

Nama : Nabilah Salwa

NIM : 1103204060

Analisis Data Eksploratif (EDA) adalah proses yang digunakan dalam ilmu data untuk mengeksplorasi dan menganalisis data sebelum melakukan analisis statistik lebih lanjut atau membuat model prediktif. Tujuan utama EDA adalah untuk memahami sifat dasar kumpulan data, mengidentifikasi pola, anomali, dan tren, serta menghasilkan hipotesis awal tentang data sebelum memasuki tahap analisis lebih dalam.


Exploratory Data Analysis

John Tukey (1961) EDA Focus on:


- Understand the data's **underlying structure**
- Develop **intuition** about the data set
- How the data were collected (to aid in **cleaning**)
- How to further **investigate** with statistical methods

EDA is Any Initial Investigations of Data:

1. **Basic:** Data Visualization and Numerical Summary
2. **Intermediate:** Statistical Hypothesis Tests with Confirmatory Analysis
3. **Advanced:** Unsupervised learning, PCA, SVD and Clustering



John Tukey, seorang ahli statistik terkenal, adalah salah satu tokoh terkemuka dalam pengembangan analisis data eksplorasi (EDA). EDA merupakan pendekatan statistik yang diperkenalkan oleh Tukey pada tahun 1970an yang berfokus pada pemahaman dan eksplorasi data sebelum melakukan analisis statistik formal. Tukey berpendapat bahwa EDA merupakan langkah penting dalam memahami data sebelum melanjutkan ke analisis lebih lanjut.



What we will learn


- **LDA: Linear Discriminant Analysis**

- **Approach:** Find a linear combination of features that characterizes or separates two or more classes of objects or events.

- **PCA: Principal Component Analysis**

- **Approach:** Uses an orthogonal transformation to convert bigger set of features into smaller set of linearly uncorrelated variables called principal components.

- **SVD: Singular Value Decomposition**

- **Approach:** Use matrix decomposition to find best projection axis with minimum reconstruction error.
- 

LDA (Linear Discriminant Analysis) adalah metode statistik untuk analisis multivariat. LDA digunakan untuk pemilihan fitur dan pemisahan kelas dalam data.. Ini adalah teknik yang digunakan dalam pembelajaran mesin untuk reduksi dimensi dan analisis statistik..

PCA adalah teknik reduksi dimensi yang digunakan untuk menganalisis dan mengurangi kompleksitas data.. Teknik ini biasa digunakan dalam analisis data multivariat dan pengolahan gambar.. Tujuan utama dari PCA adalah mengidentifikasi komponen-komponen utama (principal components) dalam data yang menjelaskan sebagian besar variasi dalam data, sehingga dimungkinkan untuk merepresentasikan data dalam dimensi yang lebih rendah daripada dimensi aslinya, tetapi tetap mempertahankan informasi yang signifikan..

SVD adalah singkatan dari "Singular Value Decomposition" (Pemisahan Nilai Tunggal) dalam matematika dan analisis data.. SVD adalah teknik dekomposisi matriks yang digunakan untuk berbagai tujuan dalam analisis data, pemrosesan sinyal, pengenalan pola, dan banyak aplikasi lainnya.. SVD adalah alat yang kuat dalam analisis data dan pemrosesan sinyal, dan sering digunakan dalam berbagai aplikasi ilmu data dan teknik lainnya