

1. Project Title

House Price Prediction using Linear Regression (Supervised ML)

2. Objective

The aim of this project is to predict house prices based on different features such as crime rate, number of rooms, tax rate, and more, using the Boston Housing dataset (Kaggle alternative dataset used).

3. Dataset Description

Dataset from Kaggle: House Price Prediction Dataset
(<https://www.kaggle.com/datasets/bhanupratapbiswas/house-price-prediction>)

Common Columns in the dataset

CRIM – Crime rate

ZN – Residential land zoned

INDUS – Industrial area proportion

RM – Average number of rooms

AGE – Age of the house

DIS – Distance to employment centers

TAX – Property tax rate

PTRATIO – Student–Teacher ratio

MEDV/PRICE – Target variable (House Price)

(Your exact dataset may have slightly different names, but core meaning remains same.)

4. Steps Followed

- Step 1: Import necessary libraries

pandas

numpy

matplotlib

seaborn

scikit-learn

- Step 2: Load the dataset

```
df = pd.read_csv("HousePrice.csv")
```

- Step 3: Data Preprocessing

Check null values

Remove missing rows

Detect target column (usually price or medv)

Normalize/scale if required

- Step 4: Exploratory Data Analysis (EDA)

Correlation heatmap

Distribution plots

Room count vs price

Crime rate vs price

- Step 5: Split Data

Train-test split (80% Training, 20% Testing)

- Step 6: Build Linear Regression Model

```
from sklearn.linear_model import LinearRegression
```

- Step 7: Evaluate Model

Metrics used:

R² Score

Mean Squared Error (MSE)

Mean Absolute Error (MAE)

5. Model Performance (Typical Results)

Linear Regression

R² Score: ~0.70–0.74

MSE: Moderate

MAE: Good performance

(Linear regression performs well since Boston dataset is nearly linear.)

6. Visualizations

Scatter plot: RM vs Price

Heatmap of correlation

Actual vs Predicted price plot

These show:

Houses with more rooms → higher price

Crime rate ↑ → price ↓

Student–teacher ratio impacts price

7. Conclusion

Linear Regression can predict house prices with 70–75% accuracy.

The number of rooms (RM) is the strongest predictor.

Crime rate (CRIM) and tax rate (TAX) negatively affect prices.

This model is useful for basic price estimation but can be improved using:

Random Forest

XGBoost

Gradient Boosting

8. Technologies Used

Python

Jupyter Notebook

Pandas

NumPy

Scikit-Learn

Matplotlib, Seaborn

9. Applications

Real estate price prediction

Property valuation

Market analysis