Analisis Kluster (Clustering) - Hierarchical Clustering

# Konsep Dasar:

Hierarchical Clustering adalah metode clustering yang membangun hierarki cluster. Ada dua jenis utama hierarchical clustering:

* **Agglomerative:** Mulai dengan setiap titik data sebagai cluster terpisah, lalu secara iteratif menggabungkan cluster terdekat hingga titik data berada dalam satu cluster.
* **Divisive:** Mulai dengan semua titik data dalam satu cluster, lalu secara iteratif membagi cluster menjadi sub-cluster lebih kecil hingga setiap titik data menjadi cluster terpisah.

**Contoh Kode (Scikit-Learn):**

from sklearn.cluster import AgglomerativeClustering

# ... (kode untuk membaca dan mempersiapkan dataset)

# Membuat model Agglomerative Clustering

agg\_clustering = AgglomerativeClustering(n\_clusters=3) # 3 cluster sebagai contoh

# Melatih model

labels = agg\_clustering.fit\_predict(X)

**Tugas:**

1. Terapkan Agglomerative Clustering pada dataset Iris dengan jumlah cluster (k) = 3.
2. Visualisasikan hasil clustering dengan dendrogram.
3. Evaluasi hasil clustering dengan menghitung Silhouette Score.

**Selamat Mengerjakan Tugas! 🙂**

**Tugas Hari 37:**

1. **Terapkan Agglomerative:**

**import numpy as np**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**from sklearn import datasets**

**from sklearn.cluster import AgglomerativeClustering**

**from sklearn.decomposition import PCA**

**from sklearn.metrics import silhouette\_score, adjusted\_rand\_score**

**from scipy.cluster.hierarchy import dendrogram, linkage**

**# ... (kode untuk membaca dan mempersiapkan dataset)**

**# Membuat model Agglomerative Clustering**

**agg\_clustering\_ir = AgglomerativeClustering(n\_clusters=3) # 3 cluster sebagai contoh**

**# Melatih model**

**labels = agg\_clustering\_ir.fit\_predict(X)**

1. **Visualisasikan:**

**# Visualisasi Hasil Hierarchical Clustering dengan Dendrogram**

**plt.figure(figsize=(10, 7))**

**linked = linkage(X, 'ward')**

**dendrogram(linked, orientation='top', labels=y, distance\_sort='descending', show\_leaf\_counts=True)**

**plt.title('Dendrogram for the Iris Dataset')**

**plt.show()**

**# Visualisasi Hasil Hierarchical Clustering dengan PCA**

**pca = PCA(2)**

**X\_pca = pca.fit\_transform(X)**

**plt.figure(figsize=(8, 6))**

**plt.scatter(X\_pca[:, 0], X\_pca[:, 1], c=labels, cmap='viridis', marker='o', edgecolor='k', s=100)**

**plt.title('Hierarchical Clustering of Iris Dataset (2D PCA)')**

**plt.xlabel('Principal Component 1')**

**plt.ylabel('Principal Component 2')**

**plt.show()**

1. **Evaluasi:**

**# Evaluasi Hierarchical Clustering dengan Silhouette Score**

**silhouette\_avg = silhouette\_score(X, labels)**

**print(f'Silhouette Score: {silhouette\_avg}')**

**# Evaluasi Hierarchical Clustering dengan Adjusted Rand Index**

**ari = adjusted\_rand\_score(y, labels)**

**print(f'Adjusted Rand Index: {ari}')**