EXPLORATORY DATA ANALYSIS (EDA) PADA DATA REAL ESTATE UNTUK PREDIKSI HARGA RUMAH

PRAKTIKUM PEMROSESAN DATA

Oleh

Nama

NIM

(Program Studi Informatika)

A black background with a black square

Description automatically generated with medium confidence

UNIVERSITAS TRISAKTI  
September 2024

I. Pendahuluan

Pada praktikum ini akan dilakukan pembelajaran mengenai fundamental pada pemrosesan data dan *exploratory data analysis* (EDA). Hal ini berfungsi untuk mengetahui tindakan yang harus dilakukan pada saat diberikan sebuah *dataset*, baik untuk membersihkan data atau melakukan transformasi/*scaling* terhadap data yang diberikan.

Pada praktikum ini dilakukan untuk menjawab pertanyaan:

1. Meng-*load* data yang diberikan pada komputer.
2. Melakukan pengelompokkan data yang diberikan merupakan Fitur atau target.
3. Melakukan pembersihan dan penanganan data jika diperlukan, seperti penanganan *missing value*, penanganan data *outlier*, penanganan data tidak konsisten, dsb.

II. Hasil Praktikum

II.1 *Import Library*

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.2 Menyiapkan *Dataset*

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan |

II.3 Mengubah Atribut *Transaction Date* menjadi *integer*

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan |

II.4 Mengecek dengan fungsi pandas.DataFrame.head()

|  |
| --- |
|  |

II.5 Menampilkan nama kolom pada dataset

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.6 Mengubah nama atribut menjadi nama atribut baru pada *dataframe*

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.7 Melihat Jumlah data unik dari setiap atribut

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.8 Mengecek data *duplicate*

|  |
| --- |
|  |

|  |
| --- |
|  |

|  |
| --- |
| **Penjelasan:** |

II.9 Melihat tipe data numerik pada kolom setiap atribut

|  |
| --- |
|  |

|  |
| --- |
| **Penjelasan:** |

II.10 Menampilkan gambar *boxplot* dari setiap atribut yang digunakan

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.11 Membuang data *outlier* berdasarkan gambar *boxplot*

[Menggunakan sintaks untuk membuang data *outlier*, sertakan screenshot dan penjelasan mengapa atribut tersebut yang dipilih.]

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.12 Melakukan cek linearitas data atribut dengan data target menggunakan *ScatterPlot*

Data sebelum dibuang outliernya :

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

Data setelah sudah dibuang outliernya :

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.13 Melakukan filtering berdasarkan kondisi pada kolom

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.14 Menampilkan *heatmap* dari data

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.15 Melakukan transformasi data dengan fungsi logaritma

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.16 Pengecekan nilai korelasi antar prediktor dan antara prediktor dengan target

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.17 Pengecekan Nilai Null

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.18 Korelasi dengan HeatMap

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.19 Menentukan Kolom Prediktor dan Target

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.20 Menunjukkan perbedaan regresi linar dengan data orisinal dan data hasil transformasi

[Menggunakan fungsi LinearRegression() dan r2\_score() dari library sklearn, sertakan screenshot serta beri penjelasan mengenenai perbedaan nilai r2\_score().]

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.21 Membuat histogram dari beberapa kolom dalam df\_4

|  |
| --- |
|  |

II.22 Membuat scatter plot (diagram sebaran)

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.23 Membuat *regression plot*

|  |
| --- |
|  |

|  |
| --- |
| Penjelasan: |

II.24 Kesimpulan ( minimal 5 baris )

|  |
| --- |
|  |