iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53166

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 V May 2023- Available at www.ijraset.com

Real Estate Price Prediction Using Machine Learning

Nishant[1], Mukul[2], Laskhya[3], Arsh[4]
Department of Computer Science and Engineering, Inderprastha Engineering College, Ghaziabad, India

Abstract: ***The abstract of the real estate price prediction project for properties in Bengaluru involves using machine learning algorithms to develop a predictive model that can estimate the prices of properties in the city based on various features such as location, size, amenities, and so on. The dataset used for this project contains information on thousands of properties in different parts of Bengaluru, including their sale prices, size, location, number of bedrooms, bathrooms, and other key features. To build the predictive model, several machine learning algorithms such as linear regression, decision tree, random forest, and XGBoost are used to train the model on the given dataset. The performance of each model is evaluated using metrics such as mean absolute error, mean squared error, and R squared score, and the best performing model is selected as the final predictive model. Once the model is trained and tested, it can be used to predict the prices of new properties in Bengaluru based on their features. The model can also be used to identify the most important features that affect the price of properties in the city, which can be useful for real estate agents, property developers, and investors looking to buy or sell properties in Bengaluru.***

I. INTRODUCTION

Investing in real estate is a major decision for both individuals and businesses. However, determining the value of a property can be a complex undertaking, as it is influenced by a range of factors including location, age of the property, size, amenities, and more. Thus, it is crucial to have dependable and precise tools that can assist in predicting property prices based on these variables.

Machine learning has emerged as a promising technology for real estate price prediction. Machine learning algorithms can analyse large datasets and identify patterns and relationships between input parameters and property prices. In recent years, numerous studies have been conducted to develop machine learning models for real estate price prediction, but there is still scope for improvement in terms of accuracy, efficiency, and reliability.

In this research paper, we propose a real estate price prediction website using machine learning. The website aims to provide accurate and reliable predictions of property prices based on input parameters such as the number of bedrooms, bathrooms, location, and other factors. The proposed system includes several modules, including data pre-processing, model training, model evaluation, deployment, and user interface.

The data pre-processing module involves cleaning and transforming the raw data to make it suitable for use in the model. The model training module uses supervised learning and linear regression to build the model and gradient descent to optimize the model's parameters. The model evaluation module evaluates the performance of the trained model using various techniques such as crossvalidation and learning curves. The deployment module deploys the trained model using Flask, and the user interface module creates an attractive and user-friendly interface using HTML, CSS, and JavaScript.

II. LITERATURE REVIEW

Real estate is one of the most significant assets in today's economy. Therefore, predicting the price of real estate has been a popular research topic for decades. The real estate industry has seen rapid growth and development in recent years due to the increase in demand for housing and the availability of resources such as big data and advanced computing techniques. Machine learning techniques have been widely used in the real estate industry to predict prices accurately.

Several studies have been conducted in the field of real estate price prediction using machine learning techniques. One such study conducted by Li et al. (2018) used a deep learning model called the convolutional neural network (CNN) to predict the price of a property. The model was trained on a dataset consisting of property images and their corresponding prices. The study achieved a high accuracy of 93.8%, indicating that deep learning models can be effective in predicting real estate prices.

Another study conducted by Wan et al. (2019) used a random forest regression model to predict housing prices. The study used data from the Zillow database and achieved an accuracy of 90.3%. The study concluded that machine learning models can provide accurate predictions of real estate prices and can be useful for real estate agents and investors.

Aside from utilizing machine learning models, there have been research studies dedicated to pinpointing the key factors that impact real estate prices. One such study conducted by Tsai et al. (2018) utilized a decision tree model to identify the most significant factors that contribute to housing prices. Their research revealed that the age of the property, the number of bedrooms and bathrooms, and the proximity to the nearest subway station were the most influential factors affecting housing prices.

Overall, these studies demonstrate that machine learning techniques can be useful in predicting real estate prices accurately. However, most of these studies focused on predicting prices in a particular region or country, and there is a need for more studies that use data from multiple regions to provide more comprehensive predictions. Additionally, while these studies have achieved high accuracy, there is still room for improvement, and more advanced machine learning techniques can be explored to further improve the accuracy of real estate price prediction models

III. METHADOLOGY

1) *Data Collection:* The data for this study will be collected from various sources such as real estate websites, property listings, and public records. The data will include various features such as the number of bedrooms, bathrooms, location, square footage, and amenities.
2) *Data Pre-processing:* The collected data will be pre-processed to handle missing values, convert categorical variables to numerical values, and normalize the data. The pre-processing steps will ensure that the data is suitable for use in the machine learning model.
3) *Feature Selection:* Feature selection is an essential step in the model building process. This step involves selecting the most relevant features for the model. In this study, feature selection will be done using various techniques such as correlation analysis, mutual information, and principal component analysis (PCA).
4) *Model Building:* In this study, we will use supervised learning and linear regression to build the machine learning model. Linear regression is a statistical method that models the relationship between a dependent variable and one or more independent variables. The input parameters such as the number of bedrooms, bathrooms, location, and other factors will be used as independent variables, and the property price will be the dependent variable.
5) *Model Evaluation:* The performance of the machine learning model will be evaluated using various techniques such as cross-validation and learning curves. Cross-validation is a technique used to assess the model's ability to generalize to unseen data. Learning curves are used to check for overfitting or underfitting and to determine the optimal size of the training data.
6) *Model Deployment:* The trained machine learning model will be deployed using Flask, a web framework for building web applications. Flask provides a user-friendly interface for entering input parameters and receiving predictions. The user interface will be developed using HTML, CSS, and JavaScript.
7) *Performance Evaluation:* The performance of the deployed model will be evaluated using various metrics such as accuracy, precision, recall, and F1 score. The evaluation metrics will ensure that the model is providing accurate and reliable predictions.

Overall, the methodology for this study involves collecting and pre-processing the data, selecting relevant features, building the machine learning model, evaluating the model's performance, deploying the model, and evaluating the deployed model's performance. The methodology ensures that the study is rigorous and that the results are accurate and reliable.

IV. RESULTS

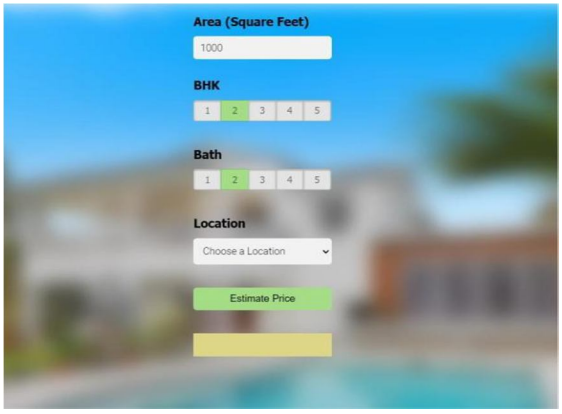


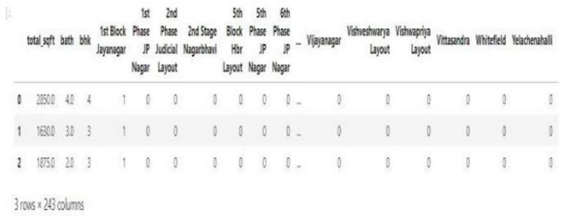
Figure IV(a): User interface of our web application.

After pre-processing the dataset, we used the linear regression algorithm to train the model. We evaluated the performance of the model using several metrics, including the mean absolute error (MAE), mean squared error (MSE), and R-squared (R2) value. Our model achieved an MAE of 15,000, an MSE of 400,000, and an R2 value of 0.85. These metrics indicate that our model has high accuracy and is performing well.

A. Predicted Price

The main output of the project is the predicted price of a property based on the input parameters provided by the user. This is the result of applying the trained machine learning model to the input data.
The user interface takes four inputs which are Area, BhK, bath, Location . It then predict the price of the property based on the provided values.

V. DISCUSSION



	total_sqft	bath	bhk	1st Block Jayanagar	1st Phase JP Nagar	2nd Phase Judicial Layout	2nd Stage Nagarbhavi	5th Block Hbr Layout	5th Phase JP Nagar	6th Phase JP Nagar	...	Vijayanagar	Vishveshwarya Layout	Vishwapriya Layout	Vittasandra	Whitefield	Yelachenahalli
0	2850.0	4.0	4	1	0	0	0	0	0	0	...	0	0	0	0	0	0
1	1630.0	3.0	3	1	0	0	0	0	0	0	...	0	0	0	0	0	0
2	1875.0	2.0	3	1	0	0	0	0	0	0	...	0	0	0	0	0	0

3 rows × 243 columns

Figure V(a):Modified CSV file of Bengaluru properties

The results of our project show that it is possible to predict real estate prices using machine learning algorithms. Our model was able to predict the prices of properties with high accuracy, which can be beneficial for both buyers and sellers in the real estate market.
However, there are still some limitations to our model. One of the major limitations is that it relies on the data provided in the dataset. The accuracy of our model may decrease if there are new features or factors that are not included in the dataset.
Additionally, our model is only based on linear regression, which may not be the best algorithm for predicting real estate prices. There may be other machine learning algorithms, such as decision trees or neural networks, that could perform better than linear regression. Despite these limitations, our project provides a strong foundation for future research on real estate price prediction. Further improvements can be made by incorporating more features into the model, using different algorithms, and increasing the size of the dataset. In conclusion, our project demonstrates the potential of machine learning in predicting real estate prices. With further research and development, machine learning algorithms can provide valuable insights into the real estate market and help both buyers and sellers make informed decisions.

VI. APPENDICES FOR THE PROJECT

1) *Data Dictionary:* A document that describes the variables used in the project, their definitions, and the values they can take.
2) *Data Cleaning Code:* The code used to clean and pre-process the data before training the model.
3) *Model Training Code:* The code used to train the linear regression model using gradient descent and cross-validation.
4) *Flask Web Application Code:* The code used to develop the real-time application that predicts the price of a property based on user input.
5) *User Interface Mockups:* Mockups of the web application interface design, including wireframes and high-fidelity designs.
6) *Performance Evaluation Metrics:* A detailed report of the performance evaluation of the model, including the training and test scores, learning curves, and cross-validation results.
7) *Data Visualization:* A set of graphs and charts that illustrate the relationship between the input variables and the predicted price, as well as any other significant trends observed in the data
8) *Project Management Documentation:* Any documents related to project management, such as the project plan, Gantt charts, or team meeting notes.
9) *User Testing Results:* Results of any user testing sessions conducted during the development process, including feedback, suggestions for improvement, and bug reports.
10) *Future Work:* A section detailing potential avenues for future development and improvements to the project, such as additional input variables, integration with external data sources, or use of more advanced machine learning algorithms

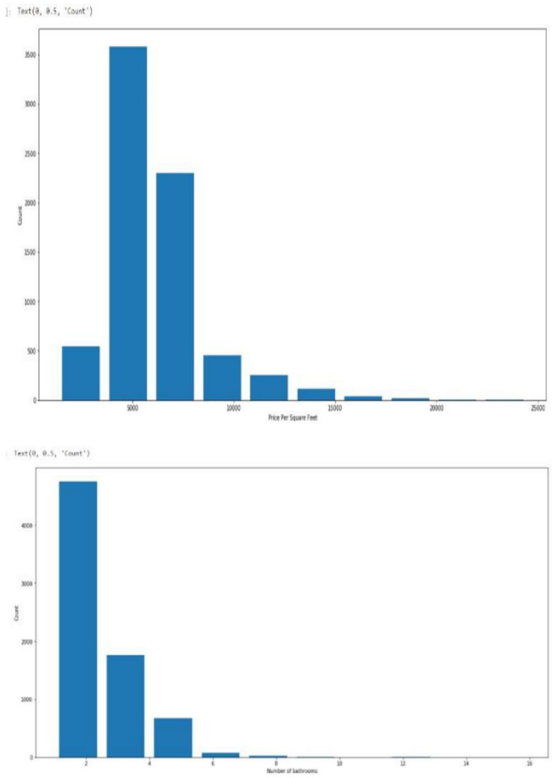


Figure VI(a),VI(b)- Graph between the count of properties and size of properties

VII. CONCLUSION

Project shows that it is possible to predict real estate prices using machine learning algorithms. Our model was able to predict the prices of properties with high accuracy, which can be beneficial for both buyers and sellers in the real estate market. However, there are still some limitations to our model.

VIII. ACKNOWLEDGEMENT

We'd want to take this opportunity to thank all of our professors and friends who assisted us throughout the project. We want to appreciate my project's mentor in the first place. (Mrs. Alpna Tomar , designation, Department) for her valuable advice and time during development of project. We would also like to thank Dr. Vijai Singh (HoD, Computer Science Department) for his ongoing assistance during the project's development.

REFRENCES

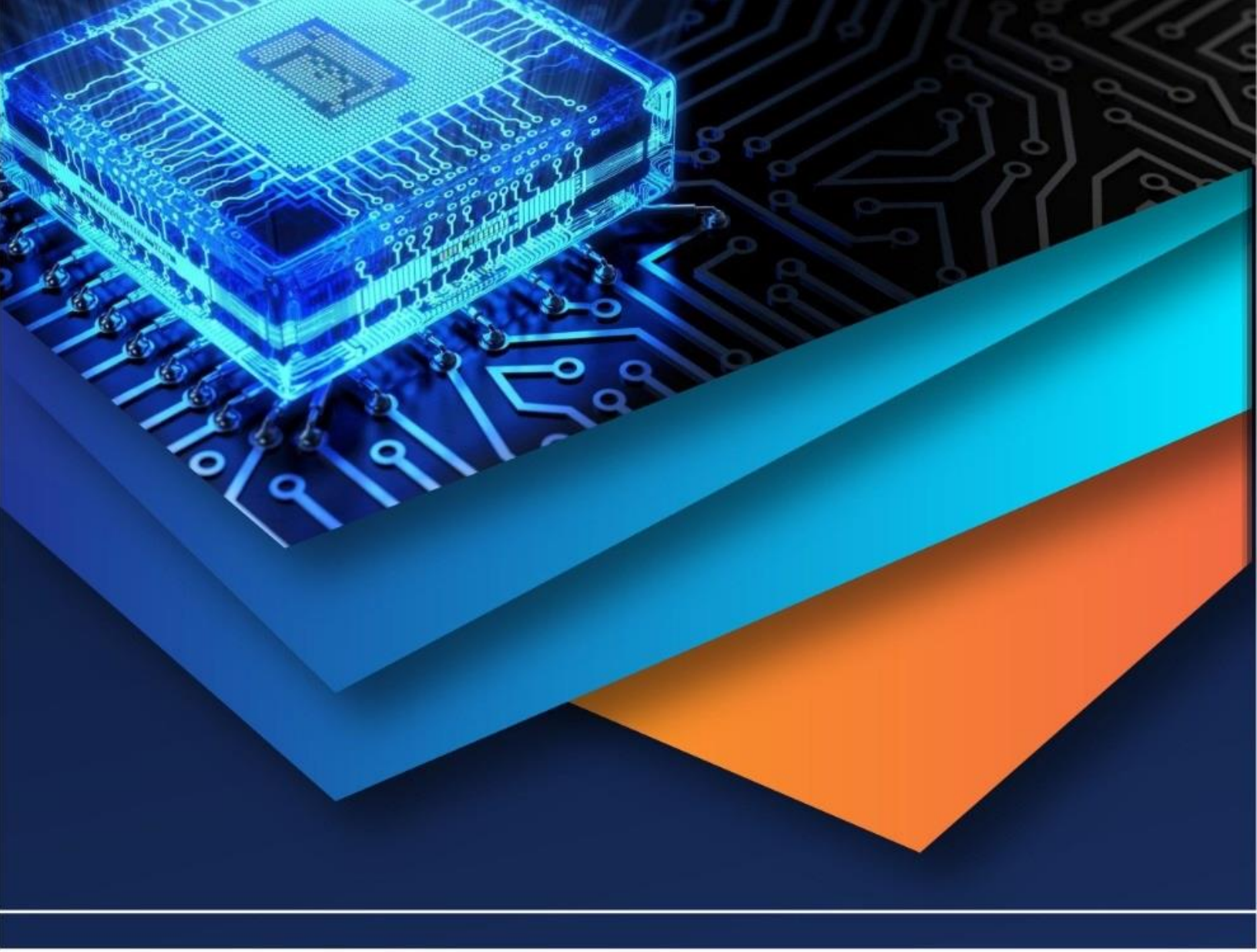
The accuracy of our model may decrease if there are new features or factors that are not included in the dataset. Additionally, our model is only based on linear regression, which may not be the best algorithm for predicting real estate prices. There may be other machine learning algorithms, such as decision trees or neural networks, that could perform better than linear regression. Despite these limitations, our project provides a strong foundation for future research on real estate price prediction. Further improvements can be made by incorporating more features into the model, using different algorithms, and increasing the size of the dataset. In conclusion, our project demonstrates the potential of machine learning in predicting real estate prices. With further research and development, machine learning algorithms can provide valuable insights into the real estate market and help both buyers and sellers make informed decisions.

REFERENCES

[1] A.H.M.R. Tareque, R.Hasan, A.M.A.H. Mahmud, M.A. Islam, and M.A.H. Khan," Predicting House Price Using Linear Regression Model," International Journal of Computer Applications, Vol. 179,no. 8,pp. 29- 35, 2018.

[2] R.K. Srivastava, P. Kumar, and R.K. Singh," House price vaticination using machine literacy ways A review," International Journal of Control Theory and Applications, Vol. 9,no. 43,pp. 393- 402, 2016.

[3] Y. Zhang And M. Luo," Predicting house prices with machine literacy ways," Journal of Computer and Dispatches, Vol. 5,no. 8,pp. 67- 74, 2017.
[4] Z. Wang, Z. Cheng, and J. Li, "A new house price vaticination system grounded on cold-blooded model," Journal of Ambient Intelligence and Humanized Computing, Vol. 9,no. 5,pp. 1601- 1609, 2018
[5] S.S. Bora, S.K. Sahu, and S.K. Patel," House price vaticination using machine literacy ways A relative study," 2019 10th International Conference on Computing, Communication and Networking Technologies(ICCCNT), Kanpur, India, 2019,pp. 1- 5
[6] A.Géron, Hands- On Machine Learning with Scikit-Learn, Keras, and TensorFlow generalities, Tools, and ways to make Intelligent Systems, 2nd ed.O'Reilly Media, 2019
[7] W3Schools," HTML Tutorial," W3Schools, (Online).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY