

**HOUSING PRICE PREDICTION REPORT**



**Submitted by:**

**Pratiksha Avinash Potghan**

**(Internship 30, Flip Robo Technology)**

**ACKNOWLEDGMENT**

It is my pleasure to present this report. Working on this project was a great experience that gave me very informative knowledge of data analysis.

Flip Robo Technologies (Bangalore) provided all the required information and dataset, which helped me complete the project.

I want to thank my SME Mohd. Kashif for giving me the dataset and instructions to perform the complete case study process.

**INTRODUCTION**

Houses are one of the necessary need of each and every person around the globe and therefore housing and real estate market is one of the markets which is one of the major contributors in the world’s economy. It is a very large market and there are various companies working in the domain. Data science comes as a very important tool to solve problems in the domain to help the companies increase their overall revenue, profits, improving their marketing strategies and focusing on changing trends in house sales and purchases. Predictive modelling, Market mix modelling, recommendation systems are some of the machine learning techniques used for achieving the business goals for housing companies. Our problem is related to one such housing company.

A US-based housing company named Surprise Housing has decided to enter the Australian market. The company uses data analytics to purchase houses at a price below their actual values and flip them at a higher price. For the same purpose, the company has collected a data set from the sale of houses in Australia. The data is provided in the CSV file below.

The company is looking at prospective properties to buy houses to enter the market. You are required to build a model using Machine Learning in order to predict the actual value of the prospective properties and decide whether to invest in them or not. For this company wants to know:

• Which variables are important to predict the price of variable?

• How do these variables describe the price of the house?

* ***Business Problem Framing***
* In this “Housing Price Prediction” project we need to build a machine learning regression model which can predict the house prices accurately.
* We have many independent/input attributes present in given dataset from which we need to find the best attributes which effect the property prices and nature of those attributes on property price.
* This project will help the companies in real estate investments in the Australian market, since it will predict the price of the houses located in any particular area with the help of which investor can buy the houses in good prices.
* ***Conceptual Background of the Domain Problem***
* As this project is all about the price prediction of houses located in the city of Australia, so knowledge of housing or real estate plays a very important role in this project while choosing the main attributes for the price prediction.
* Like we all know that housing prices mainly dependent on the locality in which house is located, various facilities like road, electricity supply, transportation connectivity also plays major role in the housing prices.
* The main object of these problem is :

1. To apply data pre-processing to obtain a clean data.
2. To build the machine learning models which are able to predict the housing prices based on its features.
3. To analyze and select the best model among all.

* ***Review of Literature***
* House costs increment consistently, so there is a need for a framework to anticipate house costs later on. House cost expectation can assist the designer with deciding the selling cost of a house and can assist the client with orchestrating the perfect chance to buy a house. There are three factors that impact the cost of a house which includes **physical condition, idea and area.**
* House price forecasting is a crucial topic of land. Machine learning techniques are applied to research historical property transactions in World to get useful models for house buyers and sellers.
* For the choice of prediction methods we compare and explore various prediction methods.
* We utilize **regression** as our model due to its adaptable and probabilistic methodology on model selection. Our result exhibit that our approach of the difficulty got to achieve success, and has the power to process predictions that might be comparative with other house cost prediction models. This study utilizes machine learning algorithms as a search method that develops housing price prediction models.
* We create a housing cost prediction model in sight of machine learning algorithm models. House price prediction on a knowledge set has been done by using all the above mentioned techniques to seek out the simplest among them.
* A number of the related factors that impact the value were also taken into considerations like physical conditions, concept and site etc. Here we aim to form our evaluation supported every basic parameter that's considered while determining the worth.
* We use linear regression techniques during this pathway, and our result aren't sole determination of one technique rather it's the weighted mean of varied techniques to offer most accurate results.
* T he results proved that this approach yields minimum error and maximum accuracy than individual algorithm applied.

**Keywords:** House price prediction, Machine Learning, AI, Model, Linear Regression, Algorithm.

* ***Motivation for the Problem Undertaken***

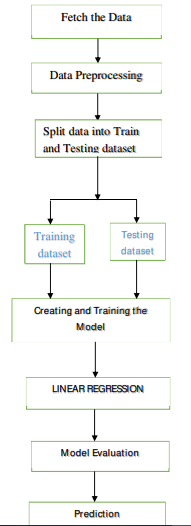
My main objective behind in building this project was to get a best Machine learning model which can be a good help and will be an accurate answers to all the questions raised by the house buyers and sellers.

**Analytical Problem Framing**

* ***Mathematical/ Analytical Modeling of the Problem***
* A US-based housing company named **Surprise Housing** has decided to enter the Australian market. The company uses data analytics to purchase houses at a price below their actual values and flip them at a higher price. For the same purpose, the company has collected a data set from the sale of houses in Australia.
* In this project we have many independent variables or features with the help of which we want to predict our target variable “**Sales Price**” using different machine learning algorithms.
* ***Data Sources and their formats***
* The dataset utilized during this work is obtained from land Agents within the US to gauge the prices of homes.
* This dataset was intended to help them find the foremost appropriate cost of the house in the locations within the US.
* These attributes, or predictor variables, are the factors which are majorly considered during house purchase and thus influence the pricing of homes. The dataset utilized in this work is obtained from land Agents within the US to judge the costs of homes.
* ***Data Preprocessing Done***
* As the first step I first checked the null values using **main\_data.isnull().sum()** in our dataset and then visualized them using heatmap.
* Then filled all the missing data of our dataset using mean and mode.
* After that I checked the statistical description of data using **main\_data.describe(),** which gave me count, mean, standard deviation, minimum, 25%, 50%, 75% and maximum.
* Then encoded the data so that object data will be converted to numeric values.
* Then checked correlation using **main\_data.corr(),** also checked for outliers and skewness.
* ***Data Inputs- Logic- Output Relationships***
* In this project we have many input/independent features some of them have numerical datatype, some of them have categorical datatype. By means of encoding we converted all categorical datatype into binary 0/1 datatypes. Later we fed this independent feature to our machine learning model to predict our target variable “Sales Price”
* ***Hardware and Software Requirements and Tools Used***
* We have used jupyter notebook as an environment of model building.
* Programming language python 3 is used in this project,along with that we have used many python packages also which are listed below:
  1. Numpy→Package for scientific computing .
  2. Pandas→Package for data manipulation and analysis.
  3. Matplotlib→Visualisation package
  4. Seaborn→Visualisation package
  5. Sklearn →software machine learning package

**Model/s Development and Evaluation**

* ***Testing of Identified Approaches (Algorithms)***
* Linear Regression Model
* AdaBoost Regressor Model
* Gradient Boosting Regressor Model
* Support Vector Regressor Model
* Random Forest Regressor Model
* Decision Tree Regressor Model
* KNearest Neighbour Regression Model
* ***Key Metrics for success in solving problem under consideration***
* Key metrics for success are MAE(Mean Absolute error) and MSE(Mean squared error), r2 score and cross validation score.
* We selected GradientBoosting as our best model as r2 score for it was 89% and cross validation score was 86% whereas RandomForest gave us r2 score as 87% and cross validation score as 84%.
* ***Visualizations***
* We did visualization of data using :
* count plot
* scatter plot
* bar plot
* distribution plot.
* ***Interpretation of the Results***



**CONCLUSION**

* ***Key Findings and Conclusions of the Study***
* In this project we built several regression models to predict the best price of a house with given input features.
* We evaluated and compared all the models and then selected the best out of them.
* We also checked the rank wise importance of all the features.
* In this project we followed all the necessary data science steps to obtain model with full accuracy.
* The process includes data selecting, data cleaning, data preprocessing, data visualization, splitting data into train and test, then creating and training model, using all regression models and finally model evaluation and prediction.
* ***Learning Outcomes of the Study in respect of Data Science***
* To obtain clean data.
* Analyze data properly.
* Use of appropriate models for getting accurate scores.

.