REPORT

By – Swagat Sekhar Pati,

Silicon Institute of Technology

**Dataset** - Boston Housing Dataset

**Objective** – Develop a simple ML model to predict housing prices based on various features.

**Method** **used** – REGRESSION

* Why Regression?
* Allows model to relate between various features of the dataset
* Predict continuous outcomes , specify numerical values
* Flexible with multiple variables
* High predictive accuracy
* Identifies which features are more significant predictors

**Models used** – 1) Linear Regressor

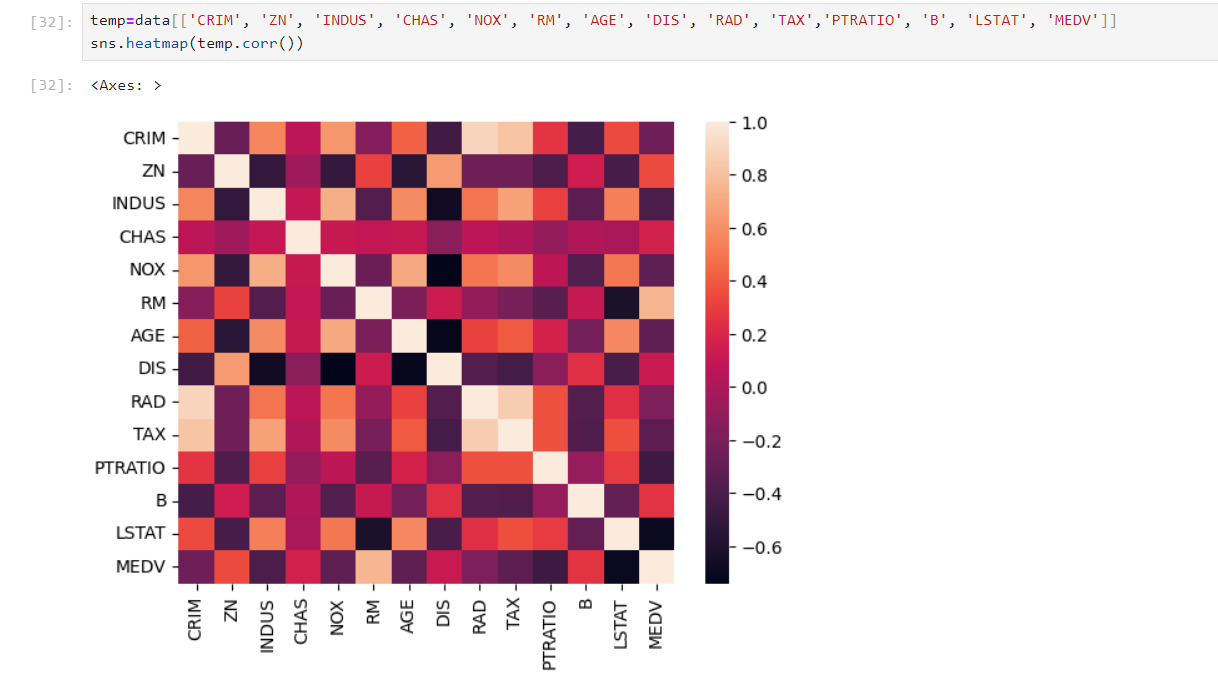
2) Decision Tree Regressor

3) Extra Tree Regressor

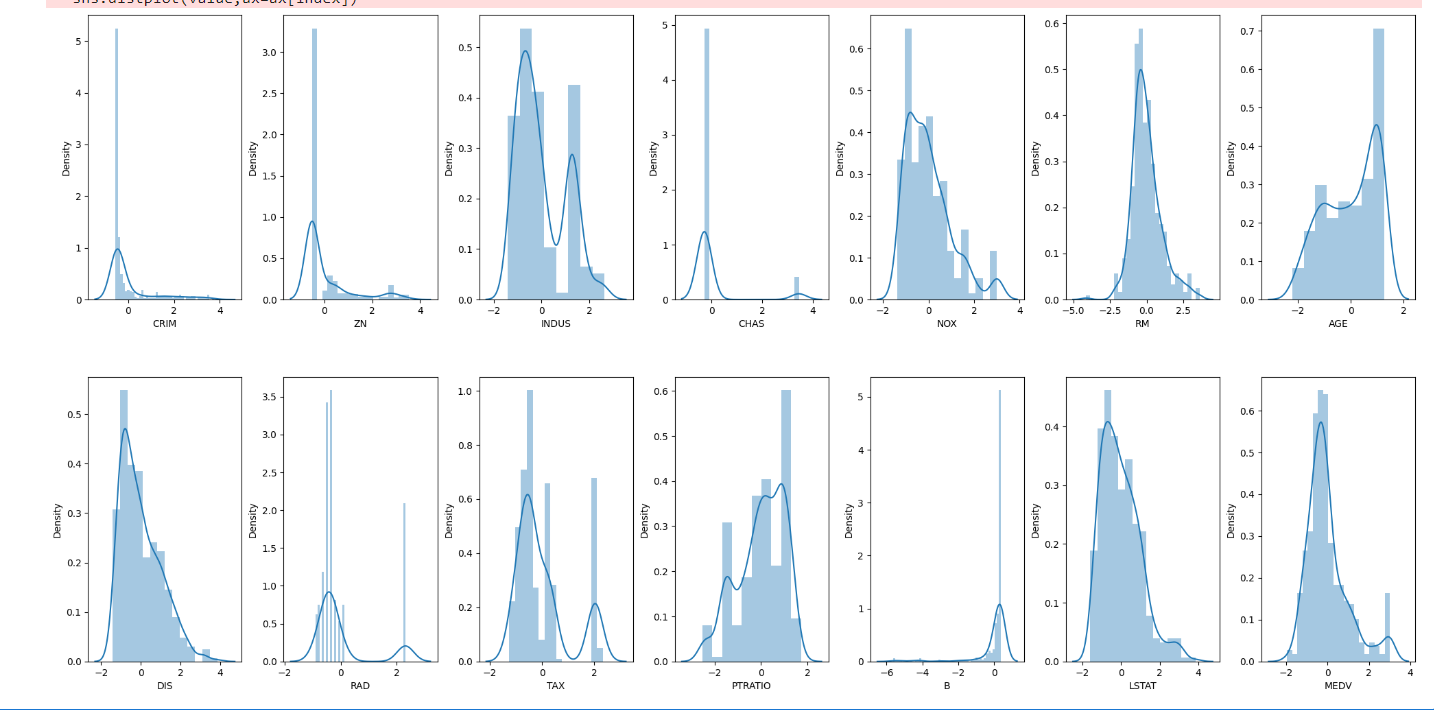
4) Random Forest Regressor

5) Extreme Gradient Boost Regressor

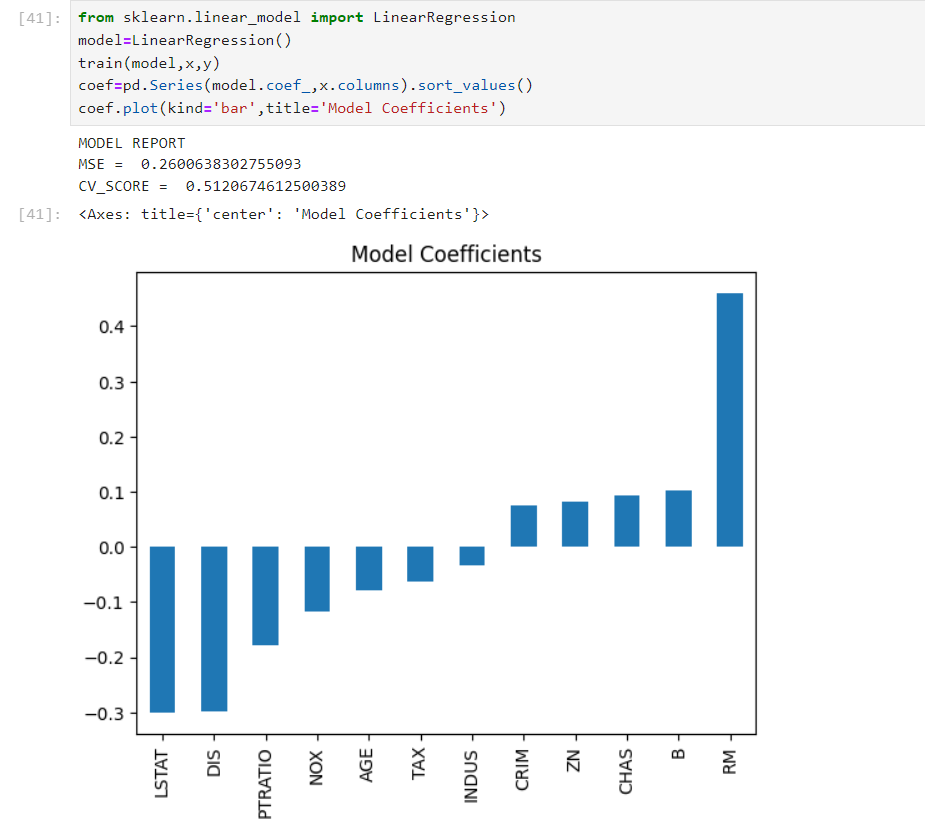
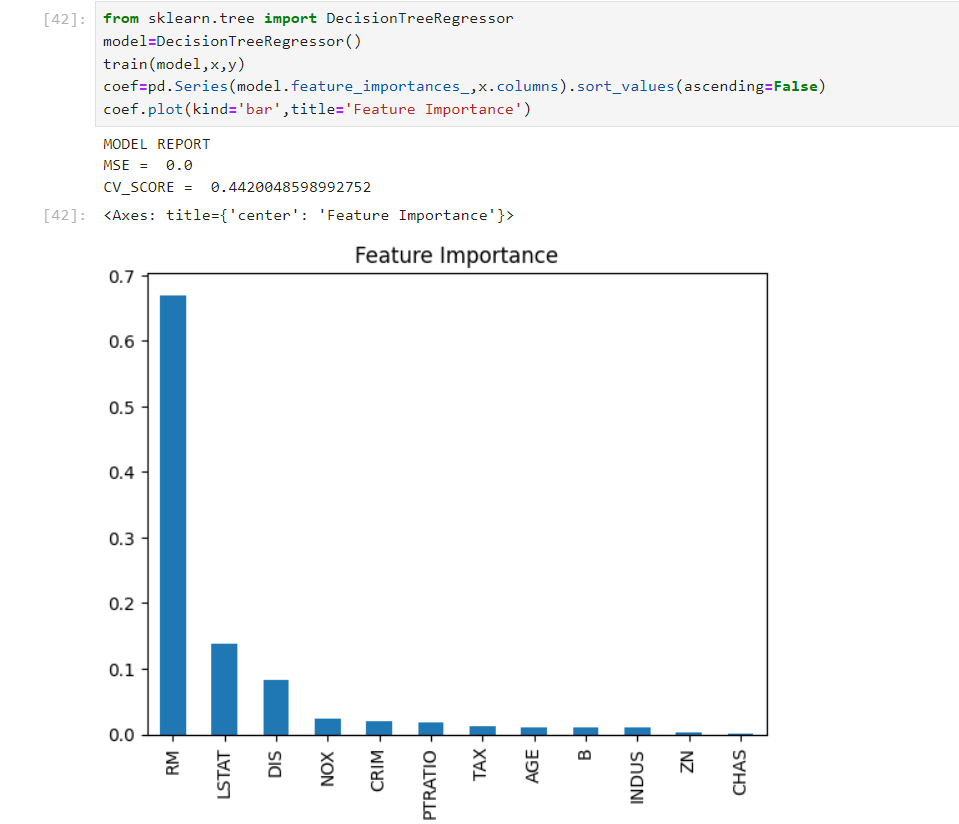
1. Preprocess and clean the dataset from null values
2. Remove outliers
3. Find correlation between the features of the dataset



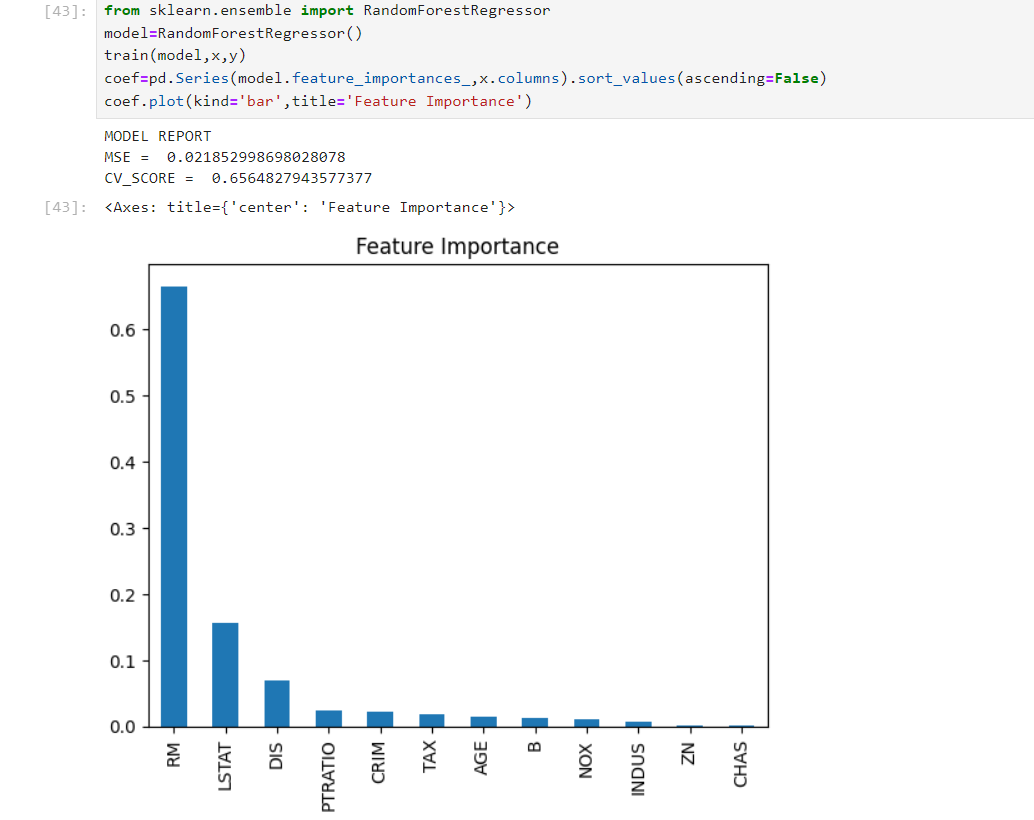
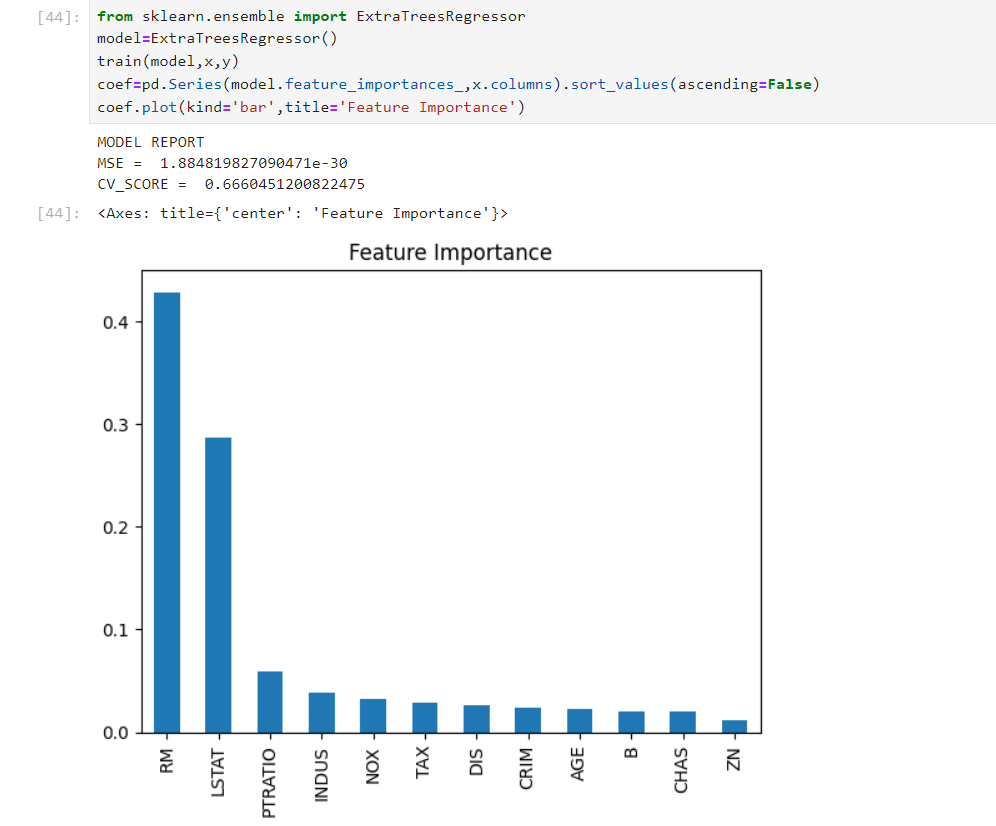
1. The distribution plot of each individual feature for better understanding of how deviated the features are



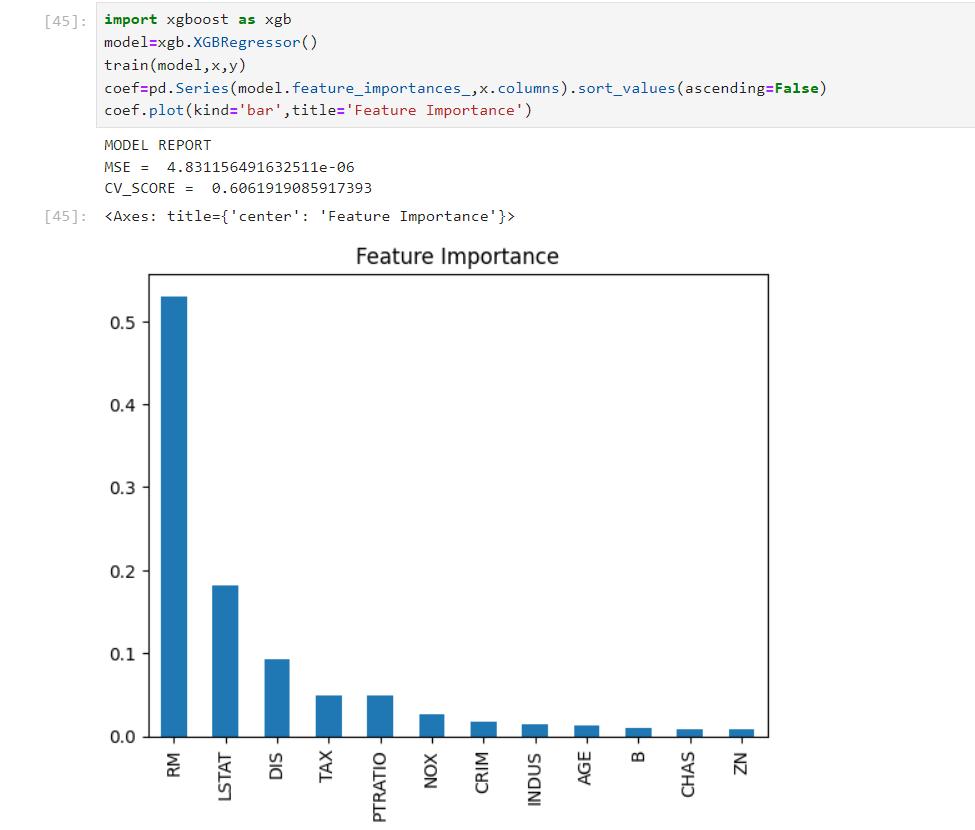
1. Compare all the models used in the dataset and find the best model that can predict the house price based on feature importance

1)Linear Regression 2)Decision Tree Regression

3)Random Forest 4) Extra Trees



1. Extreme Gradient Boost Regression

***Conclusion***

From the above 5 models’ values , we can conclude that **Extreme Gradient Boost** and **Extra** **trees Regressor** are the best fitted and predictive model for the given dataset as in both the cases 70% of the total factors are dependent on “RM” and “LSTAT”

i.e, 70% of the price will be dependent on these columns.