House Price Prediction

Submitted by:

Fazal Ahamed

Batch:1826

ACKNOWLEDGMENT

I have done it by myself , referred few resources such as kaggle, towards data science, my younger brother has given me some input in Data analysis.

INTRODUCTION

- ## Business Problem Framing

The Dataset consists of various houses data ranging from 1975 to 2010.
Based on Construction model, years, built area in we are going to predict the sales price of the house.

- ## Conceptual Background of the Domain Problem

As it is a housing dataset, we will mainly focus on connectivity, city, near by location as major factors of the datset.

- ## Review of Literature

As We are going to analyse the train data set given to us in csv format, by loading it in jupyter notebook.
I will analyse the data and then will perform EDA steps to have a basic knowledge on how the data is distributed.
By using univariate , Bivariate ,MultiVariate analysis, we will gain some insights.

Going further we will check for the Null values present in the dataset, we will treat them using mean median or mode based upon the data is categorical or numerical.

Then we will check for skewness removal.
Once all the EDA is done the we will convert categorical data into numerical data using Label Encoder.

As we all are aware that machine only understands the binary form of the data.

- ## Motivation for the Problem Undertaken

As it is the Housing dataset price prediction am very much excited in analyzing, building a model which can predict the prices.
It gives me a so much pleasure in working on projects like this, Surely this will boost my confidence a bit more towards doing project in machine learning.
With this project I have gained a lots of ideas how to approach a dataset, how to project your graphs etc.

Analytical Problem Framing

- ## Mathematical/ Analytical Modeling of the Problem

In this project as it contains continious data to predict we will use Linear Regression to build the model.

- ## Data Sources and their formats

No data sources referred , its my own analysis of the data.

- ## Data Preprocessing Done

What During the Cleaning process I have clearly checked null values, If found applied Mean median and mode techique to fill those nan values.

Later Using Label encoder I have tranformed data into 0,1 so that it can fit in the model.

- ## Hardware and Software Requirements and Tools Used

Python jupyter notebook used as a tool for this housing project.
As it contains many libraries to visualize the data and build a ML model.

- ## Model/s Development and Evaluation

Listing The Lasso and Ridge Model used for Regularization and MAE, MSE,Rmse, used for evaluation metrics.

- ## Interpretation of the Results

As we can see accuracy is good, predicted result was giving almost nearest values in all the test data set given to us.

CONCLUSION

- ## Learning Outcomes of the Study in respect of Data Science

Data visualization gives a better understanding of the dataset.

Data cleaning is a major part, As we have cleared few columns with max number of null values and filled few of them usinf fillna using mean.

As it is a Regression problem, where the output is continious so I have used linear regression model.

Lasso and Ridge regularization for improvement and Ensemble technique to boost efficiency of the model.

And finally hyper tuned to get the best state.

- ## Limitations of this work and Scope for Future Work

The limitations the project is It consists of various types of data, hence It took lots of time for analysis.

The model built here can only be used for house prediction.

Future Scope of work:

This model can be used in future as well to predict the output as far as the data remains in same format.

Techniques can be followed to further extend this study and improve the results.

As we gather little more data then we can improve in our models efficiency.