

Analisis EDA House Pricing Dataset

1. Persiapan Data:

Langkah pertama adalah memuat dataset ('kc_house_data.csv') ke dalam DataFrame Pandas dan melakukan eksplorasi awal data. Kode ini menyoroti anomali pada atribut 'id' dan melakukan pengecekan nilai yang kosong. Selanjutnya, 'date' diubah menjadi 'month' dan 'year', dan kolom yang tidak diperlukan seperti 'id' dan 'date' dihapus. Nol pada atribut 'bedrooms' dan 'bathrooms' diperiksa, dan baris yang memiliki keduanya bernilai nol dihapus.

2. Analisis Data Eksploratif (EDA):

Skrip ini melakukan analisis EDA baik secara tunggal maupun bersama-sama. EDA tunggal melibatkan visualisasi fitur numerik menggunakan histogram, sedangkan EDA bersama-sama menggunakan pair plot dan peta panas matriks korelasi untuk memahami hubungan antar fitur numerik.

3. Pemisahan dan Standarisasi Data:

Dataset dibagi menjadi set pelatihan dan pengujian. Fitur numerik distandarisasi menggunakan StandardScaler. Rata-rata dan deviasi standar dari set pelatihan digunakan untuk menormalkan kedua set tersebut.

4. Pelatihan dan Evaluasi Model:

Tiga model regresi dilatih dan dievaluasi: Decision Tree Regressor (DTR), Random Forest Regressor (RFR), dan KNeighbors Regressor (KNR). Kode ini menghitung skor R-kuadrat untuk masing-masing model pada set pelatihan. Random Forest Regressor dikonfigurasi dengan 500 estimator.

5. Pengisian dan Penghitungan Root Mean Squared Error (RMSE):

Kode ini menggunakan SimpleImputer untuk menangani nilai yang hilang. Selanjutnya, dihitung nilai RMSE untuk masing-masing model pada set pelatihan dan pengujian. Hasil RMSE disimpan dalam DataFrame untuk analisis lebih lanjut.

6. Visualisasi:

Nilai-nilai RMSE untuk masing-masing model pada set pelatihan dan pengujian divisualisasikan menggunakan diagram batang horizontal.

7. Prediksi Model:

Program ini membuat prediksi pada sebuah contoh uji untuk masing-masing model dan menampilkan nilai sebenarnya bersama dengan nilai yang diprediksi.