

Nama : Rully Lukas T

NIM : 1103200181

Kelas : TK44G2

### 1. Principal Component Analysis (PCA)

PCA adalah teknik reduksi dimensi yang digunakan untuk mengidentifikasi komponen utama (principal component) dalam suatu dataset. Komponen utama ini merupakan kombinasi linear dari variabel-variabel asli yang menjelaskan sebagian besar variabilitas data.

Tujuan PCA:

- Mengurangi dimensi data tanpa kehilangan informasi penting.
- Meningkatkan interpretabilitas data dengan mengidentifikasi variabel-variabel yang paling penting.
- Mempermudah visualisasi data dengan memetakannya ke ruang dimensi yang lebih rendah.

### 2. Linear Discriminant Analysis (LDA)

LDA adalah teknik reduksi dimensi yang digunakan untuk memisahkan kelas-kelas data yang berbeda. LDA mencari kombinasi linear dari variabel-variabel asli yang memaksimalkan perbedaan antara kelas-kelas tersebut.

Tujuan LDA:

- Meningkatkan akurasi klasifikasi data.
- Mengurangi dimensi data untuk mempermudah proses klasifikasi.
- Memvisualisasi data dengan memetakannya ke ruang dimensi yang lebih rendah.

### 3. Singular Value Decomposition (SVD):

SVD adalah teknik dekomposisi matriks yang digunakan untuk menguraikan matriks data menjadi tiga matriks: matriks U, matriks S, dan matriks V. Matriks S berisi nilai-nilai singular yang mewakili variabilitas data.

Tujuan SVD:

- Mengurangi dimensi data dengan memilih nilai-nilai singular yang signifikan.
- Merekomendasikan produk kepada pengguna berdasarkan kesamaan preferensi.
- Mengidentifikasi pola dan tren dalam data teks.

### Pandas Framework

Pandas adalah library open-source berbasis Python yang powerful dan high-level untuk manipulasi dan analisis data. Pandas menyediakan struktur data yang fleksibel dan mudah

digunakan bernama DataFrame dan Series yang berperan penting dalam pengelolaan data.

#### Konsep Utama Pandas:

**Series:** Merupakan array data satu dimensi yang berlabel (labeled). Label ini seperti index yang memudahkan untuk mengakses data tertentu. Tipe data elemen dalam Series bisa homogen (misalnya, semua numerik atau semua string)

**DataFrame:** Merupakan struktur data dua dimensi yang tersusun dari kolom (column) dan baris (row). Tiap kolom dapat memiliki tipe data yang berbeda-beda. Secara umum,

DataFrame digunakan untuk menyimpan data tabular seperti hasil ekspor data dari spreadsheet atau database.

**Data Cleaning:** Pandas menyediakan fungsi untuk pembersihan data, seperti menangani missing values, duplikasi data, dan inkonsistensi format data.

**Data Analysis:** Pandas menawarkan berbagai fungsi untuk melakukan perhitungan statistik, seperti mean, median, standar deviasi, dan korelasi.

**Data Visualization:** Pandas memudahkan pembuatan visualisasi data dengan mengintegrasikan library Matplotlib dan Seaborn.

**Data Wrangling:** Pandas piawai dalam melakukan transformasi dan manipulasi data, seperti filtering, sorting, merging, joining, dan pivoting data.

#### Keunggulan Pandas:

1. Mudah dipelajari dan digunakan: Pandas memiliki syntax yang intuitif dan mudah dipahami, bahkan untuk pengguna yang baru belajar Python.
2. Performa yang cepat: Pandas dioptimasi untuk menangani data dalam jumlah besar secara efisien.
3. Fleksibel: Pandas dapat menangani berbagai tipe data dan struktur data yang kompleks.
4. Terintegrasi dengan library lain: Pandas dapat bekerja sama dengan library lain yang populer di dunia data science seperti NumPy, Matplotlib, Seaborn, dan scikit-learn.