Boston Housing Prediction Analysis - Regression

This project involves the analysis of a dataset and the construction of machine learning models to predict a target variable. The dataset is preprocessed by handling missing values and converting categorical variables to numerical ones. Exploratory data analysis is then performed to gain insights into the dataset's features and relationships with the target variable.

The data is split into training and testing sets, and five regression models are trained on the data: Linear Regression, Decision Tree Regression, Random Forest Regression, Extra Trees Regression, and XGBoost Regression. The models are evaluated using various metrics such as mean squared error, mean absolute error, and R-squared.

The feature importance of each model is then visualized using a bar chart, highlighting the most important features for predicting the target variable. Finally, the best model is selected based on its performance and feature importance, and it is used to make predictions on new data.

The project's main objective is to predict the median value of owner-occupied homes in Boston using machine learning algorithms. The dataset consists of 13 input variables, and the best performing model was XGBoost Regression, evaluated using mean squared error and mean absolute error metrics. The project provides insights into the key factors that influence the value of homes in Boston, making it a valuable resource for anyone interested in the real estate market.

Dataset Information

The Boston House Prices Dataset, which comprises information on 506 residences in different Boston suburbs, was gathered in 1978 and includes 14 characteristics.

```
The Attribute Information for the Boston Housing Dataset is listed in a specific order, which is as follows::
        - CRIM     The crime rate for each town divided by the population
        - ZN       The percentage of residential land that is zoned for plots of land exceeding 25,000 square feet.
        - INDUS    The proportion of non-commercial business land in each town
        - CHAS     A dummy variable for the Charles River, which equals 1 if a tract bounds the river and 0 if it does not.
        - NOX      The concentration of nitric oxides in the air, measured in parts per 10 million,
        - RM       The average number of rooms in each residence
        - AGE      The proportion of homes occupied by their owners that were constructed before 1940
        - DIS      The distances, weighted by five employment centers in Boston.
        - RAD      The accessibility of each residence to radial highways.
        - TAX      The property tax rate per $10,000 of the total value of each property
        - PTRATIO  The pupil-teacher ratio for each town
        - B        Proportion of Black residents in each town, which is given by the expression 1000(Bk - 0.63)^2.
        - LSTAT    The percentage of the population in each town that has a lower socioeconomic status
        - MEDV     The median value of homes occupied by their owners, measured in thousands of dollars.
```

Libraries

scikit-learn
seaborn
matplotlib
pandas

Algorithms

XGBoost
Extra Tress
Random Forest
Decision Tree

```
<li>Linear Regression

# Data Source :
Download link: https://www.kaggle.com/puxama/bostoncsv
```