Exploratory Data Analysis

Data Science Club Universitas Trunojoyo Madura Pertemuan ke Enam

09 Januari 2021

What is Exploratory Data Analysis?

Exploratory Data Analysis or (EDA) is understanding the data sets by summarizing their main characteristics often plotting them visually. This step is very important especially when we arrive at modeling the data in order to apply Machine learning. Plotting in EDA consists of Histograms, Box plot, Scatter plot and many more.

Descriptive Statistics

- **Purpose:**
  - Knowing whether there is a missing value or not
  - Find out whether there are outliers or not
  - Find out if there are differents format

- **Visualization:**
  - Know the characteristics/data pattern
  - Knowing whether there are outliers or not using plots
  - Knowing the distribution of data

Benefits

- Give insight into dataset
- Understand the underlying structure
- Extract important parameters and relationships that hold between them
- Test underlying assumptions