**Table of Contents**

2. Introduction

3. Dataset Description

3. References

**INTRODUCTION**

Humans have various needs and desires that ought to be fulfilled for quality life or life in general. Any human desires for various levels of needs in their life. Among those needs, one of the most important physiological need of a human is shelter and security. Housing provides these exact needs to a human. Housing generally refers to living spaces or construction and usage of houses or buildings collectively for sheltering people or some provisions. Housing provides not only the sheltering and security needs but also social needs and provides stability in life. Housing is not only a source of need fulfillment for humans, but a human also values their shelter (housing) as a prized possession and a place to belong i.e. home. Families often look for decent and affordable houses that not only provides them shelter but also contributes to the well being of themselves and their future generations. Proper housing also contributes to health as studies show that continuous roaming is stressful and not healthy as people become more prone to diseases while affordable and proper housing provides relaxation and better mental and physical condition. Along with these, housing also provides comfort and peace of mind with a complete freedom inside the house without interference from outside society. All these interpretations refer that housing is very important to humans thus its study and research is also valued greatly.

In the previous years, housing studies have generally been around housing prices along with affordability and its quality. Studies show that housing price is affected by macroeconomy where increase in Consumer Price Index lead to an increase in housing price along with property price raise too. Along with these, various factors like Money Supply, Best Lending Rate, Mortgage Repayments, Real Interest Rate etc. also affect housing prices. Nevertheless, housing prices are direct result of quality of a house.

The ML models used in this project are Linear Regression, Polynomial Regression, Decision Tree Regressor and Random Forest Regressor. These ML models will be compared with each other in their efficiency and productivity on accuracy, learning and prediction. These ML models are available in the open sourced “Scikit-learn” library available in Python programming language for Machine Learning. The data related to housing is provided in UCI archive. The data is also available in Kaggle. The data in Kaggle is the copy of the one in UCI which is indexed and properly arranged for easier access.

The data can be found for study or download on the following links:

**UCI Data** **:** <https://archive.ics.uci.edu/ml/machine-learning-databases/housing/>

**ABOUT THE DATA**

The data used in this project is the one provided in Kaggle, but since both data links provide same data content, it doesn’t matter what data link we use. The only difference is that the data in UCI is branched while the data in Kaggle is indexed and properly arranged. The data is about housing attributes and various factors surrounding its price. The dataset provides insight on different attributes related to housing like size, price, quality etc.

The provided dataset is a multivariate data with about 506 instances and 14 real attributes. The data file in summary has 14 columns namely CRIM, ZN, INDUS, CHAS, NOX, RM, AGE, DIS, RAD, TAX, PTRATIO, LSTAT and MEDV with 506 rows i.e. (506 x 14). The associated tasks related to this data may be Classification and Clustering.

The attributes of the data with their individual information are given below:

**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

**TAX** : full-value property-tax rate per $10,000

**PTRATIO** : pupil-teacher ratio by town B 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town

**LSTAT** : % lower status of the population

**MEDV** : Median value of owner-occupied homes in $1000's

**REFERENCES**

**SITE REFERENCES :**

[**https://archive.ics.uci.edu/ml/machine-learning-databases/housing/**](https://archive.ics.uci.edu/ml/machine-learning-databases/housing/)

[**http://www.housingvirginia.org/housing-virginia-toolkit/why-is-affordable-housing-important-is-rental-or-homeownership-more-important/**](http://www.housingvirginia.org/housing-virginia-toolkit/why-is-affordable-housing-important-is-rental-or-homeownership-more-important/)

[**https://mg.co.za/article/2008-10-10-the-importance-of-adequate-housing**](https://mg.co.za/article/2008-10-10-the-importance-of-adequate-housing)

[**https://www.mindcontroversy.com/importance-of-home/**](https://www.mindcontroversy.com/importance-of-home/)

[**https://en.m.wikipedia.org/wiki/Housing**](https://en.m.wikipedia.org/wiki/Housing)