distance

Beberapa metode Pengelompokan **Agglomerative Hierarchical** :

**a. Single Linkage (Jarak Terdekat)**

single-linkage

single-linkage

**b. Complete Linkage (Jarak Terjauh)**

complete-linkage

complete-linkage

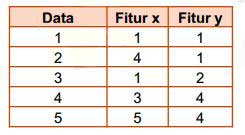
**c. Average Linkage (Jarak Rata-Rata)**

average-linkage

average-linkage

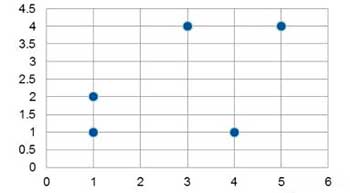
**1. Contoh Soal Perhitungan**

Perhatikan dataset berikut :



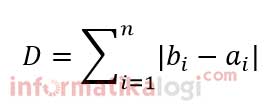
dataset-hierarchical-clustering

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



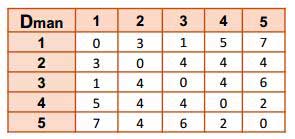
dataset-visual-hierarchical-clustering

Menghitung Jarak Pada Semua Pasangan dua data :



Persamaan Manhattan Distance

D*man* (*Data1*, *Data1*) = |1-1| + |1-1| = 0  
D*man* (*Data1*, *Data2*) = |1-4| + |1-1| = 3  
D*man* (*Data1*, *Data3*) = |1-1| + |1-2| = 1  
D*man* (*Data1*, *Data4*) = |1-3| + |1-4| = 5  
D*man* (*Data1*, *Data5*) = |1-5| + |1-4| = 7  
D*man* (*Data2,* *Data3*) = |4-1| + |1-2| = 4  
D*man* (*Data2*, *Data4*) = |4-3| + |1-4| = 4  
D*man* (*Data2*, *Data5*) = |4-5| + |1-4| = 4  
D*man* (*Data3*, *Data4*) = |1-3| + |2-4| = 4  
D*man* (*Data3*, *Data5*) = |1-5| + |2-4| = 6  
D*man* (*Data4*, *Data5*) = |3-5| + |4-4| = 2



Dman-hierarchical-clustering

**2. Metode Single Linkage**

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

min(D*man*) = min(d*13*) = 1

Terpilih kelompok 1 dan 3, sehingga kedua kelompok ini digabungkan.

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

d(13)2 = min {d12, d32} = min {3,4} = 3

d(13)4 = min {d14, d34} = min {5,4} = 4

[Baca Juga :  Contoh Sederhana Koefisien Jaccard Pada Dokumen Teks](https://informatikalogi.com/koefisien-jaccard-dokumen-teks/" \t "_blank)

d(13)5 = min {d15, d35} = min {7,6} = 6

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-hierarchical-clustering-(2)

Selanjutnya dipilih jarak dua kelompok yang terkecil.

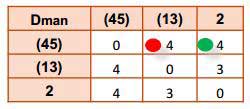
min(D*man*) = min(d*45*) = 2

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

d(45)(13) = min {d41, d43 , d51, d53}  = min {5,4,7,6} = 4

d(45)2 = min {d42, d52} = min {4,4} = 4

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



Dman-hierarchical-clustering-(3)

Selanjutnya dipilih jarak dua kelompok yang terkecil.

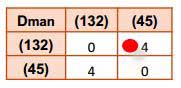
min(D*man*) = min(d(13)2) = 3

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).

d(132)(45) = min {d14, d15 , d34, d35 , d24, d25}  = min {5,7,4,6,4,4} = 4

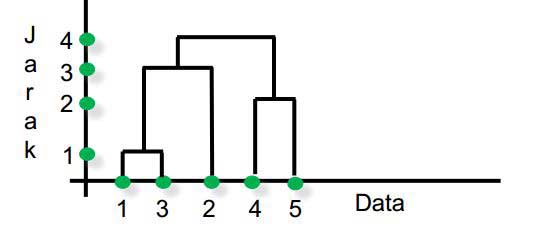
Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok (13) dan 2, serta menambahkan baris dan kolom untuk kelompok (123).



Dman-hierarchical-clustering-(4)

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

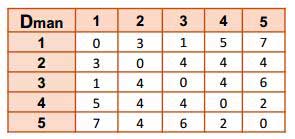
Berikut Dendogram Hasil Metode Single Linkage :



dendogram-single-linkage

**3. Metode Complete Linkage**

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



Dman-hierarchical-clustering

min(D*man*) = min(d13) = 1

Terpilih kelompok 1 dan 3, sehingga kedua kelompok ini digabungkan.

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

d(13)2 = max {d12, d32} = max {3,4} = 4

d(13)4 = max {d14, d34} = max {5,4} = 5

d(13)5 = max {d15, d35} = max {7,6} = 7

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-hierarchical-clustering-complete

Selanjutnya dipilih jarak dua kelompok yang terkecil.

[Baca Juga :  Algoritma C4.5](https://informatikalogi.com/algoritma-c4-5/" \t "_blank)

min(D*man*) = min(d45) = 2

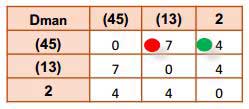
Dengan menghapus baris-baris dan kolom-kolom matrik jarak yang bersesuaian dengan kelompok 1 dan 3, serta menambahkan baris dan kolom untuk kelompok (13).

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

d(45)(13) = max {d41, d43 , d51, d53}  = max {5,4,7,6} = 7

d(45)2 = max {d42, d52} = max {4,4} = 4

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



Dman-hierarchical-clustering-complete-(2)

Selanjutnya dipilih jarak dua kelompok yang terkecil.

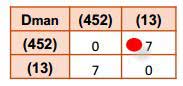
min(D*man*) = min(d(45)2) = 4

Terpilih kelompok (45) dan 2, sehingga kedua kelompok ini digabungkan.

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

d(452)(13) = min {d41, d43 , d51, d53 , d21, d23}  = max {5,4,7,6,3,4} = 7

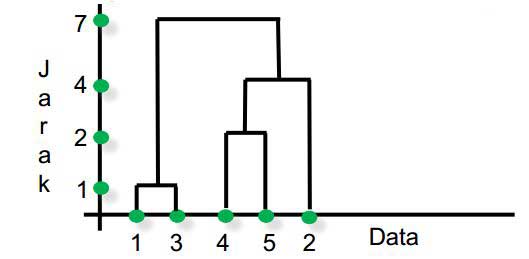
Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok (45) dan 2, serta menambahkan baris dan kolom untuk kelompok (452).



Dman-hierarchical-clustering-complete-(3)

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

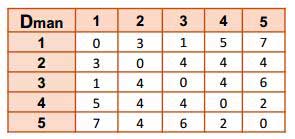
Berikut Dendogram Hasil Metode Complete Linkage :



dendogram-complete-linkage

**4. Metode Average Linkage**

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



Dman-hierarchical-clustering

min(D*man*) = min(d13) = 1

Terpilih kelompok 1 dan 3, sehingga kedua kelompok ini digabungkan.

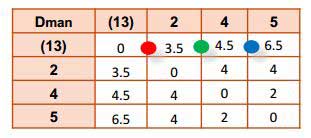
Menghitung jarak antar kelompok (1 dan 3) dengan kelompok lain yang tersisa, yaitu 2, 4 dan 5.

d(13)2 = average {d12, d32} = average {3,4} = (3+4) / 2 = 3.5

d(13)4 = average {d14, d34} = average {5,4} = (5+4) / 2 = 4.5

d(13)5 = average {d15, d35} = average {7,6} = (7+6) / 2 = 6.5

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-hierarchical-clustering-average

Selanjutnya dipilih jarak dua kelompok yang terkecil.

[Baca Juga :  Algoritma Apriori (Association Rule)](https://informatikalogi.com/algoritma-apriori-association-rule/" \t "_blank)

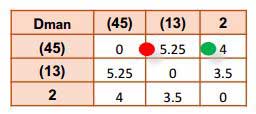
min(D*man*) = min(d45) = 2

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

d(45)(13) = average {d41, d43 , d51, d53}  = average {5,4,7,6} = (5+4+7+6) / 4 = 5.25

d(45)2 = average {d42, d52} = 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).



Dman-hierarchical-clustering-average-(2)

Selanjutnya dipilih jarak dua kelompok yang terkecil.

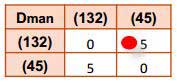
min(D*man*) = min(d(13)2) = 3.5

Terpilih kelompok (13) dan 2, sehingga kedua kelompok ini digabungkan.

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

d(452)(13) = average {d14, d15 , d34, d35 , d24, d25}  = average {5,7,74,6,4,4} = (5+7+4+6+4+4) / 6 = 5

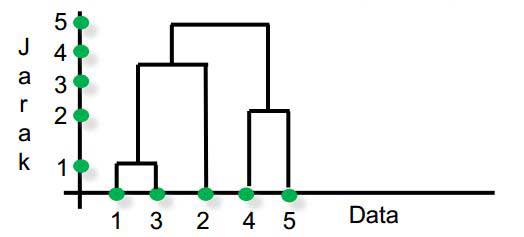
Menghapus baris dan kolom matrik yang bersesuaian dengan kelompok (45) dan 2, serta menambahkan baris dan kolom untuk kelompok (452).



Dman-hierarchical-clustering-average-(3)

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

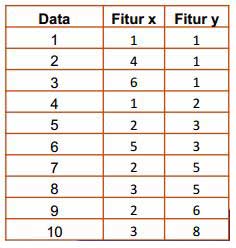
Berikut Dendogram Hasil Metode Average Linkage :



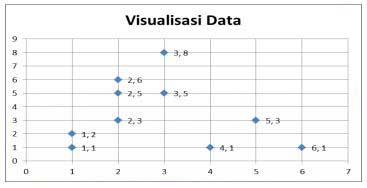
dendogram-average-linkage

**5. Latihan Soal**

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



latihan-soal-ahc-2



latihan-soal-ahc-2-2

Semoga artikel berjudul “**[Algoritma Hierarchical Clustering](https://informatikalogi.com/algoritma-hierarchical-clustering/" \t "_blank)**” bisa bermanfaat dan silahkan jika masih ada yang kurang jelas dapat ditanyakan di kolom komentar dibawah ini.

Silahkan **Like Fanspage** dan **Share** artikel ini jika menurut kamu bermanfaat untuk kamu dan orang lain.