# HOUSE PRICE PREDICITION USING ML

**Puja A. Chavan (Cholke)*, Prem Shejole, Rituraj Sharma, Shafaque Shaikh,Dheeraj Shelke, Pratham Sharma**

*Vishwakarma Institute of Technology Pune, India*
puja.cholke@vit.edu, pratham.sharma21@vit.edu, rituraj.sharma21@vit.edu, shafaque.shaikh21@vit.edu, prem.shejole21@vit.edu, dheeraj.shelke21@vit.edu

***Abstract —***
Nowadays people are trying to buy home by considering the various features like area, location, bedrooms etc. with their market strategies and budgets. Our house price model predict the price of house for peoples based on the features of house and their budget. Data mining is now widely used in the real estate market. In this paper house price dataset of Bengaluru is used which involves various parameters like area, location, bedrooms etc. also linear regression algorithm is used to train our dataset out of which 80% data goes for training and 20% data is goes for testing. As a result, user give text as input means which type of area, how many bedrooms, etc. user wants on the basis of that user will get the predicted price of house. According to research, changes in housing prices are often a cause for concern for homeowners and the housing market. Book reviews are done to evaluate key features and most effective models for real estate forecasting. This study demonstrates the use of Artificial Neural Network, Relocation Line as one of the most effective models compared to others. In addition, our findings suggest that land building factors are important factors in predicting house prices.

Keywords - Machine Learning Model, Qualities, Linear Regression.

# I. INTRODUCTION

A house is one of the most basic necessities of life, as well as food, water, and other necessities. As living standards improved, demand for housing increased rapidly. Although some people build houses for money and property, most people around the world buy a house as a shelter or as a way to earn money. Every year, there is an increase in demand for housing, leading to an increase in housing prices. The issue arises when there are many variables such as location and the need for a building that may affect the value of the house; as a result, most stakeholders including buyers and developers, real estate developers, and the real estate industry would like to know the specific factors or factors that influence house prices to help investors make decisions and homeowners set house prices.Previous research has focused on analyzing the factors that influence house prices and predicting house prices using a separate machine learning model. However, this article covers both real estate forecasts and real estate. Many speculative models (Machine Learning Models) such as vector retreat, sensory network, and others can be used to predict house prices. The real estate model offers many benefits to real estate buyers, real estate investors, and real estate agents. This model will provide a wealth of information and information to homeowners, real estate investors, and real estate developers, such as estimating real estate prices in current markets, which will help them define real estate prices.

For this project, we have used the Bengaluru real estate database which contains the prices of different real estate based on different parameters such as location, total square meter Real estate, size, BHK, etc. By analyzing the data our model is able to predict. house price.

# II. LITERATURE REVIEW

Following are the works related to House price prediction:

In paper [1] they discussed about the the the several Algorithms after training and testing these algorithms they selected XGBoost regression Algorithm also discussed about the datatset and the attributes which necessary in dataset also discussed about that remove those column and rows which don’t have any effect on final prediction results.

In paper [2] they compared the hedonic price model with an artificial network model both the models tested using 200 house information from the results the hedonic price models show poorer results by comparing with the results of neural network model.

In paper [3] they presented different machine learning methods performance in predicting the selling price of houses based on a number of features such as the area, the number of bed- and bathrooms and the geographical position.

In paper [4] they discussed about the prediction of future housing prices that is generated by using various regression algorithms in machine learning. In-order to select the prediction methods they compared and explored various prediction methods. To predict the future price, the previous market trends, price ranges and also upcoming development will be analyzed.

In paper [5] they present the model which predict the price of house by considering various features which are given the datset.in this model they used random forest regression to train the dataset also they added new features like tax, air quality etc.

In paper [6] they examines the SVM algorithm in machine learning for house price prediction. It takes data from users, processes it, classifies it using pre-available data, classifies the data using various classification algorithms, and predicts the exact price of a property.

In paper [7] they discussed about the different models for predicting house prices. We compare three different types of machine learning, including Random Forest, XGBoost, and LightGBM, and two machine learning techniques, including Hybrid Regression and Stacked Generalized Regression, to analyze the best solution. Even with, all of these methods gave desirable results, but each model has its own strengths and weaknesses. The random forest method has the lowest error on the training set, but is prone to overfitting. High time complexity as data record needs to be adjusted several times.

In paper [8] they presented The study about predicting house prices based on NJOP houses in Malang city with regression analysis and particle swarm optimization (PSO). Regression analysis is used to identify the ideal coefficient in prediction, and PSO is used to choose the affected factors. The findings of this study demonstrated the suitability of combination regression and PSO, obtaining IDR 14.186 as the lowest possible prediction error. For the other model, the error prediction values are still large. Future studies will employ a variety of techniques that are compatible with time-series data to produce fewer error prediction values and more accurate results.

In paper [9] they discussed about the concept of machine learning and its numerous applications are presented in general terms. Utilizing Future applications of the data modeling and analysis in this work in lodging value-prediction systems are possible. machine learning techniques like clustering and regression to examine the quality of the output or solution, prices for homes have been projected based on a variety of factors. Python has also been used for multicollinearity reduction using VIF to remove mistakes and redundancy from the dataset. A high level of accuracy can be achieved with the correct applications using ML techniques.

In paper [10] they explain that Data is the building block of Machine learning and using this we can predict the house prices using Leaner Regression, Decision Tree Regression, K- Means Regression and Random Forest Regression. Earlier manual prediction was done for predicting the house prices but now various methods have been introduced as many researchers proposed different methods of algorithms based on resale price of houses, neural networks, hybrid regression techniques, etc. This model is divided in particular phases through which collectively we can achieve accurate house price prediction model which includes Data Processing, Looking for correlations, To fill missing attributes, Fitting the model. In the process of developing this model , it is trained on Linear regression, Decision tree, K-mean and Random forest algorithms. Out of this Random Forest give a highest accuracy in prediction of housing prices.

In paper [11] they explain the exact procedure that how to analyze the dataset and how to find the correlations between the parameters, they used four algorithms in their house price model i.e. logistic regression, support vector regression, lasso regression and decision tree. they compared the result of each algorithm and conclude that Decision Tree algorithm gives highest accuracy.

In paper[12] they discussed about the LASSO Regression, Elastic Net Regression grading boosting regression after they compared the accuracy of all these regression algorithm and they

conclude gradient boosting algorithm has high accuracy and they used MSE and RMSE to calculate the accuracy of algorithm on the dataset.

In paper [13] they discussed about all the regression algorithms and conclude that decision tree algorithm has highest accuracy for their dataset.

In paper [14] they discussed about the convolutional neural network algorithm for the prediction of house prices also they discussed about the deep learning algorithm LSTM and GRU to predict the price of house correctly. In paper [15] they distinguish the features and used the CNN algorithm to predict the accurate the price of house effectively. In paper[16]they discussed about the Emotions classification System by using deep learning also they used CNN-LSTM architecture for the classification.

## III. METHODOLOGY

We collected the dataset of Bengaluru which consist of 13320 records with 40 parameters out of 40 parameters we choose 8 parameters. Parameters such as area type, location, society, bathroom, balcony etc. after that we preprocess the data in the dataset we found that there are no missing values in dataset. we used python and it's libraries to apply model to our dataset.

The following are the libraries that we used for our model:

- Pandas is a data analysis program.
- NumPy is a Python package that performs mathematical operations.
- Matplotlib is a library for graphing datasets.
- Scikit Learn how to use statistical modelling, including classification, to solve problems.

We construct an algorithm based on Linear regression method to predict house prices. To train and test the parameters of this Linear regression model, we applied dataset for predicting house prices

- Regression:

If we have to establish a relationship between dependent and independent variables then the best thing we need to do is analyze the regression. basically the dependent variability continues naturally. In this project we have used line data so line deployment is the best option for our model. In the case of linear regression, the relationship between the dependent variations and the independent variations that we provide for the model, that relationship is linear by nature.

### a. Attributes

In terms of house prices we have certain advantages.

i) Location

The most important factor in determining the price of a house is its location. In his research, we found the importance of local factors in determining home prices. A fixed location attribute is used to separate the location.

ii) Structure

A building element is an element that people can recognize, such as the number of bedrooms and bathrooms, the amount of floor space etc. These features of the building, often promised

by real estate agents or developers to lure potential buyers, thus fulfilling potential buyers' desires.

## b. System Architecture

In this model we take input from user as text and then data pre-processing will start in the dataset so

We collected the dataset of Bangalore House Price which is structured and well categorized.

In Data pre-processing, we clean up the data from the dataset and fill the missing values from the dataset.

We classify the data into training and testing after extracting the required information from the dataset.

In Training and Testing of model, data in dataset is broken down into two parts that is training and testing. We use 80% of our data in dataset for training and remaining 20% for testing. Training of model is done by using various algorithms of machine learning for getting predicted output. Out of these we used linear regression which help us to analyse the price with different parameters using graphs.

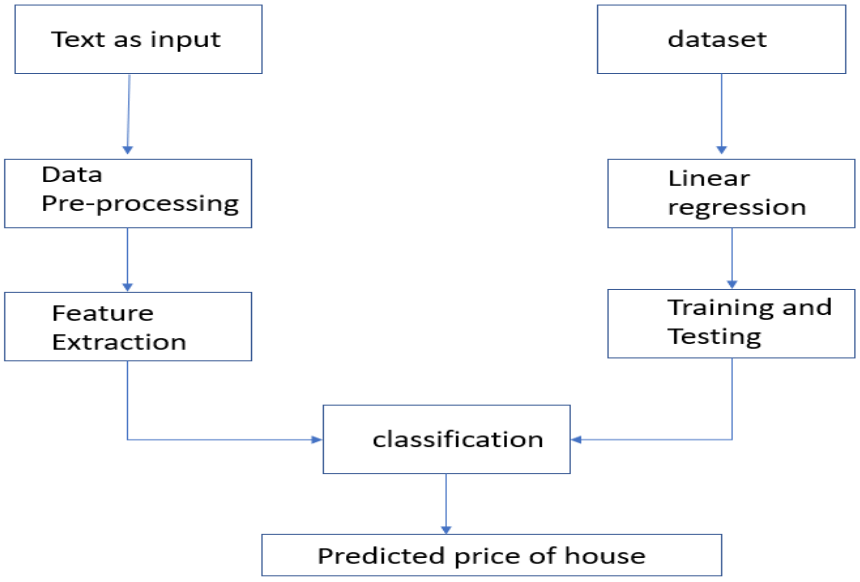



Figure 1: system architecture diagram

# IV. RESULTS AND DISCUSSIONS

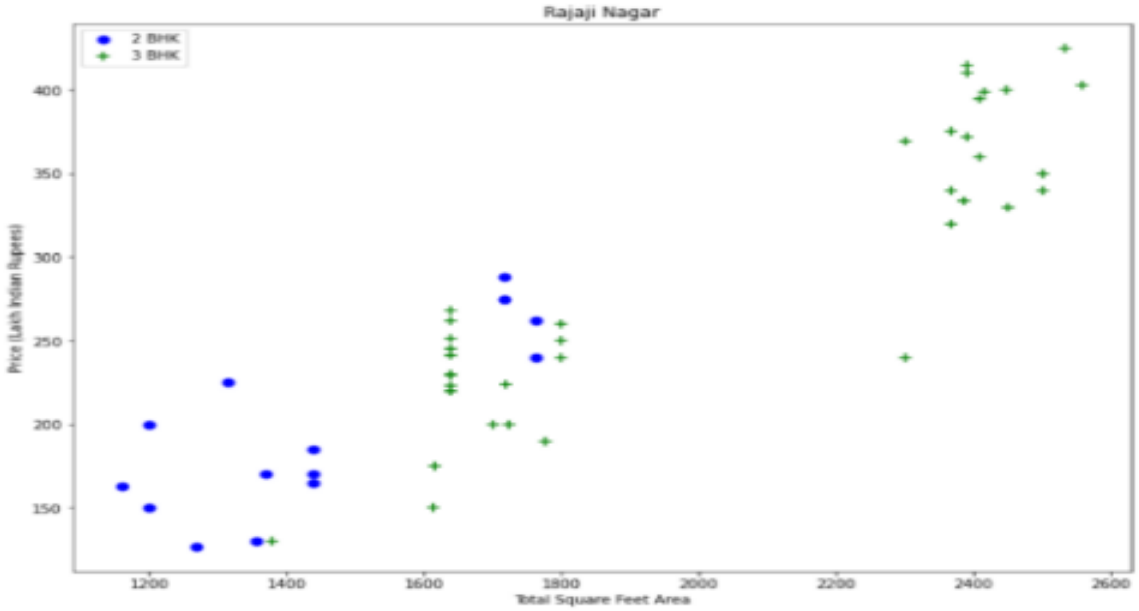
In this section, the relationship between house price and forecasting model has been investigated. In addition, the effect of different attributes on a particular model was assessed and shared with the database. After that data preprocessing and data analysis was happened to fill the missing values In the dataset and to remove the outliers from it. after doing operations on various regression algorithm we found that linear regression algorithm is the best with accuracy of 87.60%.

| | area_type | availability | location | size | society | total_sqft | bath | balcony | price |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Super built-up Area | 19-Dec | Electronic City Phase II | 2 BHK | Coomee | 1056 | 2.0 | 1.0 | 39.07 |
| 1 | Plot Area | Ready To Move | Chikka Tirupathi | 4 Bedroom | Theanmp | 2600 | 5.0 | 3.0 | 120.00 |
| 2 | Built-up Area | Ready To Move | Uttarahalli | 3 BHK | NaN | 1440 | 2.0 | 3.0 | 62.00 |
| 3 | Super built-up Area | Ready To Move | Lingadheeranahalli | 3 BHK | Soiewre | 1521 | 3.0 | 1.0 | 95.00 |
| 4 | Super built-up Area | Ready To Move | Kothanur | 2 BHK | NaN | 1200 | 2.0 | 1.0 | 51.00 |
| 5 | Super built-up Area | Ready To Move | Whitefield | 2 BHK | DuenaTa | 1170 | 2.0 | 1.0 | 38.00 |
| 6 | Super built-up Area | 18-May | Old Airport Road | 4 BHK | Jaades | 2732 | 4.0 | NaN | 204.00 |
| 7 | Super built-up Area | Ready To Move | Rajaji Nagar | 4 BHK | Brway G | 3300 | 4.0 | NaN | 600.00 |
| 8 | Super built-up Area | Ready To Move | Marathahalli | 3 BHK | NaN | 1310 | 3.0 | 1.0 | 63.25 |
| 9 | Plot Area | Ready To Move | Gandhi Bazar | 6 Bedroom | NaN | 1020 | 6.0 | NaN | 370.00 |

Figure 2: dataset of house price prediction model

When the code is applied, it generates the first exit sites, followed by forecasts. These graphs help to understand the relationship between target variables (value) and various predictive variables. The following are some examples of representations in other areas (area):

Figure 3.1: Total Square Feet Area Vs Price (Lakh Indian Rupees) of Rajaji Nagar




In figure 3.1,we can observe that dataset is not trained so it was unable to predict the accurate price of house but after training the dataset we can observe graph which was given in figure 3.2.by observing that graph we can say now our model is trained.

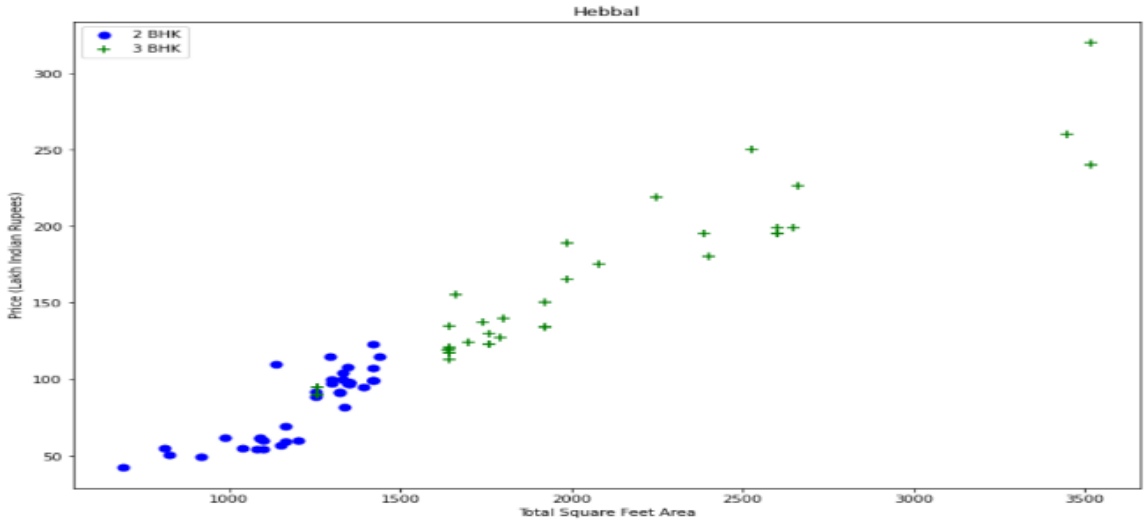



Figure 3.2: Total Square Feet Area Vs Price (Lakh Indian Rupees) of hebbal

In this model, user can give input various parameters like area, location, bathrooms, bedrooms etc. on the basis of that user get the predicted price of house. In figure 3.3 you can see the predicted price of house by considering the different parameters.

```
location=str(input("location:"))
sqft=float(input("sqft:"))
bath=float(input("bath:"))
bhk=float(input("bhk:"))


def predict_price(location,sqft,bath,bhk):
    loc_index = np.where(x.columns==location)[0][0]

    x = np.zeros(len(x.columns))
    x[0] = sqft
    x[1] = bath
    x[2] = bhk
    if loc_index >= 0:
        x[loc_index] = 1

    return lr.predict([x])[0]
```

```
location:Electronic City Phase II
sqft:1020
bath:3
bhk:2
```

```
[59] c=predict_price(location,sqft,bath,bhk)
     print("predicted price of house is",c,"lakh rupees")
      #predict_price(,1440,2,3)

     predicted price of house is 35.91054164675815 lakh rupees
```

Figure 3.3: Result

## V. CONCLUSION

A paper entitled “House Price Prediction using ML” presented to predict price of house based on various factors from the data provided. We have used Linear regression algorithm to train the dataset that helps to predict accurate price of house. To test the attributes and parameters of this Relocation model we used a database to predict house prices. This model helps people to buy houses at reasonable prices and is not deceived by the seller or agents.

## VI. FUTURE SCOPE

The purpose of this model is to present the prices of the most valuable home to every passerby. Future prices can be predicted by considering past market trends and price categories, as well as future developments.

## VII. ACKNOWLEDGMENT

The authors would like to express special thanks to our teacher, who gave us the opportunity to work on this project. Her efforts was greatly appreciated because of whom many new terms came to be known for which the entire team is thankful.

## VIII. REFERENCES:

[1] Smith Dabreo , Shaleel Rodrigues , Valiant Rodrigues , Parshvi Shah "Real Estate Price Prediction "ISSN: 2278-0181,Issue 04, April-2021

[2] Visit Limsombunchai, Christopher Gan and Minsoo Lee,"House Price Prediction: Hedonic Price Model vs. Artificial Neural Network", American Journal of Applied Sciences 1 (3):193-201, 2004, ISSN 1546-9239, 193 – 201.

[3] ALAN IHRE ISAK ENGSTRÖM "Predicting house prices with machine learning methods", June 6,2019.

[4] Bindu Sivasankar,Arun P.Ashok,Gouri Madhu,Fousiya S. "House Price Prediction", jul 31,2020.

[5] Anand G. Rawool ; Dattatray V. Rogye ; Sainath G. Rane ; Dr. Vinayak A. Bharad" House Price Prediction Using Machine Learning", May 8,2021.

[6] Mr. Piyush Chordia, Mr. Pratik Konde, Ms. Supriya Jadhav, Hrutik Pandhare, Prof. Shikha Pachouly "Prediction of House Price Using Machine Learning"

[7] Quang Truong, Minh Nguyen, Hy Dang, Bo Mei "Housing Price Prediction via Improved Machine Learning Techniques" (IIKI2019).

[8] Nikita Malik, Vidhu Gaba and Priyansh "Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization."

[9] Adyan Nur Alfiyatin, Hilman Taufiq, Ruth Ema Febrita and Wayan Firdaus Mahmudy "Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization", Vol. 8, No. 10, 2017.

[10] ANAND G. RAWOOL1 , DATTATRAY V. ROGYE2 , SAINATH G. RANE3 , DR. VINAYK A. BHARADI4 "House Price Prediction Using Machine Learning" MAY 2021,IRE Journals ,Volume 4 Issue 11,ISSN: 2456-8880

[11] NEELAM SHINDE and KIRAN GAWANDE "VALUATION OF HOUSE PRICES USING PREDICTIVE TECHNIQUES" ISSN: 2393-2835, Volume-5, Issue-6, Jun.-2018.

[12] CH.Raga Madhuri, 2 Anuradha G, 3 M.Vani Pujitha "House Price Prediction Using Regression Techniques: A Comparative Study" 978-1-7281-0027-2/19/$31.00 ©2019 IEEE.

[13] 1 P. Durganjali, 2 M. Vani Pujitha "House Resale Price Prediction Using Classification Algorithms"978-1-7281-0027-2/19/$31.00 ©2019 IEEE.

[14] Choujun Zhan, Zeqiong Wu, Yonglin Liu, Zefeng Xie, Wangling Chen "Housing prices prediction with deep learning: an application for the real estate market in Taiwan" 978-1-7281-4964-6/20/$31.00 ©2020IEEE.

[15] Yong Piao, Ansheng Chen, Zhendong Shang "Housing Price Prediction Based on CNN" 978-1-7281-2106-2/19/$31.00 @2019 IEEE

[16] Puja A. Chavan and Dr. Sharmishta Desai "A Review on BCI Emotions Classification for EEG Signals Using Deep Learning" © 2021 The authors and IOS Press.