Title: ***House Price Prediction Using Machine Learning***

Team Members

CH.VishwanathRao-2303A51095

G.V.N Hasini-2303A51099

E.Shivani-2303A51098

Course: Applications Of Data Mining.

Name of the instructor: E.Rama Krishna.

Submission Date: 24-04-2025.

House Price Prediction Using Machine Learning

Introduction

Accurate house price prediction is crucial for both buyers and sellers in the real estate market. This project utilizes machine learning to predict house prices based on various features like area, number of rooms, location, and more.

Objective

To build a regression model that accurately predicts house prices based on key property features.

Dataset Description

- Source: Kaggle - House Prices: Advanced Regression Techniques

- Features:

  - LotArea

  - OverallQual

  - OverallCond

  - YearBuilt

  - TotalBsmtSF

  - GrLivArea

  - FullBath

  - GarageCars

  -GarageArea

  - SalePrice (Target)

Tools & Libraries

- Python 3.x

- NumPy

- Pandas

- Matplotlib / Seaborn

- Scikit-learn

Data Preprocessing

```
import pandas as pd

import numpy as np


df = pd.read_csv("/content/india_housing_prices.csv")


df.fillna(df.median(numeric_only=True), inplace=True)
```

```
from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()

categorical_cols = df.select_dtypes(include=['object']).columns

for col in categorical_cols:

    df[col] = le.fit_transform(df[col])
```

Exploratory Data Analysis

```
import seaborn as sns

import matplotlib.pyplot as plt


plt.figure(figsize=(10, 6))

sns.histplot(df['Price_in_Lakhs'], kde=True, bins=30)

plt.title("Distribution of Housing Prices")

plt.xlabel("Price")

plt.ylabel("Count")

plt.show()


numeric_cols = df.select_dtypes(include=['int64', 'float64']).columns

plt.figure(figsize=(12, 10))

sns.heatmap(df[numeric_cols].corr(), annot=True, cmap="coolwarm", fmt=".2f")

plt.title("Feature Correlation Matrix")

plt.show()


if 'Bedrooms' in df.columns:

    plt.figure(figsize=(8, 5))

    sns.boxplot(x='Bedrooms', y='Price_in_Lakhs', data=df)
```

```
plt.title("Price by Number of Bedrooms")
plt.xlabel("Bedrooms")
plt.ylabel("Price")
plt.show()
```

Model Building

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.preprocessing import StandardScaler

X = df.drop(['Price_in_Lakhs'], axis=1)
y = df['Price_in_Lakhs']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

model = RandomForestRegressor(n_estimators=20, random_state=42, n_jobs=-1)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
```

Evaluation

```
from sklearn.metrics import mean_squared_error, r2_score
```

```
mse = mean_squared_error(y_test, y_pred)

r2 = r2_score(y_test, y_pred)


print(f"Mean Squared Error: {mse:.2f}")

print(f"R² Score: {r2:.2f}")
```

Conclusion

- The model achieved strong predictive performance, showing high R^2 and low RMSE.

- Top features influencing price include OverallQual, GrLivArea, and GarageCars.

- Enhancements can include advanced models like XGBoost, feature selection, and interactive dashboards.

Future Enhancements

- Deploy model using Streamlit or Flask.

- Feature engineering for better predictions.

- Add user input interface for dynamic prediction.

Appendix

- Full code in house_price_prediction.py

- Dataset file: ` india_housing_prices.csv`

- GitHub Repository: [https://github.com/2303a51095/House-Price-Prediction]