

Nama : Ilham Muhamad Firdaus

NIM : 1103202001

Tugas Minggu 5 Understanding EDA

Homework

- StatQuests ([PCA](#), [LDA](#), [SVD](#))
- Learn [PCA](#), [LDA](#) and [SVD](#) in ScikitLearn
- Learn Pandas Framework

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 ilmuwan statistik terkenal, dianggap sebagai salah satu tokoh utama dalam pengembangan EDA. Pada tahun 1977, beliau menerbitkan bukunya yang berjudul "Exploratory Data Analysis," di mana ia memperkenalkan berbagai teknik dan metode EDA, termasuk grafik dan visualisasi data.



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.
- 

PCA, LDA, dan SVD adalah tiga teknik yang digunakan dalam machine learning dan analisis data untuk mereduksi dimensi data atau mengekstraksi informasi penting dari data.

1. PCA (Principal Component Analysis)

Tujuannya adalah untuk menghilangkan varian yang kurang penting dalam data dan mempertahankan varian yang paling signifikan. Dengan menghilangkan beberapa komponen yang kurang signifikan, kita dapat mengurangi dimensi data.

2. LDA (Linear Discriminant Analysis)

Tujuannya adalah untuk menemukan kombinasi linear dari fitur yang memaksimalkan jarak antara kelas-kelas yang berbeda dan meminimalkan dispersi dalam setiap kelas. Dengan kata lain, LDA mencoba untuk menciptakan proyeksi yang membuat kelas-kelas terpisah dengan baik.

3. SVD (Singular Value Decomposition)

SVD dapat digunakan dalam berbagai konteks analisis data, termasuk dalam reduksi dimensi dan pemisahan sinyal dari noise. Kita dapat menggunakan komponen singular values tertentu untuk merepresentasikan data asli dengan dimensi yang lebih rendah.

Ketiganya merupakan teknik yang penting dalam analisis data dan machine learning, dan pemilihan teknik tergantung pada tujuan analisis Anda. PCA dan LDA terutama digunakan untuk mereduksi dimensi, sementara SVD adalah teknik dekomposisi matriks yang lebih umum digunakan untuk berbagai aplikasi analisis data.