

Nama : Zaky Ibnu Kusumah
NIM : 1103204213
Kelas : TK4404

Analisis EDA untuk Dataset House Pricing

Analisis EDA (Exploratory Data Analysis) adalah langkah awal yang penting dalam pemahaman dan persiapan dataset sebelum membangun model prediksi harga rumah atau tugas lainnya. Berikut adalah contoh analisis EDA yang dapat Anda lakukan untuk dataset harga rumah

1. Memahami Data Awal:

Baca dataset harga rumah dan tampilkan beberapa baris pertama data. Lihat info dataset untuk memeriksa jenis data, jumlah entri, dan apakah ada nilai-nilai yang hilang. Hitung statistik deskriptif untuk variabel numerik seperti harga, luas tanah, jumlah kamar tidur, dll.

2. Visualisasi Data:

Buat histogram atau plot distribusi untuk variabel harga. Buat scatter plot untuk memahami hubungan antara harga dengan variabel numerik seperti luas tanah, jumlah kamar tidur, dll. Buat box plot untuk menilai distribusi harga berdasarkan kriteria tertentu seperti tipe rumah, lokasi, dll. Gunakan peta panas korelasi untuk melihat korelasi antara variabel numerik.

3. Penanganan Nilai Hilang:

Identifikasi variabel dengan nilai-nilai yang hilang. Tentukan strategi penanganan nilai-nilai yang hilang, seperti pengisian dengan mean/median/modus atau penghapusan data yang hilang.

4. Pembersihan Data:

Deteksi dan penanganan outlier jika ada (misalnya, menggunakan metode IQR). Encode variabel kategori jika diperlukan, seperti mengubah kategori menjadi angka.

5. Analisis Kategori:

Buat grafik batang untuk variabel kategori seperti tipe rumah, lokasi, dan lain-lain. Hitung distribusi frekuensi untuk kategori-kategori ini.

6. Analisis Waktu (Jika Ada):

Jika dataset berisi data waktu, buat plot deret waktu untuk memahami tren dan pola.

7. Analisis Korelasi:

Hitung korelasi antara variabel-variabel numerik dan harga. Identifikasi variabel yang paling berpengaruh dalam memprediksi harga rumah.

8. Analisis Multivariat:

Gunakan teknik seperti analisis komponen utama (PCA) untuk mengurangi dimensi jika dataset memiliki banyak fitur numerik.

9. Analisis Distribusi Harga pada Subkelompok:

Pisahkan data menjadi subkelompok berdasarkan kriteria tertentu (misalnya, tipe rumah) dan analisis distribusi harga dalam setiap subkelompok.

10. Ringkasan Statistik Akhir:

Buat ringkasan akhir dari hasil analisis EDA dan tentukan apakah ada langkah tambahan yang perlu diambil sebelum memulai pemodelan.