ML Ops Report

Name : Ashutosh Gupta

Roll : M23CSE009

Boston Housing Price Prediction

Overview

The objective of this experiment is to predict housing prices using the **Boston Housing Dataset**. Two machine learning models were trained and evaluated:

1. **Linear Regression**
2. **Random Forest Regressor**

Both models were evaluated based on their **Mean Squared Error (MSE)**. The experiment was tracked using **MLflow** to log model parameters, metrics, and models.

Dataset Overview

Dataset Name: Boston Housing Dataset

Features:

- The dataset consists of 13 features, including:
    - crim: per capita crime rate by town
    - zn: proportion of residential land zoned for lots over 25,000 sq. ft.
    - indus: proportion of non-retail business acres per town
    - chas: Charles River dummy variable (1 if tract bounds river; 0 otherwise)
    - nox: nitric oxides concentration (parts per 10 million)
    - rm: average number of rooms per dwelling
    - age: proportion of owner-occupied units built prior to 1940
    - dis: weighted distances to five Boston employment centers
    - rad: index of accessibility to radial highways
    - lstat: percentage of lower-status population
- **Target Variable**:
    - medv: median value of owner-occupied homes in $1000s.

Models and Evaluation

1. Linear Regression

- **Mean Squared Error (MSE)**: 24.29
- **Model Summary**: Linear Regression is a simple model that assumes a linear relationship between the features and the target variable. Despite its simplicity, it produced a reasonable MSE for this dataset.

Key Result:

- MSE of 24.29 indicates that on average, the squared differences between the predicted and actual house prices are around 24.29 thousand dollars.

.

2. Random Forest Regressor

- **Mean Squared Error (MSE)**: 8.18
- **Model Summary**: Random Forest is an ensemble method that uses multiple decision trees to improve prediction performance. This model significantly outperformed Linear Regression with a much lower MSE.

Key Result:

- MSE of 8.18 indicates a better fit for this dataset compared to Linear Regression. The lower MSE means that Random Forest has more accurate predictions.

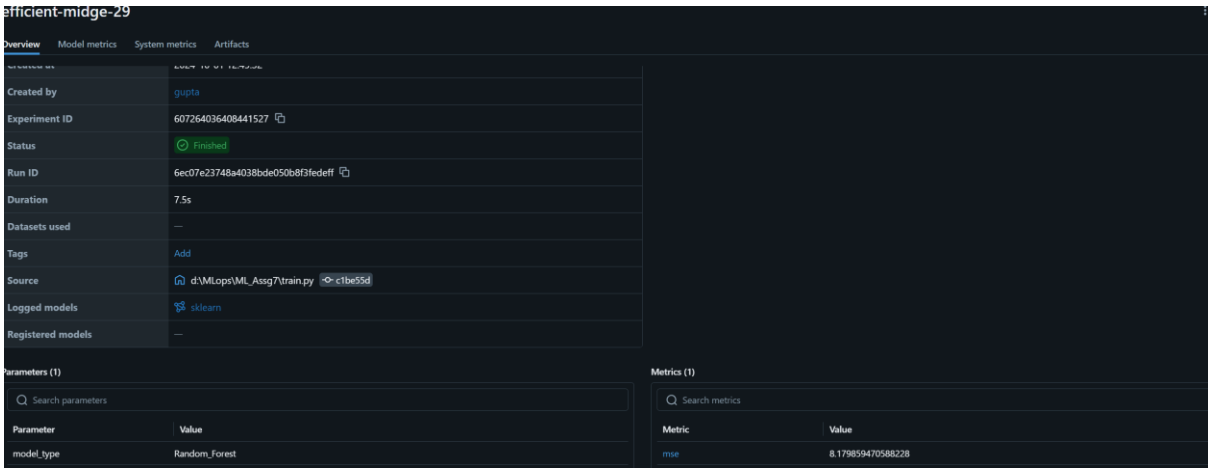
Summary of Results

Model	Mean Squared Error (MSE)
Linear Regression	24.29
Random Forest	8.18

- The **Random Forest** model performed significantly better than the **Linear Regression** model in terms of prediction accuracy, as indicated by the much lower MSE.
- Both models were logged in **MLflow** for future reference, comparison, and reproducibility. However, future improvements can include logging the model signature and input example for both models.

MLflow Artifacts

Random forest



Linear Regression