# Housing Price Prediction

## Project Title:

Housing Price Prediction

## Objective:

To predict the price of house using different parameters like Crime Rate, Room number, House Age, Tax, etc. so that real estate companies can use them to make profit in buying and selling activities. For example, they can buy a house that shows higher predicted price compared to its actual cost and profit their business and also vice-versa.

## Link to the Dataset:

https://archive.ics.uci.edu/ml/machine-learning-databases/housing/housing.data

## Variables in the Dataset:

The dataset contains the following

- 1. CRIM      per capita crime rate by town
- 2. ZN        proportion of residential land zoned for lots over

- ▪ 25,000 sq.ft.
- 3. INDUS    proportion of non-retail business acres per town
- 4. CHAS    Charles River dummy variable (= 1 if tract bounds
  - ▪ river; 0 otherwise)
- 5. NOX    nitric oxides concentration (parts per 10 million)
- 6. RM    average number of rooms per dwelling
- 7. AGE    proportion of owner-occupied units built prior to 1940
- 8. DIS    weighted distances to five Boston employment centres
- 9. RAD    index of accessibility to radial highways
- 10. TAX    full-value property-tax rate per $10,000
- 11. PTRATIO  pupil-teacher ratio by town
- 12. B    $1000(Bk - 0.63)^2$ where Bk is the proportion of blacks
  - ▪ by town
- 13. LSTAT    % lower status of the population
- 14. MEDV    Median value of owner-occupied homes in $1000's

# Variables used from the Dataset:

Same as above variables in the dataset.

# List of independent variables:

- CRIM
- ZN
- INDUS
- CHAS
- NOX
- RM
- AGE
- DIS

- RAD
- TAX
- PTRATIO
- B
- LSTAT

## List of dependent variables:

- MEDV

## ML Models used:

Linear Regression

Decision Tree Regressor

Random forest Regression

## Accuracy of the models:

Linear Regression model:

Accuracy on Traing set:  0.7465991966746854

Accuracy on Testing set:  0.7121818377409193

Decision Tree Regressor model:

Accuracy on Traing set:  1.0(overfit)

Accuracy on Testing set:  0.6687323989899613

Random Forest Regression model:

Accuracy on Traing set:  0.978487690385328

Accuracy on Testing set:  0.840930157856634


Here, accuracy value of forest regression model is the best and so I selected this model to find the best possible solution.


## Steps implemented and description:


Initially, I imported required libraries and gathered dataset from sklearn datasets. Then, I loaded the data into the jupyter notebook and analyzed the data. After that, I analyzed all the features by constructing various plot diagrams and correlation matrix.

Then, I split the data into training and testing data and tried different models on training dataset and selected the best one.