

Nama : Rheza Ilham Firmansyah
NIM : 1103204033

Tugas Minggu 5 Understanding EDA

Homework

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

Analisis Data Eksploratif (EDA)

EDA atau Analisis Data Eksploratif adalah langkah penting dalam proses analisis data. Ini memberikan mekanisme untuk memeriksa data secara visual dan kuantitatif, mengungkap pola, mengidentifikasi anomali atau outlier, menguji asumsi, dan mengembangkan model awal. Tujuan utama dari EDA adalah untuk membantu analis memahami sifat data dan membuat keputusan yang tepat selama analisis atau pemodelan berikutnya.

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





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 (Principal Component Analysis) yaitu Teknik reduksi dimensi yang mengubah fitur asli menjadi set fitur baru yang ortogonal dan mencakup sebagian besar variasi dalam data.

LDA (Linear Discriminant Analysis) yaitu Metode klasifikasi yang juga dapat digunakan untuk reduksi dimensi dengan tujuan memaksimalkan perbedaan antar kelas.

SVD (Singular Value Decomposition) yaitu Faktorisasi matriks yang memecah sebuah matriks menjadi tiga matriks lainnya, sering digunakan dalam sistem rekomendasi dan reduksi dimensi.