

# Real Estate Price Prediction System

Raj Gudhka<sup>1</sup>, Karan Rathod<sup>1</sup>, Manish Parmar<sup>1</sup>, Dr. Sunil Karamchandani<sup>2</sup>

<sup>1</sup> U. G. Student, Department of Electronics and Telecommunication, D.J.Sanghvi College of Engineering, Vile Parle (W), Mumbai - 400056

<sup>2</sup> Associate Professor, Electronics and Telecommunication Department D.J.Sanghvi College of Engineering, Vile Parle (W), Mumbai - 400056

E-mail: [rajgudhka104@gmail.com](mailto:rajgudhka104@gmail.com), [karanrathod2834@gmail.com](mailto:karanrathod2834@gmail.com), [manish2580parmar@gmail.com](mailto:manish2580parmar@gmail.com), [sunil.karamchandani@djsce.ac.in](mailto:sunil.karamchandani@djsce.ac.in)

**Abstract:** In this research, we tend to make predictions about the value of houses purchased after being evaluated using machine learning algorithms. Numerous factors, such as the property's size, location, building materials, Real estate markets, time of the year, lifestyle, length of stay, number of bedrooms and so on, affect how much a home is worth. In this research paper, the Real Estate Price Prediction Model for houses is created by machine learning methods.

**Keywords:** Dataset, Machine Learning, Linear Regression Model.

## I. INTRODUCTION

These days, real estate property not only satisfies a person's basic needs but also serves as a symbol of their position and riches. Real estate investments often appear beneficial as property values don't drop sharply. Variations in real estate prices will impact a multitude of stakeholders, including lenders, insurance companies, and house investors. Investing in the real estate market seems like a tempting substitute for other forms of investing. Predicting the significant estate price is hence a critical economic indicator.



Fig.1 Factors that affect housing affordability

Asia comes in second place in the globe with a variety of 24.67 crore homes, according to the 2011 census. Asian economies are also the fastest growing major economies, surpassing even China's, which grew at a pace of 7.2 percent this year and is predicted to rise at a rate of seven percent next year. The top cities for development and investment are city and metropolis, according to the 2017 edition of emerging Themes in Realty Asia Pacific. These cities have taken the place of Sydney as the national capital. Using the National Housing Bank's Residential Index (Residex), the housing expenses of twenty-two towns out of twenty-six born within the period from April to the Gregorian month of April are

compared to the quarter from Jan to March. A wider spectrum of investors have an interest in real estate now that the Benami Property Act and the Real Estate Regulation Development Act have been introduced throughout the Asian nation. India has become a desirable place to invest due to the nation's economy's modernization and strengthening throughout Asia. But historical recessions demonstrate that real estate prices are fundamentally fixed. The state's economic circumstances are linked to the expenses of the significant estate property. In spite of this, we frequently lack appropriate, regulated methods for sustaining significant estate property values.

## II. LITERATURE SURVEY

The project aims to build a Machine Learning Model for predicting House prices. Real Estate in this today's age has become more than a basic requirement. Not just for those considering purchasing real estate, but also for the businesses that market these properties. India's Real Estate industry is expected to reach a value of US\$48.24 trillion by 2028 in India, growing at an annual pace of 4.19% between 2024 and 2028. Real Estate Price fluctuations can have an impact on a wide range of people who are household investors, bankers, legislators and many more. It appears that investing in the real estate industry is a desirable option. Thus, Artificial Intelligence and Machine Learning techniques were developed and an algorithm was developed that uses specific input features to forecast house prices. This algorithm's business purpose is that classified websites can use it to immediately forecast prices of upcoming properties to be advertised by using a few input variables to determine the accurate and reasonable pricing.

## III. PROPOSED DESIGN

The goal of this system is to determine the price of a property by looking at the many criteria that the user provides. After receiving these features, the machine learning model predicts something based on how these elements affect the label. The first step in achieving this will be to locate a good dataset that satisfies the needs of both developers and users. The dataset will also go through a process known as data cleaning after it has been chosen, which involves removing any extraneous data and converting the original data into a CSV file. The data will also go through data preparation, which entails label encoding as well as addressing missing data

necessary. Additionally, this will undergo data transformation to become a NumPy array, which will allow it to be transmitted for model training at the end. A final algorithm and model that can produce precise predictions will be developed by extracting the error rate of the numerous machine learning techniques that were utilized for training the model. After logging in, users and businesses may fill out a form with various property qualities for which they wish to get an estimate for a price. In addition, the application will be sent in following a careful selection of qualities. The user will then be able to examine the estimated cost of the property they provided in a matter of seconds once this data has been loaded into the model.

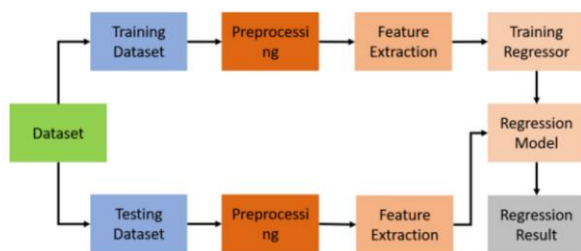


Fig.2 Block Diagram of the prediction model

The block diagram above illustrates the traditional machine learning methodology. Its two primary components are the evaluation and the training. The components of the training include the label, inputs, features extractor, and the machine learning algorithm. The components that comprise the testing portion are the input, features extractor, predictive model, and output label.

**Information:** The information is gathered from several sources.

**Feature Extraction Tool:** Only significant elements that have an impact on the prediction outcomes are retained. Other unnecessary features, such as name or ID, are removed.

**Features:** Only certain inputs that significantly influence the model's prediction are taken into account after feature extraction.

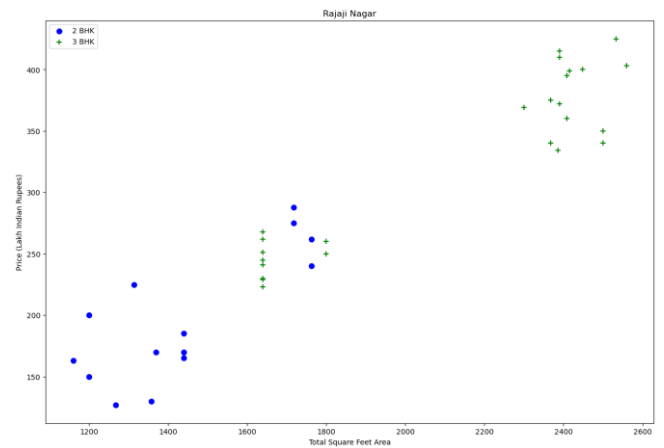


Fig.3 Plot of Total square feet area v/s Price

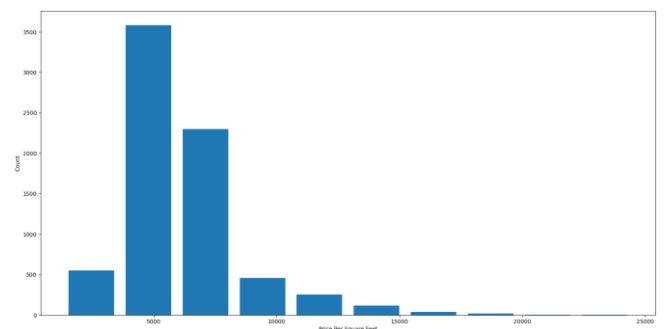


Fig.4 Plot of Price per sqft v/s Count

**Machine learning algorithms:** To accomplish their goals, AI systems make use of machine learning algorithms. The most common application of it is for predicting output values from supplied values for input. The Linear Regression is a fundamental process in machine learning. The Regression Model by utilizing an array of machine learning methods, the regression model allows us to foresee a label variable (y) based on the values for any number of attribute/feature factors (x). The primary goal of a model based on regression is to develop an equation in mathematics that represents y as an expression of the model's x variables.

**Label:** The output that the model produces following training is the label.

First, the training input is the dataset's data, from which appropriate training features are collected. To obtain a normalized dataset, these training features are preprocessed, and the data row is labeled. The algorithm for machine learning receives the output from the training dataset. Regression model receives the output from a Machine Learning. Using an algorithm, a trained regressor or model is created. Given the new data—that is, the retrieved feature from the test dataset—this regressor may predict the output label after training.

Once the model is ready, the user will input every feature or characteristic of the home they want to get a pricing estimate for. Moreover, all attributes will be tested to see if they correspond to the same type of data as required once the user hits submit. This will check all attributes for null values. The data will then be transmitted for prediction and the projected price will be shown to the user on the website if all requirements have been met.

#### IV. IMPLEMENTATION

We have used Jupyter Notebook for this Machine Learning Project. The Jupyter Notebook is the online web application which is also open source that helps us to code live, create and share it. It offers an easy-to-use, document-centric experience. The user will input their criteria to receive the pricing of the residences they are interested in. In order to obtain references for houses, users may also obtain a sample blueprint of the house. It indicates that the purchase prices alone, without taking additional factors into account for the model, can be used to predict the selling prices. We took a Real Estate dataset for Bangalore City from Kaggle website and built a ML Model. Data Cleaning, Feature Engineering, One Hot Encoding, Outlier Detection, Dimensionality Reduction, GridSearchCV were the core concepts used. Pandas is used for Data Cleaning, Python is the programming language which is used majorly in the project, Matplotlib is used for Data Visualisation, Sklearn is used for model building. Area (in sqft), BHK, Bath and Location are the inputs and eventually the predicted price will be the output. K-Fold Cross Validation and GridSearchCV to come up with the best algorithm and best parameters.

#### V. RESULTS AND CONCLUSION

We have predicted the expenses of the house using a machine learning algorithm in our study report. We have discussed the methodical approach to investigating the dataset and determining the degree of connection among the various parameters existing. Therefore, we often compute each model's performance using entirely distinct performance criteria and compare them using these metrics. In our opinion, using a large dataset for future work would produce a more accurate and realistic representation of the model.

```
In [81]: predict_price('1st Phase JP Nagar',1000, 2, 2)
Out[81]: 83.49904677167721

In [85]: predict_price('1st Phase JP Nagar',1000, 2, 3)
Out[85]: 81.72616900743024

In [83]: predict_price('Indira Nagar',1000, 2, 2)
Out[83]: 181.27815484007024

In [84]: predict_price('Indira Nagar',1000, 3, 3)
Out[84]: 184.58430202033577
```

Fig.5 Final output of the prediction model(in lakhs)

#### VI. SOCIAL IMPACT & FUTURE

The social effect of real estate price prediction algorithms is complex, affecting dynamics of community development, investment patterns, and housing affordability. These systems can improve market transparency and give useful information for consumers, sellers, and policymakers by utilizing sophisticated algorithms and large databases. To guarantee fair results, however, issues with confidentiality of information, algorithmic biases, and disparities in socioeconomic status have to be addressed. These technologies could be essential in the future for creating more sustainable and inclusive urban environments, encouraging balanced growth, and resolving the housing crisis. To minimize any unfavorable effects and advance equity in real estate transactions, ethical standards and legal frameworks will be crucial. In the end, the development of such prediction systems has the potential to improve economic efficiency, empower people, and create vibrant communities in the actual world.

#### ACKNOWLEDGMENT

Dr. Sunil Karamchandani contributed a great and thorough support to this project and technical paper. His expert guidance and extensive help were indispensable for us in all aspects of the present work. In the research, we are thankful to our colleagues who helped us along the way and provided their expertise and knowledge to the study.

#### REFERENCES

- [1] Dabreo, S., Rodrigues, S., Rodrigues, V., & Shah, P. (2021). Real estate price prediction. *International Journal of Engineering Research & Technology (IJERT)*, 10(04), 2278-0181.
- [2] Sayan Putatunda, "PropTech for Proactive Pricing of Houses in Classified Advertisements in the Indian Real Estate Market".
- [3] Bindu Sivasankar, Arun P. Ashok, Gouri Madhu, Fousiya S, "House Price Prediction", *International Journal of Computer Science and Engineering(IJCSE)*, Vol: 8 Issue: 7, 2020.

- [4] Chaturvedi, S., Ahlawat, L., Patel, T., & Talha, M.  
(2021). Real Estate Price Prediction. *EasyChair*, 4926.