iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VI **Month of publication:** June 2023

DOI: https://doi.org/10.22214/ijraset.2023.53841

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

House Price Prediction Using Machine Learning

Deepanshu Patel[1], Parmeshwar Nayak[2], Shubham Gupta[3], Mr. Jayanth C[4], (Guide)
Electronics and Telecommunication, Dayananda Sagar College of Engineering, (Autonomous VTU-Affiliated), Bangalore, India

Abstract: ***Forecasting the appropriate house pricing for real estate customers while taking into consideration their priorities and financial situation is the goal. By examining previous market patterns, price ranges, and approaching changes, future prices may be predicted. Research indicates that house price discrepancies are a common source of concern for both homeowners and the real estate industry. Several interrelated factors influence the price at which real estate sells in places like Bengaluru. The size, location as well as and amenities of the property are significant considerations that could have an impact on the price. The analysis' findings supported the use of boosting algorithms like Extreme Gradient Boost Regression (XG Boost), Support Vector Regression, Multiple Linear Regression (Least Squares), and Machine Learning Lasso and Ridge regression models among other regression techniques in modelling explorations.***

Keyword: Machine Learning, Decision tree, House Price Prediction, Regression Methods

I. INTRODUCTION

One of man's most basic needs is a place to call home, along with other basic essentials like water, food, and many other things. As people's living conditions improved throughout time, the demand for housing rose fast. Although some individuals buy homes as real estate or investments, the majority of people buy homes for occupancy or as a source of support. Modelling uses machine learning techniques, which allow computers to acquire information from data and forecast new data. The predictive analytic model that is most often used is regression. The demand for housing is increasing yearly, which indirectly raises property prices. The problem occurs when a number of parameters, including location and housing demand, may influence the price of a home. As a result, the majority of stakeholders involved, including buyers, developers, home builders, and the real estate sector, want to know the precise traits or the accurate factors influencing the house price to help investors in their decisions and to assist home builders set the house price. Multiple linear regression is one statistical technique for establishing a relationship between the dependent target variable and various kinds of independent factors. All of these point to the developing field of regression research known as "house price prediction" which calls for machine learning expertise. This inspired us to work in this field. The five regression models that we have considered are ordinary least squares, SVR, Lasso and Ridge, and XGBoost. Furthermore, a comparative research utilizing assessment methods was carried out. Once the model and data are reasonably aligned, we may use it to forecast the financial worth of that particular City housing property.

II.PREVIOUS RELATED WORK

The two types of predictions of house prices are those that emphasize the characteristics of the home and those that emphasize the model that is employed to calculate the value of the house. A number of researchers have developed a model for predicting housing prices.

1) For forecasting the property's pricing value, Nor Hamizah Zulkifley, Shuzlina Abdul Rahman, Nor Hasbiah Ubaidullah, and Ismail Ibrahim employed four regression techniques: Multiple Linear Regression, Support Vector Regression, Artificial Neural Network, and XGBoost. With the least amount of inaccuracy, the ensemble technique forecasted prices. Numerous research have also concentrated on the methods for gathering characteristics and extracting them. M. Thamarai and S. P. Malarvizhi have contrasted several feature extraction and feature selection methods.

2) Support vector machine (SVM) regression is used by Jirong, Mingcang, and Liuguangyan to forecast China's property prices from 1993 to 2002. To fine-tune the hyper-parameters of the SVM regression model, they used the genetic algorithm. Less than 4% of the SVM regression model's error scores were recorded.

3) To predict the values of the houses, Drs. M. Thamaraia and S. P. Malarvizhi employ Decision Tree Regression along with Multiple Linear Regression techniques. The number of bedrooms (1BHK, 2BHK, and 3BHK) that are available in a house, the age of the house, the travel options from the location, such as bus, train, and flight options, and the school options nearby, such as the presence of Government schools, matriculation, and CBSE, several shopping options close to the residence, including local markets, general stores, and retail centers have all been shown to be positively and significantly correlated with the log price of houses.

©IJRASET: All Rights are Reserved

4) To determine the resale price of properties based on their key characteristics, Fan et al. used the decision tree technique. For determining the association between the prices of the properties and their important attributes, this study applies a hedonic based regression technique.

III. ATTRIBUTE SELECTION MEASURES

The train set and test data for the project were obtained through the Machine Hackathon platform. It is composed of traits that define real estate in Bengaluru. Both data sets have 9 attributes. The traits may be summed up as follows:

1) Area type: Describes the area.
2) Availability: When it is possessed or when it is prepared.
3) Price: The property's value in lakhs.
4) Size: In BHK or Bedroom (1-10 or more).
5) Society: To which it belongs.
6) Total sqft: Property's square footage measurement.
7) Bath: Number of bathrooms
8) Balcony: Number of balconies
9) Location: Where in Bengaluru it is located.

IV. SYSTEM DESIGN AND ARCHITECTURE

1) *Phase I: Data collection*

Information that has been appropriately organized and categorized must be gathered. Every machine learning challenge starts with the need for data. Data analysis would be pointless without the legitimacy of the data set. Data collection is the deliberate process of gathering information about variables. It supports the search for solutions to issues, generates a lot of hypotheses, and evaluates results. It is now possible to resolve related issues and assess the results by using data collection to support social interactions and estimate data on specific parameters within the already existing framework.

2) *Phase II: Pre-processing of data*

This stage involves cleaning up our data. There might be missing values in our dataset. There are three different ways to complete our missing values: 1) Eliminate the missing data points. 2) Remove the entire attribute. 3) Set the value

3) *Phase III: Training the model*

This phase separates the data into two categories: training and testing. An estimated 80% of the information is utilized for training, while the remaining 20% is utilized for testing. The training set contains the target variable. The model is trained using a variety of machine learning approaches to provide the results. These Random Forest Regressions yield more encouraging results.

4) *Phase IV: Model testing*

The trained model is then used to estimate home prices on a test dataset.

V. METHODOLOGY

A. *Linear Regression*

A statistical technique known as simple linear regression enables us to analyse and explore relationships among two continuous parameters:

1) One variable is considered to be either a predictor, explanatory, or independent variable and is indicated by the letter x.
2) The other variable, indicated y, is known as a response, result, or dependent variable.

B. *Lasso Regression*

LASSO stands for Least Absolute Shrinkage and Selection Operator.

Lasso regression is a regularization technique. For a more precise forecast, it is preferred over regression techniques. Shrinkage is used in this model. When data values shrink towards the mean, this is referred to as shrinkage.

The L1 regularization process used by Lasso regression results in a penalty equivalent to the absolute value of the magnitude of coefficients.

A sparse model with few coefficients may be produced by this kind of regularization; certain coefficients may go to zero and be removed from the model. Greater penalties provide coefficient values that are closer to zero, which is great for creating simpler models.

C. RMSE

When training regression models or time series models, one of the most often used metrics is called root-mean-square error, or RMSE. It is used to determine how accurately our forecasting model predicts values compared to the real or observed values.
It is also a crucial consideration for narrowing down the forecasting models that perform the best among those that may have been trained on a certain dataset. To achieve this, just compare the RMSE values of each model, then choose the one with the lowest RMSE value.

D. SVR (Support Vector Regression)

SVR aims to fit the error into a defined range, unlike the conventional linear regression where the goal is to lower the error. Regression analysis is performed using this approach, which works in a similar way to Support Vector Machines (SVM). Real values in a continuous range make up the data from regression. While taking into consideration the model complexity and error rate, the SVR model approximates the optimal values with a predetermined margin known as a tube (epsilon-tube, which specifies a tube width).

E. XGBoost Regression Model

The best approach for classification or regression issues is extreme gradient boosting, also referred to as XGBoost. Decision trees serve as the foundation for this gradient boosting-based method. It provides features that significantly affect the performance of the model. This technique helps to produce models that are more stable and have reduced variation. Additionally, the execution time is quite speedy in comparison to other algorithms.

VI. FLOW CHART

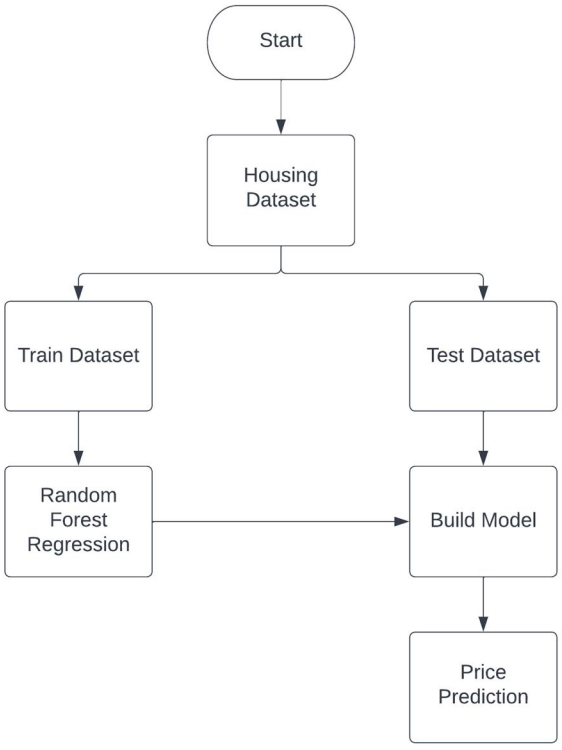


Figure 1: Flow chart

VII. CONCLUSIONS AND FUTURE SCOPE

In this paper, decision tree classifier, decision tree regression, and multiple linear regression are some of the most fundamental machine learning techniques that are employed. In this work, a machine learning programme called Scikit-Learn was employed.
The availability and price of houses in the city may be predicted by users using this work. Multiple linear regression and decision tree regression were the two methods used to predict home prices. Multiple linear regression is proven to function in comparison to the decision tree regression for predicting property values. The information might eventually include the house and additional features.

VIII. RESULT

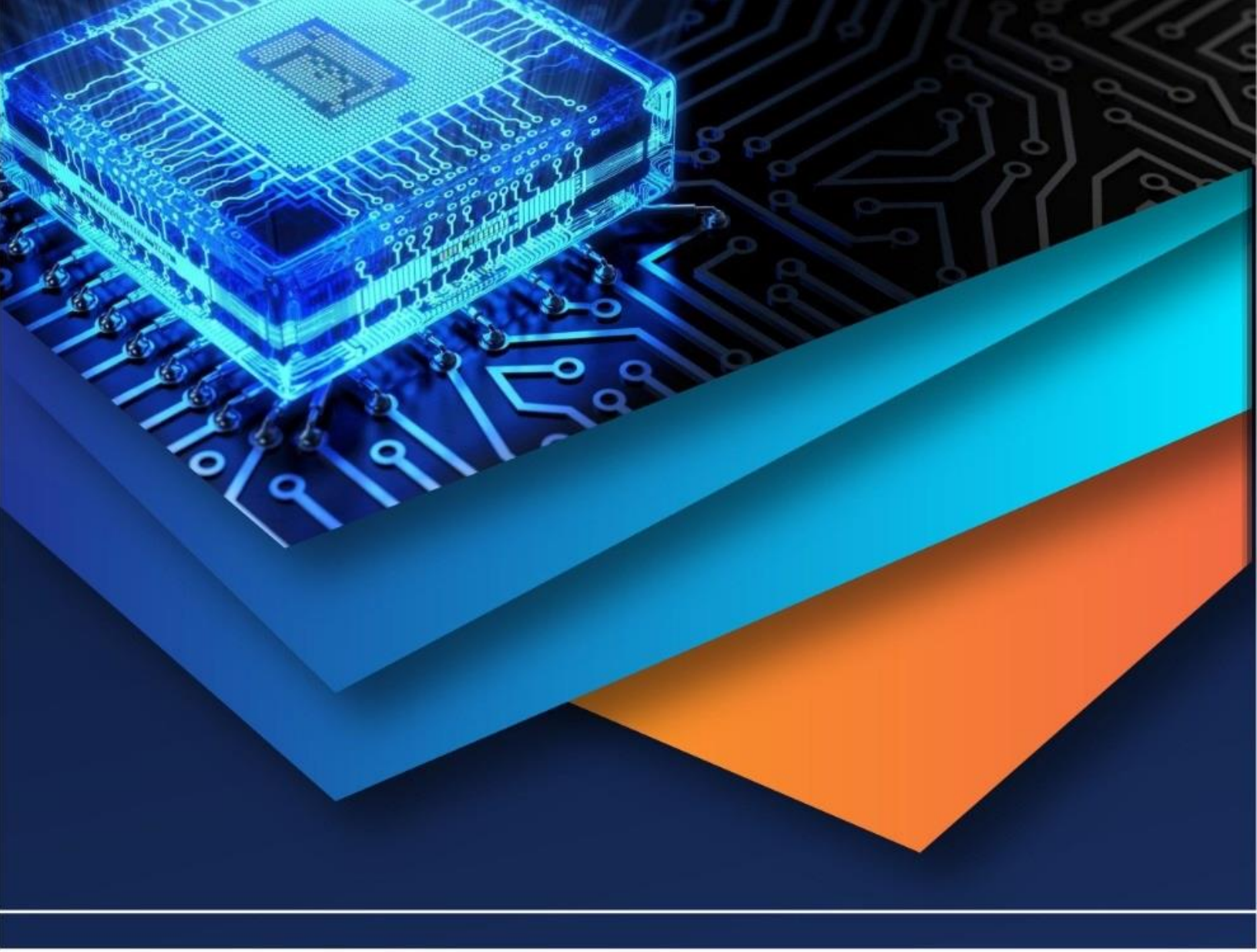
Depending on the dataset, model, and evaluation standards used, the results of using machine learning to predict housing prices can differ. Important outcomes include the ability to predict prices with accuracy, the identification of noteworthy features, the avoidance of over- or underfitting, the comparison of models, and insights into the factors affecting pricing. Careful analysis and interpretation are necessary in order to create reliable predictions and prudent actions.



Figure 1 Result

REFERENCES

[1] Nor Hamizah Zulkifley, Shuzlina Abdul Rahman, Nor Hasbiah Ubaidullah, Ismail Ibrahim, " House Price Prediction using a Machine Learning Model: A Survey of Literature", International Journal of Modern Education and Computer Science.
[2] Gu Jirong, Zhu Mingcang, and Jiang Liuguangyan. (2011). “Housing price based on genetic algorithm and support vector machine”. In: Expert Systems with Applications 38 pp. 3383-3386.
[3] M Thamarai, S P Malarvizhi, " House Price Prediction Modeling Using Machine Learning", International Journal of Information Engineering and Electronic Business(DJIEEB), VoL12, No.2, pp. 15- 20, 2020. DOI: 10.5815/ijieeb.2020.02.03
[4] G.-Z. Fan, S. E. Ong, and H. C. Koh, "Determinants of house price: A decision tree approach," Urban Studies, vol. 43, pp. 2301-2315, 2006.
[5] G.Naga Satish, Ch.V.Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu “House Price Prediction Using Machine Learning”. IJITEE, 2019.
[6] N. N. Ghosalkar and S. N. Dhage, "Real Estate Value Prediction Using Linear Regression," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-5.
[7] Ayush Varma, Abhijit Sarma, Sagar Doshi, Rohini Nair - “Housing Price Prediction Using Machine Learning and Neural Networks” 2018, IEEE.
[8] G.Naga Satish, Ch.V.Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu “House Price Prediction Using Machine Learning”. IJITEE, 2019.
[9] CH. Raga Madhuri, G. Anuradha, M. Vani Pujitha -” House Price Prediction Using Regression Techniques: A Comparative Study” 2019 in (ICSSS), IEEE.
[10] Alisha Kuvalekar, Sidhika Mahadik, Shivani Manchewar, Shila Jawale, “House Price Forecasting using Machine Learning”, 3rd International Conference on Advances in Science & Technology (ICAST) – 2020.
[11] Lu. Sifei et al,A hybrid regression technique for house prices prediction. In proceedings of IEEE conference on Industrial Engineering and Engineering Management: 2017.
[12] CH. Raga Madhuri, G. Anuradha, M. Vani Pujitha -” House Price Prediction Using Regression Techniques: A Comparative Study” 2019 in (ICSSS), IEEE.
[13] R. Victor, Machine learning project: Predicting Boston house prices with regression in towards data science.
[14] Ong, S. E., Ho, K. H. D. and Lim, C. H., “A constantquality price index for resale public housing flats in Singapore”, Urban Studies, 40(13), 2003, pp. 2705 – 2729.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)