

HOUSE PRICE PREDICTION

Submitted by:

HUMAYUN KABIR

**ACKNOWLEDGMENT**

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

**INTRODUCTION**

* Business Problem Framing

For starting a real estate business, a company want to predict the price of house. Hence the company collected data and want to analyse and get a ML model to predict the price of house.

* Conceptual Background of the Domain Problem

Many times, it is observed for the real estate projects that the house made is in not demand and does getting price according to its costing. Hence, the company must bear the losses. Hence a proper data analyses and ML model can help the developer to develop the right product.

* Review of Literature

Data has been collected with 1168 rows × 81 columns. In the spread sheet house price provided based 80 parameters. Based on null value present and corelation with sale price few data were deleted, and matrix become in to (1168, 46). Exploratory data analysis and Machine learning model with the data has been made which successfully worked with accuracy of around 80%.

* Motivation for the Problem Undertaken

The project is a challenging project for me. Initially it was difficult for me to analysis this big number of parameters. Anyways the problem has been overcome and a successful model has been prepared which can predict house price with around 80 % accuracy.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

Exploratory data analysis of the data has been done found that Sale price of house is strongly corelated with External Quality, Kitchen Quality, Garage Area, Gross Living Area, and Overall Quality.

Machine Learning with the data has been done using different algorithm and details are as follows:

|  |  |
| --- | --- |
| Algorithm | Accuracy |
| Linear Regression | 81% |
| GridSearchCV | 81% |
| Lasso | 81% |
| RandomForestRegressor | 85% |

Graphical representation for actual house price and predicted house price for Linear regression is given below:

Chart, scatter chart

Description automatically generated

* Data Sources and their formats

Data was collected from Flip Robo in CSV format. The data is having 1168 rows × 81 columns.

* Data Preprocessing Done

First data has been checked for any kind of null value. Wherever null value is found to be present, the column has been deleted. Than corelation between the sale price and other parameters has been calculated and lesser corelation parameters eliminated from the DataFrame.

* Data Inputs- Logic- Output Relationships
* Refer to section ‘Mathematical/ Analytical Modeling of the Problem’.
* State the set of assumptions (if any) related to the problem under consideration

No assumption made.

* Hardware and Software Requirements and Tools Used

Python and its libraries only used for the project.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

Exploratory data analysis of the data has been done found that Sale price of house is strongly corelated with External Quality, Kitchen Quality, Garage Area, Gross Living Area, and Overall Quality.

Machine Learning with the data has been done using different algorithm and details are as follows:

|  |  |
| --- | --- |
| Algorithm | Accuracy |
| Linear Regression | 81% |
| GridSearchCV | 81% |
| Lasso | 81% |
| RandomForestRegressor | 85% |

Graphical representation for actual house price and predicted house price for Linear regression is given below:

Chart, scatter chart

Description automatically generated

.

* Testing of Identified Approaches (Algorithms)

|  |  |
| --- | --- |
| Algorithm | Accuracy |
| Linear Regression | 81% |
| GridSearchCV | 81% |
| Lasso | 81% |
| RandomForestRegressor | 85% |

* Run and Evaluate selected models

Refer to attached Python file.

* Key Metrics for success in solving problem under consideration

Refer to attached Python file.

* Visualizations

Refer to attached Python file.

* Interpretation of the Results

Exploratory data analysis of the data has been done found that Sale price of house is strongly corelated with External Quality, Kitchen Quality, Garage Area, Gross Living Area, and Overall Quality.

Machine Learning with the data has been done using different algorithm and details are as follows:

|  |  |
| --- | --- |
| Algorithm | Accuracy |
| Linear Regression | 81% |
| GridSearchCV | 81% |
| Lasso | 81% |
| RandomForestRegressor | 85% |

Graphical representation for actual house price and predicted house price for Linear regression is given below:

Chart, scatter chart

Description automatically generated

.

**CONCLUSION**

* Key Findings and Conclusions of the Study

Model found with good accuracy more than 80%. This model will help to estimate price of the housing projects.

* Learning Outcomes of the Study in respect of Data Science

This model will help to calculate risk to reward ratio calculation in a better way.

* Limitations of this work and Scope for Future Work

This may be old and not in line the sentiment of the people where current project will be developed hence some variation can be seen in field.