House Price Prediction

The "House Price Prediction" project focuses on predicting housing prices using machine learning techniques. By leveraging popular Python libraries such as NumPy, Pandas, Scikit-learn (sklearn), Matplotlib, Seaborn, and XGBoost, this project provides an end-to-end solution for accurate price estimation.

Project Overview

The "House Price Prediction" project aims to develop a model that can accurately predict housing prices based on various features. This prediction task is of great significance in real estate and finance, enabling informed decision-making for buyers, sellers, and investors. By employing machine learning algorithms and a curated dataset, this project provides a powerful tool for estimating house prices.

Key Features

- Data Collection and Processing: The project utilizes the "California Housing" dataset, which can be directly downloaded from the Scikit-learn library. The dataset contains features such as house age, number of rooms, population, and median income. Using Pandas, the data is processed and transformed to ensure it is suitable for analysis.
- Data Visualization: The project employs data visualization techniques to gain insights into the dataset. Matplotlib and Seaborn are utilized to create visualizations such as histograms, scatter plots, and correlation matrices. These visualizations provide a deeper understanding of the relationships between features and help identify trends and patterns.
- Train-Test Split: To evaluate the performance of the regression model, the project employs the train-test split technique. The dataset is split into training and testing subsets, ensuring that the model is trained on a portion of the data and evaluated on unseen data. This allows for an accurate assessment of the model's predictive capabilities.
- Regression Model using XGBoost: The project utilizes the XGBoost algorithm, a popular gradient boosting framework, to build the regression model. XGBoost is known for its ability to handle complex relationships between features and achieve high predictive accuracy. The Scikit-learn library provides an implementation of XGBoost that is utilized in this project.
- Model Evaluation: The project assesses the performance of the regression model using evaluation metrics such as R-squared error and mean absolute error. R-squared error measures the proportion of the variance in the target variable that can be explained by the model, while mean absolute error

quantifies the average difference between the predicted and actual house prices. These metrics provide insights into the model's accuracy and precision. Additionally, a scatter plot is created to visualize the predicted prices against the actual prices.

Getting Started

To run this project locally, follow these steps:

1. Clone the repository: `gh repo clone MYoussef885/House_Price_Prediction`
2. Install the required libraries: `If you're using Google Colab, you don't need to pip install. Just follow the importing the dependencies section.`
3. Launch Google Colab: `https://colab.research.google.com/`
4. Open the `House_Price_Prediction.ipynb` file and run the notebook cells sequentially.

Conclusion

The "House Price Prediction" project provides a practical solution for estimating housing prices based on various features. By leveraging data collection, preprocessing, visualization, XGBoost regression modeling, and model evaluation, this project offers a comprehensive approach to addressing the price prediction task. The project utilizes the "California Housing" dataset from Scikit-learn, ensuring a reliable and widely accessible data source.