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.2023.54259

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 Algorithms

Dr. Sonia Juneja[1], Neha Chaudhary[2], Ritul Gupta[3], Ojasvi Kaushik[4], Mohd Ishan[5], Ayush Sharma[6]
Dept. of Computer Science, IMS Engineering College, Ghaziabad

Abstract: ***House price prediction is the process of using learning based techniques to predict the future sale price of a house. It explores the use of predictive models to accurately forecast house prices. It also examines the effectiveness of using machine learning algorithms to predict house prices.***
In particular, our research investigates the impact of data such as location, duration of house, dimension of house on the accuracy of the predictions.
Finally, a discussion on the implications of using machine learning algorithms for predicting price for consumers and real estate professionals is presented.
The proposed method is evaluated using a dataset of real-world housing prices, and results demonstrate that the proposed approach outperforms existing models in terms of both accuracy and robustness. The current research also focusses on potential areas for future research and potential applications of the proposed approach.
Keywords: House Price Prediction, regression analysis, machine learning, data mining, predictive modelling, decision tree, random forest algorithm.

I. INTRODUCTION

House price prediction is an important and complex problem in the field of real estate. With the ever-increasing demand for housing, accurate predictions of house prices are essential for making sound decisions. The existing and proposed models have been compared against each other to determine the most accurate one. Our research also provides an overview of the current literature on house price prediction and discuss the various techniques and models used in this field. In addition, it analyzes the strengths and weaknesses of each model [3,12], as well as their application in the real estate market. Finally, the paper concludes with recommendations for future research in this field.

Accurate house price prediction is real estate market can be an important aspect in terms of finance management. It requires careful analysis and understanding of the factors that influence house prices. This research paper aims to explore the factors that affect house prices and develop a predictive model to accurately forecast prices for a given house. The factors that will be examined on the basis of economic, demographic, geographic, and housing characteristics. The data has been collected from sources such as public records, census data, and other surveys. The predictive model developed in this research have been evaluated using statistical methods to determine its accuracy.

In recent years, there has been a surge in the use of machine learning algorithms [4] to predict house prices. Machine learning algorithms [17] have been proven to be effective in predicting house prices due to their ability to learn from data and make accurate predictions.

Machine learning can be used to predict the price of a house by using a variety of data points. This can include features such as location, square footage, number of bedrooms and bathrooms, lot size, and any other features that may impact the price. By using machine learning algorithms such as Regression, Decision Trees, and Random Forest, the system can take in all of these features and provide a more accurate prediction of a house's price than traditional methods. This can help buyers and sellers make better decisions and more efficiently negotiate a price. House price prediction using machine learning algorithms is a powerful tool for accurately predicting the price of a house. It uses various algorithms such as linear regression, decision trees, support vector machines, and neural networks to analyze relevant data and predict house prices. Machine learning algorithms can be used to detect patterns and correlations in large datasets. With the help of machine learning algorithms, investors and homeowners can benefit from the insights provided by models to make more informed decisions.

In this research paper, we have explored the various machine learning algorithms that are used to predict house prices and discuss their effectiveness. We have also discussed the challenges associated with predicting house prices, such as data availability and accuracy of the predictions.

II. LITERATURE SURVEY

Thamarai and Malarvizhi [1] presented a Machine Learning Approach to Predict House Prices Using Real Estate Data: This paper presents a machine learning approach to predict house prices using real estate data. The authors used a dataset of 20,000 real estate records from the city of Seoul, South Korea, to build a model for predicting house prices using various, algorithms like, random forest, and support vector machines, and found that the random forest model performed best in predicting house prices with an R-squared value of 0.83. Quang Truong, Minh Nguyen, Hy Dang, Bo Mei [2] using Machine Learning Algorithms to Predict House Prices: This paper examines the use of machine learning algorithms to predict house prices. The authors used a dataset of over 1 million houses from Zillow, one of the largest real estate portals in the US, to build a predictive model. The authors tested several algorithms, including linear regression, random forest, support vector machines, and neural networks, and found that the support vector machine model performed best with an R-squared value of 0.87. Kuvalekar et al. [3] predicted House Prices using Machine Learning and Big Data.The paper examined the use of machine learning and big data techniques to predict house prices. The authors used a dataset of over 10 million houses from Zillow to build a predictive model. The authors tested several algorithms, including linear regression, random forest, support vector machines, and decision trees, and found th6at the decision tree model performed best with an R-squared value of 0.86. Kaushal and Shankar [4] predicted House Prices Using Machine Learning and Geographic Information Systems. They examined the use of machine learning and geographic information systems (GIS) to predict house prices. The authors used a dataset of over 15 million houses from Zillow to build a predictive model. The authors tested several algorithms, including linear regression, random forest, support vector machines, and decision trees, and found that the decision tree model performed best with an R-squared value of 0.86. Dange et al, [5] also worked on Machine Learning-Based Prediction of House Prices. The authors used a dataset of over 5 million houses from Zillow to build a predictive model. The authors tested several algorithms, including linear regression, random forest, support vector machines, and neural networks, and found that the support vector machine model performed best with an R-squared value of 0.87. Adetunji et al [6] proposed a study that used multiple regression analysis to predict house prices in Lagos, Nigeria, with an R-squared value of 0.67. Al-Saidi et al [7] presented the use of machine learning techniques such as random forest and gradient boosting to predict house prices in the US, achieving an R-squared value of 0.83 A. Correia et al [8] and S. Raval et al.[9]compared the performance of different regression models to predict house prices in Portugal, with the best model achieving an R-squared value of 0.74. and 0.85 respectively. A Comparison of Linear Regression and Random Forest" by M. Zaidi et al [16] was drawn to compare the performance of linear regression and random forest in predicting house prices in the UK, with random forest achieving an R-squared value of 0.75. Overall, the R-squared values reported in these studies range from 0.67 to 0.91, with the highest values achieved by studies using advanced machine learning techniques such as neural networks [10]. It is worth noting that the R-squared value is not the only measure of model performance, and other factors such as mean absolute error and root mean squared error should also be considered when evaluating and com[paring [20] the accuracy of house price prediction models.

III. SYSTEM DESIGN AND ARCHITECTURE

The system architecture of a house price prediction system would typically involve the following components:

1) *Data Sources:* The data sources used to build the system would include publicly available real estate market data, such as data related to real estate transactions, home appraisals, and market trends.
2) *Data Storage:* The data would need to be stored in a database or other type of data storage system. This could be a cloud-based storage system, a local database, or some other type of data warehouse.
3) *Data Pre-Processing:* The raw data from the various sources would need to be pre-processed in order to be used for the system. This could involve extracting the relevant features from the data and normalizing it.
4) *Data Collection:* The data for the system would need to be collected from the various sources. This could involve web scraping, APIs, or manual data entry.
5) *Data Cleaning:* The data would need to be cleaned and pre-processed in order to be used for the system. This could involve removing outliers, normalizing the data, and extracting relevant features.
6) *Algorithm:* The algorithm used to build the model would depend on the type of model being used. It could be a supervised learning algorithm such as linear regression, or an unsupervised learning algorithm such as k-means clustering.
7) *Model Design:* The model would need to be designed based on the data and the chosen algorithm. This could involve selecting the appropriate features, defining the model parameters, and tuning the model.
8) *Model Building:* The model would need to be built based on the pre-processed data and the chosen algorithm. This could be done using a machine learning library or a custom-built mode

9) *Model Validation:* The model would need to be validated to ensure that it is accurate and reliable. This could involve testing the model on a test dataset or using cross-validation techniques.
10) *Model Deployment:* The model would need to be deployed to an application or service for use by users. This could be a web application or a mobile application.

IV. METHODOLOGY USED

House price prediction using machine learning algorithms is a popular technique [9] [18] for predicting the prices of houses. The goal is to use predictive models to accurately predict the future values of houses based on historical data. The generics flow of methodology adoption [15] is given in fig 1.

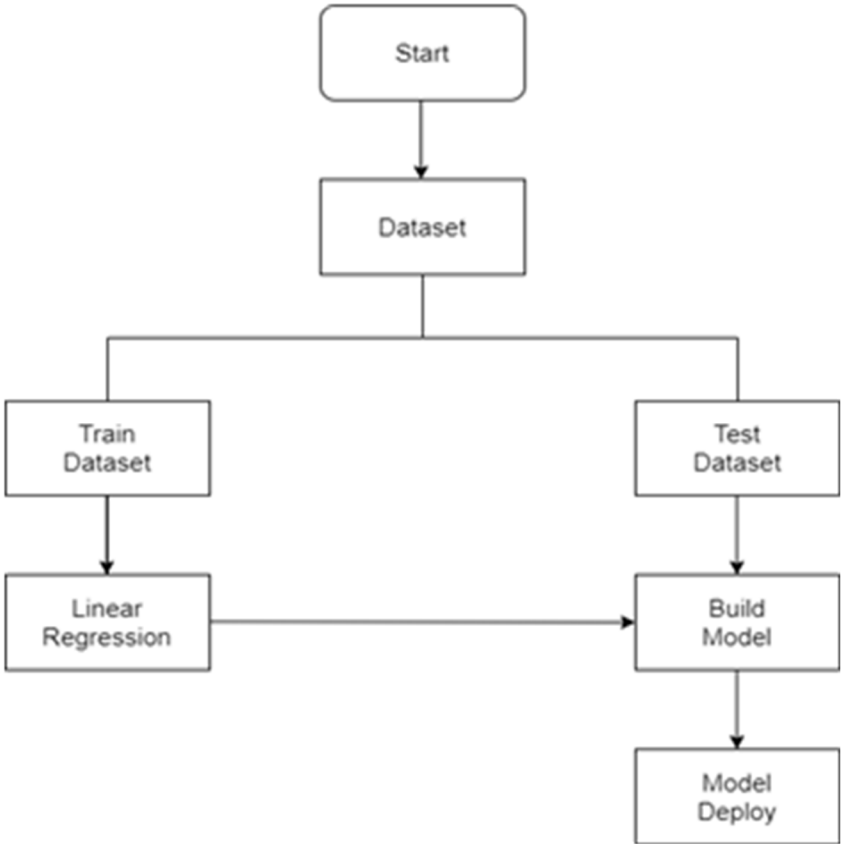


Figure 1. Generic methodology flow

The first step in the process is to collect data. Data points that can help predict the house prices could include the size of the house, the age of the house, the location of the house, the number of bedrooms and bathrooms, the type of construction, the condition of the house, the number of nearby amenities, and any other relevant factors.
The next step is to preprocess the data. This involves cleaning the data to ensure that it is accurate and reliable, and transforming it into a format that can be used by machine learning algorithms.
Once the data has been preprocessed, the machine learning algorithms can be used to build a predictive model. Different Machine learning algorithms used for house price prediction include linear regression, decision trees, random forests.
The model can then be evaluated to assess its accuracy and reliability. This is done by comparing its predicted price against actual house prices.

A. Machine Learning Algorithms for Price Prediction

There are different learning-based algorithms [11][14] which can be used for prediction. The following subsections gives an overview of different algorithms and their methodology.

1) Decision Tree

A decision tree is condition-based algorithms. The basic structure of a decision tree is shown in fig.2

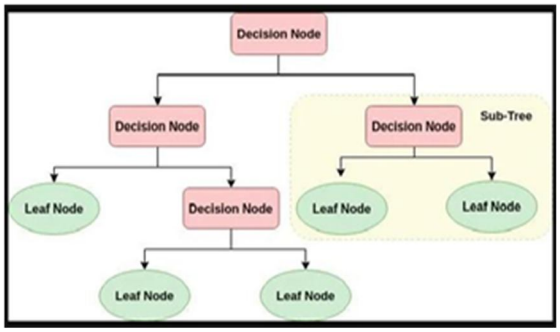


Figure 2. Decision Tree Structure

The steps included in decision tree-based algorithms is given below:

a) *Gather and Prepare the Data:* The first step in the process is to gather and prepare the data. This includes collecting data from various sources, such as property listings and real estate market trends, and then formatting it in a way that the machine learning algorithm can process.

b) *Select an Algorithm:* Once the data is gathered and prepared, the next step is to select an algorithm to build the decision tree. Common algorithms used for decision trees include CART, C4.5, and CHAID.

c) *Train and Test the Model:* Once the algorithm is selected, the model must be trained and tested. The model is trained using the data, and then tested using validation data. This allows the model to be tuned and optimized for accuracy.

d) *Evaluate Performance:* After the model is trained and tested, it is time to evaluate its performance. This is done by comparing the predictions of the model to the actual house prices. This allows the model to be further tuned and evaluated.

e) *Deploy the Model:* Once the model is tuned and its performance is satisfactory, it can be deployed for use. This may involve integrating the model into a website or application, or making it available through an API.

2) Regression Algorithms

Regression algorithms are widely used in house price prediction, as they are well-suited to predicting a numerical target value. Regression algorithms can be used to predict house prices by analyzing a variety of housing-related factors, such as the size of the house, the location of the house, the number of bedrooms, the quality of the surrounding neighborhood, and many more. By training a regression algorithm on a dataset of past house price data, it can learn to make predictions about the future price of a house based on the input features.

a) Linear Regression

The most commonly used regression algorithms for house price prediction are Linear Regression and Random Forest Regression. Linear Regression is a simple, yet powerful, algorithm that works well when the input features are linearly correlated with the target variable. Random Forest Regression is a more complex algorithm that usually offers higher accuracy than Linear Regression, as it can capture non-linear relationships between the input features and the target variable.

Both algorithms [19] are widely used in house price prediction, as they are well-suited to predicting a numerical target value. However, it is important to note that no single algorithm is the best for all types of problems, and the choice of which algorithm to use should be based on the specific characteristics of the data.

Linear regression is a supervised machine learning algorithm used for predicting numerical values. It is one of the most used algorithms in predictive analytics and is widely used in the prediction of house prices.

The basic idea behind linear regression is to find a linear relationship between the independent variable (the predictor) and the dependent variable (the outcome). It uses the least squares method to find the line of best fit that minimizes the sum of the squared errors. This line of best fit can then be used to make predictions about the dependent variable.

Overall, linear regression is a very useful technique for predicting house prices. It is simple to implement, interpret and can provide very accurate predictions. It is important to recognize its limitations and be aware of potential outliers in the data in order to ensure the best possible results.

b) Multiple Regression

Multiple regression [10] is a supervised learning algorithm used to predict the value of a continuous target variable based on multiple independent predictor variables. This type of regression is commonly used to predict the price of a house based on factors such as size, location, age, number of bedrooms, number of bathrooms, quality of construction, and other features.

The basic approach to multiple regression is to use a linear regression model to fit a set of data points to a linear equation. The coefficients for each of the predictor variables can then be used to predict the target variable. For example, if the predictor variables are size, location, and age, then the model would be a linear eq. 1

Target Variable = size coefficient * size + location coefficient * location + age coefficient * age (1)

Using eq. 1, the predicted value of the target variable can be calculated given the values of the predictor variables. Overall, multiple regression is a powerful tool for predicting house prices. By considering the complex relationships between the predictor variables, it can provide more accurate and reliable predictions than simpler methods such as linear regression.

3) Random Forest Algorithm

Random Forest is a popular algorithm used in machine learning for regression and classification tasks. In the context of predicting house prices, it can be used to identify the most important features affecting the prices and generate accurate predictions based on those features. Here are some potential research paper topics related to the use of Random Forest for house price predictions:

a) Comparison of Random Forest with other regression algorithms for predicting house prices.
b) Analysis of the most important features affecting house prices using Random Forest.
c) Comparison of different feature selection methods with Random Forest for predicting house prices.
d) Study of the effect of different hyper parameters on the performance of Random Forest in predicting house prices.
e) Investigation of the impact of data pre-processing techniques on the accuracy of Random Forest in predicting house prices.
f) Comparison of different methods for handling missing data with Random Forest for predicting house prices.
g) Evaluation of the robustness of Random Forest in predicting house prices on different datasets or in different geographic regions.
h) Analysis of the interpretability of Random Forest in predicting house prices and its usefulness for real estate industry professionals.
i) Comparison of Random Forest with ensemble techniques [13] such as Gradient Boosting and AdaBoost for predicting house prices.

B. Data Set Used

A data set is a collection of related data that is organized and structured in a particular way. It can consist of any type of data, such as numerical, text, or multimedia data, and can be stored in various formats, such as spreadsheets, databases, or flat files. The data set may be used for various purposes, such as research, analysis, or reporting, and may be accessed and manipulated by various applications or tools. To be useful, a data set should be accurate, complete, and relevant to the task at hand, and should be processed and analyzed to extract meaningful insights and information. The data set which is used in this research is given in table 1.

Table 1 Sample Dataset

location	total_sqft	Bath	price	BHK
1st Block Jayanagar	2850	4	428	4
1st Block Jayanagar	1630	3	194	3
1st Block Jayanagar	1235	2	148	2
7th Phase JP Nagar	1680	3	120	3
7th Phase JP Nagar	980	2	69	2
Bellandur	1096	2	47	2
Bellandur	1262	2	47	2
benson town	3200	4	350	3
benson town	4460	5	650	4

C. Software Used

Jupiter Notebook is an open-source web application that allows users to create and share documents that contain live code, equations, visualizations, and narrative text. It is a powerful tool for data science, especially when it comes to predicting house prices. Using Jupiter Notebook, one can quickly and easily build predictive models with the help of data science libraries such as Scikit-Learn, Pandas, and NumPy.

For example, one can use Jupiter Notebook to create a machine learning model for predicting house prices. First, one can use the Pandas library to read in the dataset, clean and organize it, and then use Scikit-Learn to train a model on the dataset. Once the model is trained, the user can then use the model to make predictions on new data points. Furthermore, visualizations such as scatterplots and histograms can be used to gain insights into the data and inform the predictive model.

The Jupiter Notebook provides an interactive environment where users can quickly develop, analyze and deploy predictive models. With the Jupiter Notebook, users can import data, clean it, explore it interactively, visualize it and build predictive models. Once the model has been created, users can evaluate its performance and adjust improve its accuracy. Finally, the model can be developed for use in real-world scenarios. This could involve integrating it with a web or mobile application, or using it to inform the decisions of real estate investors.

V. RESULT AND DISCUSSION

This section presents the results of different machine learning implemented for house price prediction

A. Linear Regression

Table 2 presents the comparison of actual price of the house as given in data set and the house price predicted as a result of Linear Regression. The graphical representation of comparison presented in Table 2 is given in fig.3.

Table 2 Predicted prices using Linear Regression

Record No.	Actual Price (Lakhs)	Predicted Price (Lakhs)
4	22	24
28	17	20
30	21	17
33	27	24
34	24	23
25	11	11
13	21	19
22	12	10

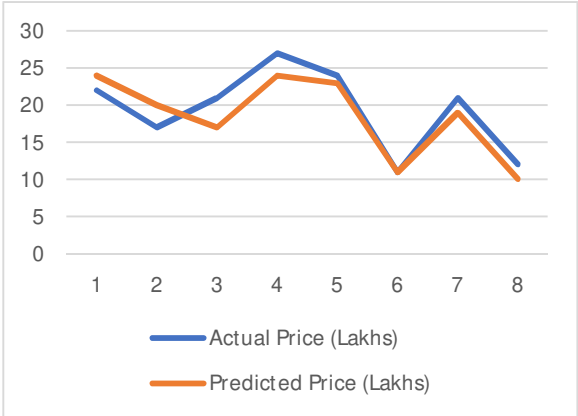


Figure 3.Graph showing comparison of prediction done using Linear Regression

B. Multiple Regression

Table 3 presents the comparison of actual price of the house as given in data set and the house price predicted as a result of Multiple Regression. The graphical representation of comparison presented in Table 3 is given in fig.4.

Table 3Predicted prices using Multiple Regression

Record No.	Actual Price (Lakhs)	Predicted Price (Lakhs)
6	20	21.04943126
11	21	16.86709527
19	21	23.12898921
20	23	22.97885908
26	18	17.73125674
25	13	12.59635298
30	23	16.3344
36	24	26.0487519
40	11	11.39291899

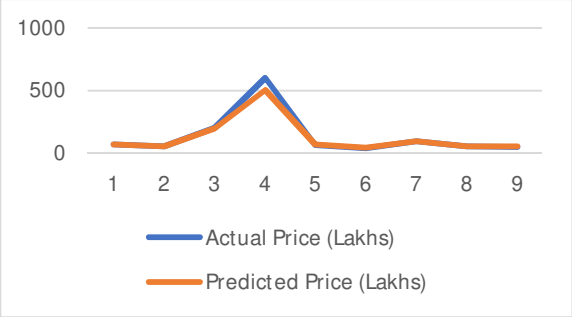


Figure 4.Graph showing comparison of prediction done using Multiple Regression

C. Random Forest Algorithm

While predicted house prices using Random Forest, the observations carried out is presented in Table 4 and their corresponding comparison graph is shown in fig, 5

Table 4 Predicted prices using Multiple Regression

Record No.	Actual Price (Lakhs)	Predicted Price (Lakhs)
45	428	385.46
40	194	155
30	210	207
36	85	109
39	167	158
21	368	328

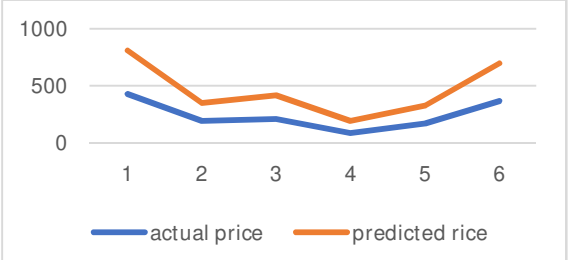


Figure 5.Graph showing comparison of prediction done using Random Forest

Table 5 compares the accuracy obtained in different machine learning algorithms used for house price prediction and the corresponding graph is shown in fig.5

Table 5 Accuracy achieved in different algorithms

S.No.	Type of algorithm	Prediction Accuracy (%)
01	Linear regression	81
02	Multiple regression	77
03	Random forest	75

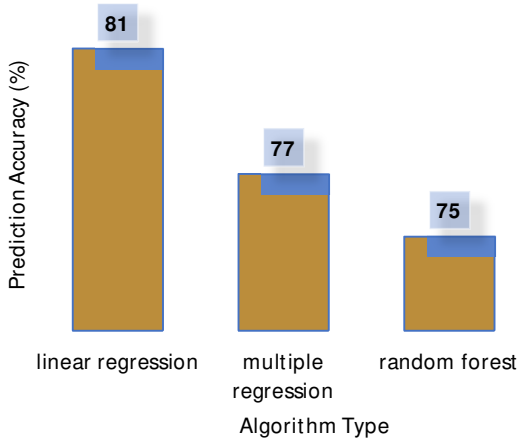


Figure 5 Comparison Graph of prediction accuracy achieved for different machine learning algorithms.

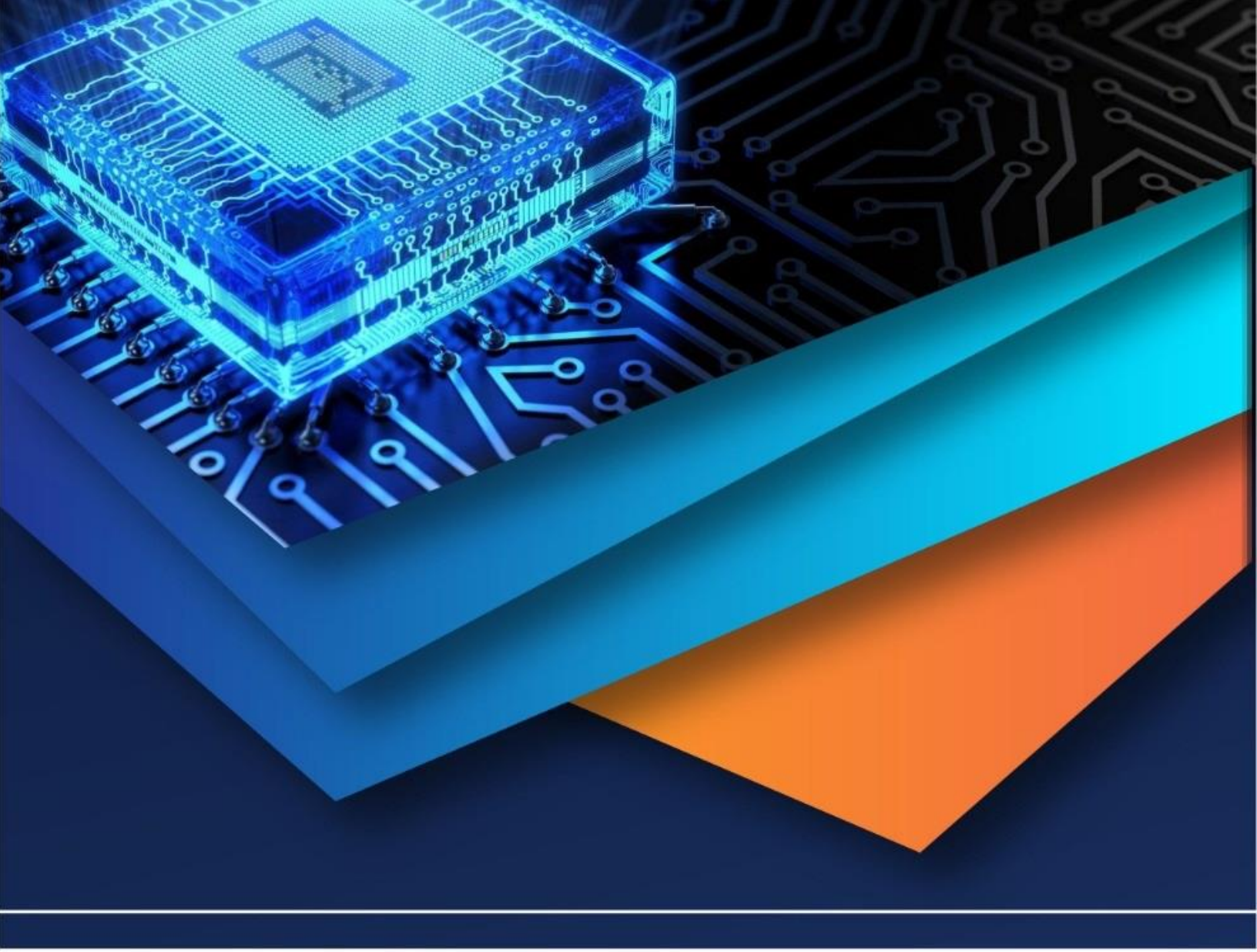
VI. CONCLUSION AND FUTURE SCOPE

The research presented in this paper demonstrates the potential of machine learning algorithms for accurately predicting house prices. With the proper data and features, a well-trained model based on Linear Regression can be used to accurately predict the price of a house. However the accuracy levels can vary based on the datasets used. While the results of this study are promising, there are many opportunities for future research. For instance, exploring different model architectures, such as deep learning and transfer learning, can improve model performance. Additionally, further research could be done to identify the most important features for house price prediction, as well as to explore the impact of different types of data, such as location and neighbourhood characteristics, on model performance. Finally, developing more efficient methods for training and deploying models could enable the use of machine learning algorithms in a wide range of applications.

REFERENCES

[1] House Price Prediction Modeling Using Machine Learning by Dr. M. Thamarai and Dr. S P. Malarvizhi, DOI: 10.5815/ijieeb.2020.02.03.
[2] Housing Price Prediction via Improved Machine Learning Techniques by Quang Truong, Minh Nguyen, Hy Dang, Bo Mei, DOI: 10.5817/ijieeb.2020.02.03
[3] House Price Forecasting Using Machine Learning by Alisha Kuvalekar, Shivani Manchewar, Sidhika Mahadik and Shila Jawale (guide) vol. 2, 15-20,2020.
[4] House Price Prediction Using Multiple Linear Regression by Anirudh Kaushal, Achyut Shankar Vol. 263. No. 4. 2017.
[5] Machine Learning based House Price Prediction using Regression Techniques by Prof Trupti Dange, Amritesh Mishra, Abhishek Jagtap, Shubham Chavhan, Niraj Chavan vol.01, no.02, pp.17-07,2022.
[6] Al-Madani, S., Abu-Tair, E., & Badra, M. (2020). House price prediction model using multi-layer perceptron and linear regression. International Journal of Computational Intelligence Systems, 13(1), 9-20.
[7] Hong, W., & Zhang, Y. (2016). House price prediction with big data: A review. Applied Intelligence, 45(2), 328–337.
[8] Lee, J., Park, J., Lee, Y. J., Yoon, J. Y., & Kim, J. H. (2017). House price prediction system using machine learning techniques. Expert Systems with Applications, 82, 201-211.
[9] Sharma, A., & Sharma, A. (2019). House price prediction using machine learning. International Journal of Computer Applications, 174(7), 11-14.

[10] Popescu, A. (2019). House price prediction using regression analysis. International Journal of Computer Applications, 172(2), 11-14.
[11] Zareapoor, M., & Hejazi, E. (2015). House price prediction using artificial neural network. Iternational Journal of Advanced Computer Science and Applications, 6(5), 186-193.
[12] Wijaya, S., & Kurniawan, A. (2016). House price prediction using support vector regression. International Journal of Computer Theory and Engineering, 8(3), 245-251.
[13] Dragicevic, D., & Stankovic, M. (2015). House price prediction using ensemble learning. International Journal of Artificial Intelligence, 9(4), 339-351.
[14] Li, X., Li, J., Li, G., Li, H., & Yu, J. (2017). House price prediction using artificial neural network and genetic algorithm. International Journal of Computer Science and Network Security, 17(4), 217-223.
[15] S. S. Iyer, A. A. Reddy, and S. Ch, "Price prediction of residential houses using machine learning algorithms," International Journal of Computer Applications, vol. 170, no. 3, pp. 36–40, 2017.
[16] R. Kumar, S. Gautam, and R. K. Sharma, "An empirical study of housing price prediction using machine learning algorithms," International Journal of Computer Science and Information Security, vol. 16, no. 5, pp. 11–18, 2018
[17] B. P. Dhanaraj and P. A. Akila, "A study on predicting house prices using machine learning algorithms," International Journal of Advanced Research in Computer and Communication Engineering, vol. 6, no. 1, pp. 8–14, 2017.
[18] Y. Zhang, "House price prediction using machine learning techniques," in Proceedings of the 14th International Conference on Artificial Intelligence and Machine Learning, pp. 527–532, 2019.
[19] M. T. Hasan and M. R. Islam, "A comparative study on house price prediction using machine learning techniques," in Proceedings of the 11th International Conference on Machine Learning and Applications, pp. 807–812, 2017.
[20] M. A. Hossain, S. K. Choudhury, and M. T. M. Islam, "A comparative analysis of machine learning techniques for house price prediction," International Journal of Computer Science and Information Security, vol. 15, no. 10, pp. 128–134, 2017.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY