

**HOUSING: PRICE PREDICTION**

Submitted by:

Sagnik Das

**ACKNOWLEDGMENT**

I would like to express my special thanks of gratitude to (Datatrained) also to my SME (Mr Sajid Choudhary). Who gave me the golden opportunity to do this wonderful project on the topic (Micro Credit Defaulter Project), which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to tall of them.

**INTRODUCTION**

* Business Problem Framing

This business problem helps us to understand the trend to increase the overall revenue, profits, of companies by improving their marketing strategies and focusing on changing trends in house sales and purchases.

Which variables are important to predict the price of variable?

How do these variables describe the price of the house?

* Conceptual Background of the Domain Problem

This problem is divided into 4 steps:-

1. Extensive data cleaning – Cleaning of zero and nan values, on train dataset and test dataset.
2. EDA Analysis, on train and test dataset.
3. Hyperparameter optimization tuning, on train dataset.
4. Predicting the value on test dataset

* Review of Literature

This is a comprehensive summary of the research done on the topic of “HOUSING: PRICE PREDICTION” and the variables on which the value of house depends.

* Motivation for the Problem Undertaken

To find out the factors on which the value of a can increase or decrease. Also, to find out the factors to increase the overall revenue, profits, of companies by improving their marketing strategies and focusing on changing trends in house sales and purchases.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem



* Data Sources and their formats

Data Provided by Flip Robo Technologies. Files provided in Microsoft Excel Comma Separated Values File (.csv) format.

* Data Pre-processing Done

• Remove duplicate or irrelevant observations.

• Fix structural errors

• Filter unwanted outliers

• Handle missing data (from both train and test dataset)

• Changed strings into integers

• Hyperparameter optimization using grid search cv

• Validate

* Data Inputs- Logic- Output Relationships

Input - HOUSING: PRICE PREDICTION

Output - Find out the factors to increase the overall revenue, profits, of companies by improving their marketing strategies and focusing on changing trends in house sales and purchases.

* State the set of assumptions (if any) related to the problem under consideration

Extensive data cleaning – Cleaning of zero and nan values. Set of columns selected that affects the value factor.

* Hardware and Software Requirements and Tools Used

Hardware – Laptop

Software –

import pandas as pd, import os, import csv, import sklearn, import numpy as np, import matplotlib.pyplot as plt, import seaborn as sns, %matplotlib inline, Hyperparameter optimization using grid search cv.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

Extensive data cleaning, EDA Method, Linear Regression and prediction, Hyperparameter optimization using grid search cv.

* Testing of Identified Approaches (Algorithms)

EDA Method, Hyperparameter optimization using grid search cv.

* Run and Evaluate selected models

EDA Algorithm, Linear Regression, prediction, Hyperparameter optimization using grid search cv.

* Key Metrics for success in solving problem under consideration

What were the key metrics used along with justification for using it? You may also include statistical metrics used if any.

* Visualizations

Contour, Heat map, pair plot, box plot, histogram, gridsearchcv.

* Interpretation of the Results

Find out the factors to increase the overall revenue, profits, of companies by improving their marketing strategies and focusing on changing trends in house sales and purchases.

The model also, shows that value of the house increase if the square feet area is bigger, condition of the house is good, basement is big and is in good condition the same is also correct for the garage. It also shows that value also depends on what type of land it is situated and it is located in which neighbourhood, also the year of building or repair is also important.

**CONCLUSION**

* Key Findings and Conclusions of the Study

This model shows that value of the house increases if the square feet area is bigger, condition of the house is good, basement is big and is in good condition the same is also correct for the garage. It also shows that value also depends on what type of land it is situated and it is located in which neighbourhood, also the year of building or repair is also important. Year of sale is also important as in histogram it is shown, that the price decreased in the year 2008 because of recession.

* Learning Outcomes of the Study in respect of Data Science

Data cleaning helped in cleaning excess nan data and EDA algorithm helped it in putting it in visualised matter. Hyperparameter optimization using grid search cv helped to create a model and predict. It also helps in identifying the variables that are dependent in increasing or decreasing the value of the house.

* Limitations of this work and Scope for Future Work

Excessive unfiltered data. 2 datasets with different attributes. Both datasets had few different attributes, which made it hard to compare the two datasets.

Can help housing organizations to predict the value of the house. Identify variables that are dependent in increasing or decreasing the value of the house. Helps in increasing the overall revenue, profits, of companies by improving their marketing strategies and focusing on changing trends in house sales and purchases by creating good models.