House Price Prediction

Project Title

House Price Prediction using XGBoost Machine Learning Algorithm

Project Objective

The objective of this project is to build a machine learning regression model that predicts house prices based on various socio-economic and physical features of houses. The model learns patterns from historical housing data and estimates the median house price for new inputs.

Problem Type

- **Machine Learning Type:** Supervised Learning
- **Task:** Regression
- **Target Variable:** House Price (Median Value)

Dataset Description

- **Dataset Name:** Boston Housing Dataset
- **Source:** UCI Machine Learning Repository / Scikit-learn
- **Total Instances:** 506
- **Number of Features:** 13
- **Feature Type:** Numerical
- **Target Variable:** Median house price (MEDV)

Features Description (Examples)

- **CRIM:** Per capita crime rate
- **RM:** Average number of rooms per dwelling
- **AGE:** Proportion of owner-occupied units built before 1940
- **TAX:** Property tax rate
- **LSTAT:** Percentage of lower status population

(All features are numerical and used as input variables.)

Input and Output

- **Input (X):** 13 numerical housing features
- **Output (y):** Predicted house price value

Algorithm Used

XGBoost Regressor

XGBoost is chosen because it:

- Handles non-linear relationships effectively
- Provides high accuracy for regression tasks
- Uses gradient boosting to reduce prediction errors
- Performs well on structured tabular data

Project Workflow

1. Load Boston housing dataset
2. Perform data exploration and analysis
3. Separate features and target variable
4. Split dataset into training and testing sets
5. Train XGBoost regression model
6. Evaluate model performance
7. Make house price predictions

Model Evaluation

The model is evaluated using regression metrics:

- **Mean Absolute Error (MAE)**
- **Mean Squared Error (MSE)**
- **R-squared Score (R^2)**

These metrics measure how closely the predicted house prices match actual values.

Conclusion

The XGBoost regression model successfully predicts house prices based on housing features. This project demonstrates a complete end-to-end machine learning regression workflow, including data preprocessing, model training, evaluation, and prediction, making it suitable for intermediate-level ML learning.

Project Level

Beginner to Intermediate Machine Learning Project