House Price Prediction System

1. Introduction

The House Price Prediction System is an application designed to predict the price of houses based on a variety of input features. The system leverages machine learning algorithms to provide accurate estimates of house prices by training models on historical data.

The primary goal of this project is to predict house prices using multiple regression models, compare their performances, and deploy the best-performing model in a web application.

2. Problem Statement

Predicting house prices accurately is a critical task for both buyers and sellers in the real estate market. This project aims to build a regression-based model that can predict the prices of houses based on features such as the number of rooms, square footage, location, and other relevant factors.

3. Dataset

The dataset used for this project includes various features that influence house prices, such as:

Number of bedrooms

Number of bathrooms

Total square footage

Year of construction

Location (e.g.city)

Additional features like Balcony etc.

The dataset was split into training and testing sets to evaluate model performance.

4. Machine Learning Models Used

Several machine learning regression models were trained and evaluated for the task of predicting house prices. The key evaluation metrics used were:

R² (Coefficient of Determination): Measures how well the model captures the variance in the target variable.

Mean Squared Error (MSE): Average squared difference between predicted and actual values.

Root Mean Squared Error (RMSE): Square root of the MSE, provides error in the same units as the target variable.

Mean Absolute Error (MAE): Average of absolute differences between predicted and actual values.

Models Trained and Results

1. Linear Regression

R²: 0.8339

MSE: 0.0616

RMSE: 0.2483

MAE: 0.1798

2. KNeighborsRegressor

R²: 0.9328

MSE: 0.0249

RMSE: 0.1579

MAE: 0.0651

3. DecisionTreeRegressor

R²: 0.9211

MSE: 0.0293

RMSE: 0.1711

MAE: 0.0605

4. RandomForestRegressor

R²: 0.9518

MSE: 0.0179

RMSE: 0.1338

MAE: 0.0503

5. XGBoost

R²: 0.9520

MSE: 0.0178

RMSE: 0.1334

MAE: 0.0562

6. Gradient Boosting

R²: 0.9545

MSE: 0.0169

RMSE: 0.1299

MAE: 0.0552

5. Model Selection

After evaluating the performance of all models, Gradient Boosting was selected as the best model due to its highest R² score of 0.9545 and lowest MAE of 0.0552. This model provides the most accurate predictions of house prices compared to other models.

6. Web Application

The system was deployed as a web application using Flask. Users can input house features such as the number of rooms, square footage, location, and other relevant details. The model then predicts the house price and displays the result on the web interface.

Key Features of the Web Application:

User-friendly Interface: Simple and intuitive design for users to input house details.

Real-time Predictions: Instant prediction of house prices based on user inputs.

Model Selection: Gradient Boosting Regressor, the best-performing model, is used to provide accurate predictions.

7. Challenges Faced

Data Preprocessing: Handling missing values and categorical variables required careful preprocessing to ensure model accuracy.

Model Tuning: Fine-tuning hyperparameters for models like XGBoost and Gradient Boosting required significant experimentation to achieve the best results.

Deployment: Integrating the machine learning model with the Flask web framework and ensuring smooth deployment took considerable effort.

8. Conclusion

The House Price Prediction System successfully predicts house prices based on various input features. Gradient Boosting Regressor was selected as the final model due to its superior performance metrics. The system provides accurate, real-time predictions through a web interface and can be beneficial for users in the real estate market.