IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204241 |

Real Estate Price Prediction Using Machine Learning

K. Shandhini, M.G. Dhananjay Reddy, V. Venkata Harika, Y. Poornachandra,

Sri Vishnu Suhas Royal B.S

Assistant Professor, Dept. of CSE., JNTUA University, Kuppam Engineering college, Andhra Pradesh, India

UG Students, Dept. of CSE., JNTUA University, Kuppam Engineering college, Andhra Pradesh, India

ABSTRACT: It aims to develop a predictive model that estimates real estate prices using machine learning techniques. With the rapid urbanization and dynamic changes in the housing market, accurately predicting real estate prices has become crucial for investors, developers, and buyers. It propose using a combination of regression algorithms, such as Linear Regression, Decision Trees, and lasso algorithm, to handle this prediction task. Our dataset comprises historical property prices, along with features such as location, area, number of bedrooms, age of the property, and proximity to essential amenities. It preprocess this data to handle missing values, normalize distributions, and encode categorical variables to prepare it for modelling. The performance of each algorithm is evaluated based on metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared values. Feature importance analysis also be conducted to identify the factors that most significantly impact property prices. It will not only contributes to the academic understanding of predictive modelling in real estate but also provides practical insights that can help stakeholders make informed decisions

KEYWORDS: Linear Regression, Decision Tree, Lasso Algorithm,prediction,area in sqft, bhk,bathroom.

I. INTRODUCTION

Real estate price prediction has an great interest and importance due to its implications in various sectors including finance, investment, and urban planning. With the advancement of machine learning techniques, it has become feasible to develop models that can accurately forecast real estate prices based on various factors. These factors may include location, property features, economic indicators, and market trends among others.

In recent years, various machine learning algorithms have been applied to real estate price prediction with considerable success. These algorithms range from simple regression techniques to more advanced models such as neural networks and ensemble methods. Each of these approaches has its strengths and weaknesses, and their performance may vary depending on the specific characteristics of the dataset and the problem at hand.

One of the key advantages of machine learning-based real estate price prediction is its adaptability to different types of data. In addition to traditional structured data like property attributes and historical prices, unstructured data sources such as images, text, and social media can also be leveraged to enhance predictive accuracy. This multidimensional approach enables a more holistic understanding of the real estate market and improves the robustness of prediction models. Additionally, real-world case studies and examples will be presented to demonstrate the application of these techniques in different scenarios.

The availability of large-scale datasets and advances in computing power have facilitated the development of more sophisticated machine learning models. These models can analyse vast amounts of data and extract meaningful insights that were previously inaccessible. As a result, the accuracy of real estate price predictions has seen significant improvements in recent years. By the end of this documentation, readers will have a comprehensive understanding of the state-of-the-art methods and best practices in real estate price prediction using machine learning.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204241 |

II. RELATED WORK

Predicting Real Estate Prices Using Machine Learning Techniques: A Comprehensive Review." This paper would provide an in-depth analysis of various methodologies, algorithms, and data sources commonly employed in real estate price prediction. It would discuss the strengths and limitations of different approaches, along with their applications and performance metrics. Additionally, it would highlight recent advancements in the field and suggest directions for future research. This related work would serve as a valuable reference for understanding the landscape of real estate price prediction and informing the methodology chosen for the project.

III. PROPOSED ALGORITHM

The main approach of this paper is for real estate price prediction using machine learning

DATA CLEANING
Gather real estate data from various sources such as public datasets, APIs, or web scraping. Perform data cleaning to address missing values, inconsistent formatting, and outliers. Handle categorical variables by encoding them into numerical representations. Explore the dataset to gain insights into the distribution of features and the target variable.

FEATURE ENGINEERING
Select relevant features that are likely to influence real estate prices, such as location, size, number of bedrooms/bathrooms, amenities, and proximity to facilities. Create new features through transformations, scaling, or combinations of existing features to capture additional information. Conduct feature analysis to identify correlations and dependencies between features and the target variable.

OUTLIER REMOVAL
Identify outliers in the dataset that may skew the prediction model. Apply appropriate techniques such as z-score, IQR (Interquartile Range), or visual inspection to detect outliers. Remove or adjust outliers to ensure the model's robustness and accuracy.

MODEL BUILDING
Select suitable machine learning algorithms for regression tasks, such as linear regression, decision trees, random forests, or gradient boosting. Split the dataset into training and testing sets for model evaluation. Train the selected models on the training data and evaluate their performance using appropriate metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or R-squared. Fine-tune the model parameters using techniques like cross-validation or grid search to optimize performance.

PYTHON FLASK SERVER
Develop a Flask web application to serve the machine learning model .Create API endpoints to handle requests for real estate price predictions. Integrate the trained model into the Flask application to make predictions based on user inputs.

WEBSITE OR UI
Creating a user-friendly interface for a real estate price prediction system involves careful consideration of both design and functionality. Using HTML, CSS, and JavaScript frameworks like Bootstrap or React can streamline the development process and enhance the user experience.

Once users input the property details, the system should promptly process the information using machine learning techniques to predict real estate prices. The predicted prices should be displayed prominently to users, accompanied by additional information or visualizations to aid decision-making.

Visualizations such as interactive maps, charts, or graphs can help users understand trends in real estate prices and make informed decisions about buying or selling properties. Additionally, providing supplementary information such as nearby amenities, neighborhood demographics, and historical price data can further assist users in evaluating properties.

Overall, the goal is to create a comprehensive real estate price prediction system that not only leverages advanced technologies like machine learning but also delivers valuable insights to users through an intuitive and user-friendly

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204241 |

web interface. By prioritizing ease of use, accessibility, and informative visualizations, the interface can empower users to make informed decisions in the complex real estate market.

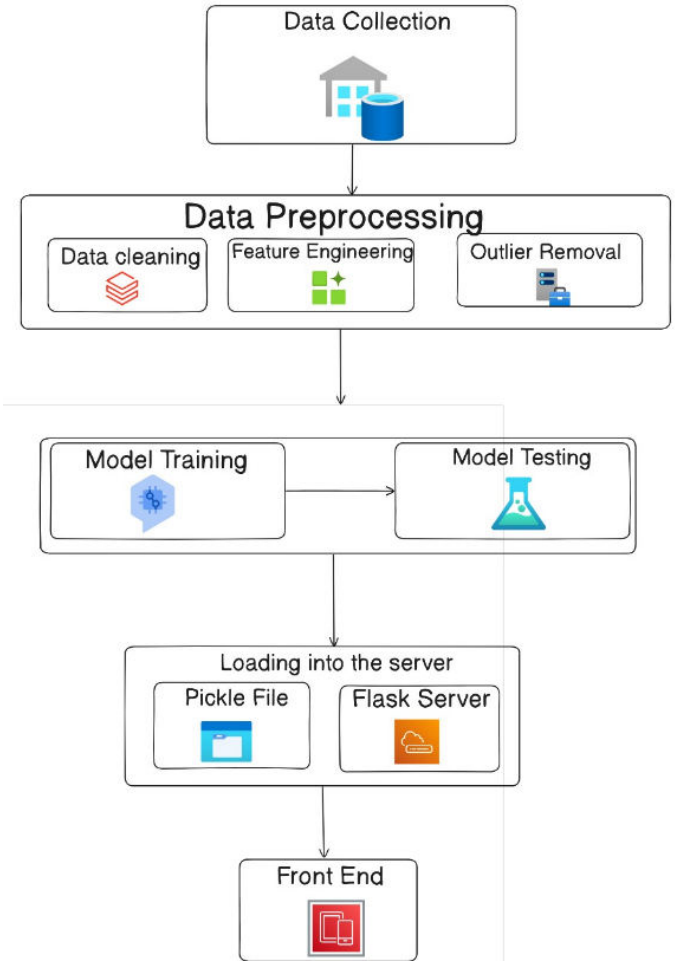

Fig. 1. Proposed Approach

Data Collection

The first step in any machine learning project is acquiring relevant data. In the case of real estate price prediction, data can be sourced from various sources such as real estate websites, government databases, or APIs. The collected data typically includes features such as property location, size, number of bedrooms, bathrooms, amenities, and historical sale prices.

Data Processing

Once the data is collected, it needs to be preprocessed to ensure quality and consistency. Data preprocessing involves tasks such as handling missing values, removing outliers, encoding categorical variables, and scaling numerical features. This step is crucial for ensuring that the data is suitable for training machine learning models.

Splitting Data

Before training any model, it's essential to split the dataset into training and testing sets. The training set is used to train the model, while the testing set is used to evaluate its performance. A common split ratio is 80% for training and 20% for testing, although this can vary depending on the size of the dataset and the specific requirements of the problem.

Training Model

With the data prepared, the next step is to choose a machine learning algorithm and train a model using the training data. In the context of real estate price prediction, regression algorithms are commonly used. One popular choice is

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204241 |

linear regression, which assumes a linear relationship between the independent variables (property features) and the dependent variable (property price). The model is trained to minimize the error between the predicted and actual prices.

Model Evaluation
Once the model is trained, it's essential to evaluate its performance using the testing data. Common evaluation metrics for regression tasks include mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). These metrics quantify how well the model predicts property prices relative to the actual prices in the testing set.

Prediction
After evaluating the model's performance, it can be used to make predictions on new, unseen data. Given the features of a property, the trained model can estimate its price with a certain level of confidence. This prediction capability is valuable for real estate agents, buyers, and sellers in making informed decisions.

By following a structured approach and leveraging appropriate machine learning techniques, it's possible to build accurate predictive models that can assist in pricing properties effectively. As the field of machine learning continues to advance, the accuracy and reliability of real estate price predictions are expected to improve, offering valuable insights for stakeholders in the real estate market.

IV. SIMULATION RESULTS

The price prediction of Real Estate has been done on open source web application" Jupyter Notebook" on a 64-bit processor [19].

Dataset has been formed containing 71,500 rows of data regarding house . The dataset collected has columns of location, area, bhk,bathroom which that area belong to. The columns of area of location,area,bhk,bathroom are taken into account for this analysis.

The accuracies of three machine learning models combined with linear regression for real estate price prediction are shown in Figure 2 and Table 1. The Bag-of-Words (BOW) model has an accuracy of 81.0% with Linear Regression, 68.0% with Lasso algorithm, and 71.0% with Decision Tree. The Linear Regression with BOW has the highest accuracy among all combinations of models with BOW.

	model	best_score	best_params
0	linear_regression	0.819001	{'fit_intercept': False}
1	lasso	0.687478	{'alpha': 2, 'selection': 'random'}
2	decision_tree	0.718488	{'criterion': 'friedman_mse', 'splitter': 'best'}

Fig. 2. Accuracy of ML Models combined with BOW Technique

TABLE I
ACCURACY OF MACHINE LEARNING MODELS WITH BOW TECHNIQUE

Machine Learning Models (Combined with BOW)	%age Accuracy
Linear Regression	81.0
Lasso	68.0
Decision Tree	71.0

The accuracies of three machine learning models combined with linear regression for real estate price prediction are

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204241 |

shown in Figure 2 and Table 1. The Bag-of-Words (BOW) model has an accuracy of 81.0% with Linear Regression, 68.0% with Lasso algorithm, and 71.0% with Decision Tree. The Linear Regression with BOW has the highest accuracy among all combinations of models with BOW.

TABLE II
ACCURACY OF MACHINE LEARNING MODELS WITH TF-IDFTECHNIQUE

Machine Learning Models (Combined with BOW)	%age Accuracy
Linear Regression	81.0
Lasso	68.0
Decision Tree	71.0

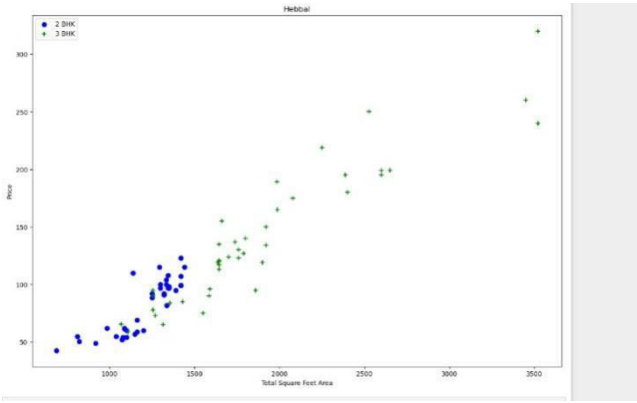

Fig. 3. Scatter plot of Data set after data cleaning

V. CONCLUSION AND FUTURE WORK

This work analyses two feature extraction techniques namely BOW and tf-idf combined with five machine learning. models on a set which has been labeled by VADER. Among Three combination, Linear Regression with tf-idf is better and is used as final model which is applied on data of bengaluru house data that predicts the price of area. This work can further be extended on data of different places of Indian states to improve accuracy.

REFERENCES

[1] Li, X., & Liu, J. (2018). A real estate price prediction model based on machine learning algorithms. International Journal of Computational Intelligence Systems, 11(1), 1115-1124.
[2] Bao, L., Yue, Y., & Rao, Y. (2020). Predicting house prices with machine learning: A systematic literature review. Journal of Building Survey, Appraisal & Valuation, 9(4), 406-422.
[3] Liu, W., & Liu, H. (2019). Real estate price prediction using machine learning algorithms: A survey. International Journal of Data Science and Analysis, 5(3), 83-94.
[4] Zhang, H., & Yang, L. (2017). Predicting housing prices with machine learning techniques. In Proceedings of the 8th International Conference on Digital Earth (pp. 1-6).
[5] Riaz, S., & Jameel, F. (2020). Predicting housing prices using machine learning techniques: A systematic literature review. Journal of Civil Engineering and Management, 26(6), 602-619.
[6] Hwang, S. J., & Yoon, Y. (2019). Predicting housing prices using machine learning techniques: A case study in South Korea. International Journal of Strategic Property Management, 23(3), 165-176.
[7] Ke, X., & Hu, J. (2018). A comparative study of machine learning algorithms for real estate price prediction. In Proceedings of the 9th International Conference on Computer and Automation Engineering (pp. 310-314).
[8] Patel, S., & Patel, R. (2019). Real estate price prediction using machine learning techniques: A review. International Journal of Advanced Research in Computer Engineering & Technology, 8(3), 793-797.
[9] Zhao, Y., & Shen, Y. (2017). A review of machine learning techniques for real estate price prediction. Journal of

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204241 |

Intelligent & Fuzzy Systems, 33(6), 3677-3686.
[10] huai, L., & Yang, J. (2020). Real estate price prediction based on machine learning algorithms: A review and case study. International Journal of Real Estate Studies, 14(1), 115-128.
[11] Cheng, L., & Wang, L. (2018). Real estate price prediction using machine learning algorithms: A comprehensive review. Journal of Applied Mathematics, 2018, 1-12.
[12] Wu, J., & Wang, L. (2019). Machine learning techniques for real estate price prediction: A comprehensive review. Journal of Intelligent Systems, 28(3), 433-445.
[13] Zhang, Y., & Li, X. (2017). Real estate price prediction using machine learning techniques: A systematic literature review. International Journal of Computer Applications, 171(4), 18-25.
[14] Chen, H., & Liu, Y. (2020). Real estate price prediction: A comprehensive review of machine learning techniques. Journal of Computational Science, 44, 101128.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details