Housing Price Prediction using XGBoost

Overview

This project involves predicting housing prices using the housing.csv dataset and the XGBoost machine learning model. The project includes data preprocessing, exploratory data analysis, hyperparameter tuning, and model evaluation, with a comparison of results before and after removing outliers.

How to Run the Notebook

1. **Open the Jupyter Notebook:**
    - The main notebook file is HOUSING_PRICE_PREDICTION.ipynb. This notebook contains all the necessary code for loading the data, preprocessing, training the XGBoost model, and evaluating the results.
2. **Upload the Dataset:**
    - Upload the housing.csv file into the Colab environment.
    - You can do this by clicking the "Files" tab on the left-hand side of the Colab interface and uploading the housing.csv file from your local machine.
3. **Run the Entire Notebook:**
    - Once the CSV file is uploaded, go to the Colab menu:
        - **Click on "Runtime" → "Run all"** to execute all the cells in the notebook.
    - The notebook will automatically read the CSV file, preprocess the data, train the model, and display results.

Dependencies

All required libraries (e.g., xgboost, sklearn, pandas, matplotlib, seaborn) are pre-installed in Google Colab, so **no installation is required**.

Summary

- The notebook compares model performance with and without outliers, and hyperparameters are optimized for the XGBoost model using grid search.
- The project concludes that keeping the outliers in this dataset yields better performance due to the valuable information carried by high-value data points.

Instructions for Use

1. Make sure to upload the housing.csv dataset.
2. Run the entire notebook from the "Runtime" menu.
3. Review the results and analysis in the output cells to understand how the model was trained and evaluated.