**TUGAS 1**

*Hierarchical Clustering*

Nama/ NIM : Whisnumurty Galih Ananta

Mata Kuliah/ Kelas : Komputasi Cerdas/ IF-02

Dosen Pengampu : Ibu Vessa Rizky Oktavia, M.Kom.

Deskripsi Tugas : Menggunakan algoritma Hierarchical Clustering pada dataset bunga iris

Link Dataset: [Iris Species | Kaggle](https://www.kaggle.com/datasets/uciml/iris)

**Hierarchical Clustering**

**Hierarchical Clustering** adalah metode **pengelompokan** **data** yang bekerja dengan mengelompokkan data **ke** dalam **pohon** **klaster**. Proses Hierarchical Clustering dimulai **dengan** memperlakukan **setiap** **titik** **data** sebagai **klaster** **terpisah**.

Kemudian, langkah-langkah berikut dilakukan secara berulang:

1. Identifikasi 2 klaster yang paling mendekat satu sama lain.
2. Gabungkan 2 klaster yang paling mirip tersebut.
3. Teruskan langkah-langkah ini hingga semua klaster digabungkan menjadi satu.

Dalam Hierarchical Clustering, **tujuannya** adalah **menghasilkan** **serangkaian** **klaster** **bersarang** secara **hierarki**. **Diagram** yang disebut **Dendrogram** (sebuah diagram **berbentuk pohon** yang **menggambarkan** **urutan** **penggabungan** atau **pemisahan**) secara grafis mencerminkan hierarki ini dan adalah pohon terbalik yang menjelaskan urutan penggabungan faktor (dilihat dari bawah ke atas) atau pemecahan klaster (dilihat dari atas ke bawah).

**Kelebihan Hierarchical Clustering:**

* Kemampuan untuk mengatasi klaster non-convex serta klaster berukuran dan berdensitas berbeda.
* Kemampuan untuk mengatasi data yang hilang dan data yang berisik.
* Kemampuan untuk mengungkap struktur hierarki data, yang berguna untuk memahami hubungan antara klaster.

Hierarchical Clustering juga memiliki beberapa kelemahan:

* Memerlukan kriteria untuk menghentikan proses pengelompokan dan menentukan jumlah klaster akhir.
* Biaya komputasi dan kebutuhan memori yang tinggi, terutama untuk dataset besar.
* Hasilnya dapat sensitif terhadap kondisi awal, kriteria penggabungan, dan metrik jarak yang digunakan.

Selain itu, Hierarchical Clustering memiliki dua jenis utama:

1. **Agglomerative**: Proses dimulai dengan memperlakukan setiap titik data sebagai **klaster** **terpisah** dan kemudian secara berulang **menggabungkan** **klaster** yang paling **mendekat** **satu sama lain** hingga hanya **tersisa** **satu** **klaster**.

A diagram of a process

Description automatically generated

1. **Divisive**: Sebaliknya, dalam Divisive Hierarchical Clustering, semua titik data **awalnya** **dianggap** sebagai **satu klaster** dan **kemudian** dalam setiap iterasi, titik data **dipisahkan** dari klaster yang tidak cocok **hingga** akhirnya tersisa **N klaster**.

A diagram of a step

Description automatically generated

Code

Link: <https://github.com/whisnumurtyga/s5-komputasi-cerdas/tree/master/Tugas1>

Step By Step:

1. Import Library

A screen shot of a computer

Description automatically generated

1. Load Dataset

A screen shot of a computer code

Description automatically generated

* Karena dataset punya kolom id dan class/label jadi di select dulu kolomnya

1. Find Distance Matrix

A screen shot of a computer code

Description automatically generated

1. Execute

A screen shot of a computer code

Description automatically generated

* Linkage berdasarkan matriks jarak dengan metode ward
* Linkage adalah salah satu parameter dalam algoritma hierarki clustering, yang digunakan untuk mengukur jarak antara dua cluster saat melakukan penggabungan (merge) dalam proses pembentukan hierarki cluster.
* Dalam Ward linkage, jarak antara dua cluster diukur dengan cara yang kompleks yang berusaha untuk meminimalkan varians dalam cluster yang baru terbentuk ketika dua cluster digabungkan. Ini cenderung menghasilkan cluster yang berukuran seimbang.
* Sebanyak 3 hierarki karena mendefinisikan truncate\_mode = level dan p = 3

1. Result

A graph with numbers and lines

Description automatically generated