

Nama : Rully Lukas T

NIM : 1103200181

Kelas : TK44G02

Exploratory Data Analysis (EDA) adalah proses awal yang penting dalam analisis data. Ini merupakan investigasi awal terhadap kumpulan data untuk memahami karakteristik, struktur, dan pola tersembunyi di dalamnya.

Tujuan utama EDA :

1. Mengenali pola dan tren: EDA membantu mengidentifikasi pola dan tren yang ada dalam data. Ini dapat berupa hubungan antara variabel, distribusi data, dan keberadaan anomali (outlier).
2. Memahami struktur data: EDA membantu memahami struktur data, seperti tipe data (numerik, kategorikal, teks), missing values, dan adanya inkonsistensi.
3. Mempersiapkan data untuk analisis selanjutnya: EDA membantu dalam persiapan data untuk analisis selanjutnya. Ini dapat meliputi pembersihan data, transformasi data, dan pemilihan fitur yang relevan.

Langkah-langkah yang dapat dilakukan dalam EDA untuk dataset house pricing :

1. Memahami Data :

- Membaca deskripsi dataset dan variabel.
- Memeriksa dimensi dataset (jumlah baris dan kolom).
- Meringkas statistik deskriptif seperti mean, median, standar deviasi, minimum, dan maksimum untuk setiap variabel.
- Mengecek missing values dan outliers.

2. Analisis Univariat :

- Memvisualisasi distribusi setiap variabel numerik dengan histogram, boxplot, dan density plot.
- Menganalisis distribusi variabel kategorikal dengan bar chart dan pie chart.
- Menghitung korelasi antara variabel numerik.

3. Analisis Bivariat :

- Menganalisis hubungan antara variabel harga rumah dengan variabel lain seperti luas rumah, jumlah kamar tidur, lokasi, dan tahun pembangunan.
- Memvisualisasi hubungan bivariat dengan scatter plot, boxplot per kategori, dan heatmap.

4. Analisis Multivariat :

- Menganalisis hubungan antara beberapa variabel secara simultan dengan menggunakan teknik seperti regresi linear, regresi pohon, dan random forest.
- Memilih model terbaik untuk memprediksi harga rumah.