Analisis Kluster (Clustering) - K-Means

# Konsep Dasar:

Clustering adalah teknik unsupervised learning yang bertujuan untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kemiripannya. K-Means adalah salah satu algoritma clustering yang paling populer dan sederhana.

# Cara Kerja K-Means:

1. **Inisialisasi:** Secara acak memilih K titik sebagai pusat cluster awal.
2. **Penugasan:** Setiap titik data ditugaskan ke cluster terdekat berdasarkan jarak Euclidean (atau jarak lainnya).
3. **Pembaruan:** Pusat cluster diperbarui menjadi rata-rata dari semua titik data dalam cluster tersebut,
4. **Iterasi:** Langkah 2 dan 3 diulang sampai pusat cluster tidak berubah secara signifikan atau mencapai jumlah iterasi maksimum.

**Contoh Kode (Scikit Learn)**

from sklearn.cluster import KMeans

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

# Membuat model K-Means

kmeans = KMeans(n\_clusters=3, random\_state=0) # 3 cluster sebagai contoh

# Melatih model

kmeans.fit(X)

# Mendapatkan label cluster untuk setiap data point

labels = kmeans.labels\_

**Tugas:**

1. Terapkan K-Means pada dataset Iris dengan jumlah cluster (k) = 3.
2. Visualisasikan hasil clustering dengan scatter plot, dengan warna titik berbeda untuk setiap cluster.
3. Evaluasi hasil clustering dengan menghitung Silhouette Score.

**Selamat Mengerjakan Tugas! 😀**

**Tugas Hari 36:**

1. **Terapkan K-Means:**

**import numpy as np**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**from sklearn import datasets**

**from sklearn.cluster import KMeans**

**from sklearn.decomposition import PCA**

**from sklearn.metrics import silhouette\_score**

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

**# Membuat Model K-Means**

**kmeans\_ir = KMeans (n\_clusters=3, random\_state=0) # 3 Cluster Sebagai Rendah**

**# Melatih Model**

**kmeans\_ir.fit(X)**

**# Mendapatkan Label cluster untuk setiap data point**

**labels = kmeans\_ir.labels\_**

**# Evaluasi KMeans dengan Silhouette Score**

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

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

1. **Visualisasikan:**

**# Visualisasi Hasil KMeans**

**pca = PCA(2)**

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

**# Plot hasil KMeans clustering**

**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('KMeans Clustering of Iris Dataset (2D PCA)')**

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

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

**# Menandai centroids**

**centers = kmeans\_ir.cluster\_centers\_**

**centers\_pca = pca.transform(centers)**

**plt.scatter(centers\_pca[:, 0], centers\_pca[:, 1], c='red', s=200, alpha=0.75, marker='x')**

**plt.show()**

1. **Evaluasi:**

**# Evaluasi KMeans dengan menggunakan Adjusted Rand Index**

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

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

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