House Price Prediction using Linear Regression

Overview

This project demonstrates how to predict house prices using **Linear Regression**. By leveraging features like the number of bedrooms, square footage, and more, the model aims to make accurate price predictions.

The project walks through data exploration, training a linear regression model, evaluating its performance, and visualizing the results.

Project Structure

```
house-price-prediction/
│
├── house_prices.csv          # The dataset used for model training
├── house_price_prediction.py   # Main script for training and evaluating the model
├── requirements.txt          # Required packages and libraries
├── README.md                  # Project documentation
└── LICENSE                    # License for the project
```

Dataset

The dataset consists of 4,600 records with the following features:

- date: Date of the sale

- price: Sale price of the house (target variable)

- bedrooms: Number of bedrooms

- bathrooms: Number of bathrooms

- sqft_living: Square footage of living space

- sqft_lot: Square footage of the lot

- floors: Number of floors

- waterfront: Indicates if the property is waterfront (1) or not (0)

- view: Rating of the house's view (0-4)

- condition: Condition of the house (1-5)

- sqft_above: Square footage of the house above the basement

- sqft_basement: Square footage of the basement

- yr_built: Year the house was built

- yr_renovated: Year the house was renovated

Requirements

- Python 3.x

- Pandas

- NumPy

- Scikit-Learn

- Matplotlib

- Seaborn

To install the required dependencies, use:

pip install -r requirements.txt

Installation

1. Clone the repository:

   git clone https://github.com/aswinrajkt586/House-Price-Prediction.git

2. Navigate to the project directory:

    cd House-Price-Prediction

3. Install the necessary dependencies:

    pip install -r requirements.txt

Usage

To train and evaluate the linear regression model, run the following:

house_price_prediction.py

The script will:

- Load and preprocess the dataset.

- Train the Linear Regression model.

- Evaluate model performance on test data.

- Output Mean Squared Error (MSE) and (R^2).

Model Evaluation

- Mean Squared Error (MSE): Evaluates the average squared difference between predicted and actual values.

- R^2 score: Represents how much variance in the target variable (house prices) is explained by the model.

Results

The linear regression model achieved:

- MSE: 111408.9

- R^2 score: 47.8

Future Work

Potential improvements:

- Implementing advanced models like Random Forest or Gradient Boosting.

- Feature engineering to improve model accuracy.

- Hyperparameter tuning for better performance.

License

This project is licensed under the MIT License. See the LICENSE file for more details.