Reduksi Dimensi dengan PCA

# Konsep Dasar:

Principal Component Analysis (PCA) adalah teknik reduksi dimensi yang dapat digunakan untuk mengurangi jumlah fitur dalam dataset sambil mempertahankan sebanyak mungkin informasi. PCA mencari kombinasi linear dari fitur-fitur asli yang menangkap sebagian besar varians dalam data. Kombinasi linear ini disebut Principal Component (Komponen Utama).

**Contoh Kode (Scikit-Learn)**

from sklearn.decomposition import PCA

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

# Membuat model PCA

pca = PCA(n\_components=2) # 2 komponen utama sebagai contoh

# Melatih model

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

**Tugas:**

1. Terapkan PCA pada dataset “Breast Cancer Wisconsin”.
2. Visualisasikan hasil PCA dengan scatter plot, dengan warna titik berbeda untuk setiap kelas.
3. Coba beberapa nilai n\_components yang berbeda untuk melihat bagaimana hal tersebut mempengaruhi hasil visualisasi.

**Selamat Mengerjakan Tugas! 😀**

**Tugas:**

1. **Terapkan PCA:**
2. **Visualisasi hasil PCA:**
3. **Mencoba beberapa nilai n\_components:**

Kode lengkap Tugas Hari 40:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.decomposition import PCA

# Fungsi untuk menerapkan PCA dan memvisualisasikan hasilnya

def plot\_pca(X, y, n\_components):

pca = PCA(n\_components=n\_components)

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

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

if n\_components == 2:

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

plt.xlabel('Principal Component 1')

plt.ylabel('Principal Component 2')

elif n\_components == 3:

ax = plt.axes(projection='3d')

ax.scatter3D(X\_pca[:, 0], X\_pca[:, 1], X\_pca[:, 2], c=y, cmap='viridis', edgecolor='k', s=100)

ax.set\_xlabel('Principal Component 1')

ax.set\_ylabel('Principal Component 2')

ax.set\_zlabel('Principal Component 3')

else:

print("Hanya mendukung visualisasi untuk n\_components = 2 atau 3.")

plt.title(f'PCA with {n\_components} Components')

plt.show()

# Coba beberapa nilai n\_components yang berbeda

n\_components\_list = [2, 3]

for n\_components in n\_components\_list:

plot\_pca(X, y, n\_components)