END-TO-END MACHINE LEARNING SOLUTION House Price Prediction System

Abstract

This project presents a complete end-to-end machine learning system for predicting house prices. The solution covers data preprocessing, feature engineering, model training, evaluation, interpretation, and deployment using a web interface.

Introduction

Real estate price estimation is a critical task for buyers, sellers, and financial institutions. Machine learning helps automate and improve pricing accuracy by learning patterns from historical data.

Business Understanding

The system helps real estate businesses automate property valuation, improve pricing strategies, and reduce manual effort.

Dataset Description

The dataset contains property-related features such as area, bedrooms, bathrooms, age, and price. It includes 300 records collected from public and simulated sources.

Exploratory Data Analysis

EDA was performed to understand feature distributions, correlations, and trends affecting house prices.

Data Preprocessing

Data cleaning, scaling, encoding, and train-test splitting were applied to prepare data for modeling.

Feature Engineering

Important features were selected and transformed to improve model performance.

Model Selection

Three models were implemented: Linear Regression, Decision Tree Regressor, and Random Forest Regressor.

Model Training

Models were trained using cross-validation and hyperparameter tuning to ensure generalization.

Model Evaluation

Random Forest achieved the best performance with MAE ■425,000 and R^2 score of 0.85.

Model Interpretation

Feature importance analysis showed area and location as the most influential predictors.

Deployment

A Flask-based web application was developed for real-time predictions.

Testing & Validation

Unit and integration tests ensured system reliability.

Limitations

Limited dataset size and lack of real-time market data.

Future Scope

Cloud deployment, larger datasets, and SHAP explainability can be added.

Conclusion

The project demonstrates a complete ML lifecycle and is suitable for real-world deployment.