International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI-Powered Real Estate Price Predictor

Roopa R 1, Murali Karthik C[2]

[1]*ASSISTANT, Assistant Professor, Department of MCA, BIET, Davanagere*

[2]*STUDENT, 4th Semester MCA, Department of MCA, BIET, Davanagere*

Abstract

Accurate property price estimation is a critical challenge in the real estate industry, where traditional methods often rely on subjective assessments and time-consuming manual analyses. This paper presents a web-based application developed using the Flask framework that leverages advanced machine learning algorithms—Random Forest, XGBoost, and LightGBM—for precise real estate price prediction. The proposed system enables users to input key property features such as the number of bedrooms, bathrooms, square footage, and location, and receive instant, data-driven price estimates. The application incorporates distinct functionalities for both Admin and User roles: Admins manage user accounts, FAQs, and user queries, while Users can register, log in, submit property details for prediction, and access support resources. By integrating robust machine learning models with an intuitive user interface and efficient role-based management, this solution enhances transparency, accuracy, and efficiency in property valuation. The results demonstrate the practical benefits of applying machine learning to real estate, streamlining the decision-making process for buyers, sellers, and professionals, and contributing to a more reliable and data-driven marketplace.

Keywords: *Random Forest, XGBoost, LightGBM, Real Estate, Machine Learning*

I. INTRODUCTION

The real estate industry plays a pivotal role in the global economy, influencing investment decisions, urban development, and individual wealth accumulation. One of the most critical aspects of this sector is the accurate estimation of property prices, which directly impacts buyers, sellers, investors, and policymakers. Traditionally, property valuation has relied on manual methods such as comparative market analysis, expert appraisals, and examination of market trends. While these approaches have been widely used, they often suffer from subjectivity, inconsistencies, and are time-consuming. The increasing complexity of real estate markets, coupled with the vast amount of data generated daily, necessitates the adoption of more objective, efficient, and data-driven methods for property price prediction.

The limitations of conventional pricing methods are evident in their dependence on human expertise and the challenge of processing multifaceted factors that influence property values. Elements such as location, size, amenities, economic conditions, and neighborhood characteristics all contribute to price determination. However, manually analyzing these variables and their interactions can be cumbersome and prone to error. Moreover, as real estate markets become more dynamic, the need for timely and accurate price estimations grows, making traditional approaches less effective in meeting the demands of modern stakeholders. This situation calls for innovative solutions that can handle large datasets, uncover hidden patterns, and provide reliable predictions in real time.

Machine learning, a branch of artificial intelligence, offers a promising avenue to address these challenges. By leveraging historical housing data, machine learning algorithms can learn complex relationships between property features and their market prices. Algorithms such as Random Forest, XGBoost, and LightGBM have demonstrated strong performance in regression tasks, making them suitable candidates for real estate price prediction. These models can analyze vast amounts of data, identify subtle trends, and adapt to evolving market conditions, thereby providing accurate and unbiased price estimates. Unlike traditional methods, machine learning models can be continuously retrained with new data, ensuring that predictions remain relevant and reflective of current market realities.

The integration of machine learning into web-based platforms further enhances the accessibility and practicality of these predictive models. By developing a web application using frameworks like Flask, users can interact with sophisticated machine learning algorithms through a simple and intuitive interface. This approach democratizes access to accurate property valuations, enabling buyers, sellers, real estate agents, and investors to make informed decisions without requiring deep technical knowledge. The web application can facilitate real-time price predictions based on user-input property features such as the number of bedrooms, bathrooms, square footage, and location. Additionally, it can incorporate functionalities such as user registration, authentication, and role-based access control to ensure secure and personalized user experiences.

This paper presents the design and implementation of a Flask-based web application that predicts housing prices using three powerful machine learning algorithms: Random Forest, XGBoost, and LightGBM. The system is designed with two distinct user roles—Admin and User—to provide comprehensive platform management and support. The Admin has the ability to manage registered users, oversee frequently asked questions (FAQs), and respond to user queries, thereby ensuring smooth operation and effective communication. Users can register and log in to the platform, input property details to receive instant price predictions, submit queries to the Admin, and access the FAQ section for quick assistance. This role-based structure enhances usability and fosters a collaborative environment between users and administrators.

The primary goal of this project is to deliver an accurate, efficient, and user-friendly solution for real estate price estimation. By harnessing the predictive power of advanced machine learning models and combining it with a seamless web interface, the application aims to reduce the risks associated with overpricing or underpricing properties. It also seeks to improve transparency in the real estate market, enabling stakeholders to make data-backed decisions that save time and resources. The system's ability to provide real-time insights and facilitate direct communication between users and administrators adds significant value to the overall user experience.

In summary, this project addresses the pressing need for modern, data-driven tools in the real estate industry by developing a machine learning-powered web application for property price prediction. It overcomes the shortcomings of traditional valuation methods by offering objective, scalable, and accurate price estimates accessible through an intuitive platform. The application not only streamlines the property buying and selling process but also exemplifies the practical application of machine learning in everyday business contexts, contributing to a more transparent and efficient real estate market.

II.RELATED WORK

Real Estate Price Prediction on Generative Language Models, Authors: Yun Zhao, Girja Chetty, Dat Tran.

This paper explores the application of transformer-based language models for predicting real estate prices, addressing the challenge of incorporating temporal and contextual factors often overlooked in traditional machine learning models. By leveraging the self-attention mechanisms of large language models and fine-tuning them with additional neural layers, the study demonstrates improved predictive performance over conventional methods. The research also highlights the potential of integrating textual insights from external sources, like expert blogs, to enrich the prediction process. Experimental results confirm that transformer models, particularly with in-context unsupervised learning, can

significantly enhance the accuracy and reliability of real estate price estimation.**[1]**

Property Price and Rent Prediction System, Authors: Atharava Pratap Singh, Abhinav Rai, Devansh Karnwal, Anupam Vats, Tripti Pandey.

This paper introduces a machine learning-based system for predicting both property prices and rental values using a combination of Support Vector Machines (SVM), Random Forest, and Gradient Boosting Machine (GBM) algorithms. The study highlights Random Forest as the most effective among the tested models, offering superior predictive accuracy. By incorporating various property-related features, the system enables users to make informed, data-driven decisions without heavy dependence on intermediaries. The research underscores the potential of machine learning to handle the complexities of real estate data and deliver actionable insights for buyers and investors in a rapidly evolving market.**[2]**

Prediction of House Price Using Machine Learning Algorithms, Authors: G. Kiran Kumar, D. Malathi Rani, Neeraja Koppula, Syed Ashraf.

This study investigates the use of various machine learning regression techniques to predict house prices based on user requirements and housing attributes such as area, square footage, and number of floors. Using a Kaggle dataset, the authors implemented algorithms like Linear Regression, KNN, Random Forest, Decision Tree, Extra Trees, and ultimately found CatBoost Regression to provide the most accurate predictions. The paper highlights how advanced machine learning models can deliver precise price estimates by analyzing correlated features, aiding prospective buyers in making financially informed housing decisions.**[3]**

Real Estate Price Prediction using Data Mining Techniques, Authors: Sayar Kumar Dey, Siddhaling Urolagin

This paper presents a comprehensive approach to predicting real estate prices using data mining techniques, focusing on multi-dimensional datasets processed through data cleansing, normalization, and model training. Various regression models such as Elastic Net, Kernel Ridge, Lasso, Random Forest, SVM, XGBoost, LGBM, and Gradient Boosting were evaluated, with Elastic Net achieving the highest accuracy of 94%. The study emphasizes the importance of preprocessing and model selection in building accurate predictive systems, aiming to empower users with data-driven decision-making while reducing reliance on intermediaries in the real estate market.**[4]**

Data-Driven House Price Prediction, Authors: N. Shalini, A. Sai Kiran, B. Sathwik, G. Trishul Reddy, Kilaru Aswini, M. Geetha Yadav.

This study introduces a robust house price prediction system utilizing the Random Forest algorithm alongside multiple and linear regression models for comparison. Drawing on historical real estate data, the model incorporates essential property attributes—like location, number of rooms, and proximity to key amenities—to forecast market values. A thorough data preprocessing and feature engineering phase ensures data quality and model accuracy. The system provides precise local price predictions and comparative insights, making it a valuable decision-support tool for real estate agents, investors, and homeowners.**[5]**

Real Estate Price Prediction Using PySpark MLlib, Authors: Kadam Prajwal Dharmaraj, Pradeep Kumar Gupta, Keerthana Ajith, Mahi Kolli, Sangita Khare, Niharika Panda.

This paper presents a scalable and efficient framework for predicting real estate prices in Indian metropolitan areas using PySpark and its MLlib machine learning library. By processing large-scale housing datasets, the study compares regression models like Random Forest, Linear Regression, Decision Tree, and Gradient-Boosted Tree to identify the most accurate predictors. The research emphasizes the need for big data solutions in the real estate domain, leveraging PySpark's distributed computing to manage data complexity and volume. The proposed system enhances urban housing price forecasts,

offering critical insights for buyers, sellers, and policy-makers.**[6]**

Revolutionizing House Price Prediction: Harnessing the Power of Data Science and Augmented Reality, Authors: Biresh Kumar, Manisha Kumari, Pallab Banerjee, Pooja Jha.

This study explores the integration of Data Science and Augmented Reality (AR) to revolutionize housing price prediction. It proposes a novel system where machine learning algorithms analyze vast real estate datasets—covering property features, neighborhood data, and market trends—and visualize results through an immersive AR interface. Users can interact with real-time predictions, view property comparisons, and personalize their experience using dynamic updates. The innovative use of AR enhances user engagement and decision-making, showing promising potential to transform the home buying and real estate process through data-driven insights and interactive technology.**[7]**

Housing Price Prediction via Improved Machine Learning Techniques, Authors: Quang Truong, Minh Nguyen, Hy Dang, Bo Mei, DOI: 10.1016/j.procs.2020.06.111, Publisher: Elsevier (Published in Procedia Computer Science, Volume 174, 2020, Pages 433–442)

This study investigates both traditional and advanced machine learning techniques for housing price prediction by emphasizing the limitations of commonly used models and exploring the effectiveness of more complex ones. Using methods like hybrid regression and stacked generalization, the authors analyze how different feature sets influence model performance. The research contributes to housing price modeling by offering a comprehensive validation of machine learning methods and demonstrating that careful model selection and integration can significantly improve prediction accuracy.**[8]**

Real Estate Price Prediction Using Machine Learning, Authors: Anil Nahak, Deepika Yadav, Shashikant Gupta.

This paper presents a comprehensive study on predicting real estate prices using a range of machine learning and data mining algorithms. Traditional methods like hedonic regression, artificial neural networks, AdaBoost, and J48 decision trees are employed as baseline models. To enhance prediction accuracy, advanced models such as Random Forest, Gradient Boosted Trees, Multilayer Perceptron, and ensemble learning are implemented using tools like Weka and RapidMiner. The research underscores the significance of combining various machine learning techniques for higher prediction accuracy in real estate pricing.**[9]**

House Price Prediction, Authors: Harshita (Assistant Professor, CSE, Inderprastha Engineering College, Uttar Pradesh, India), Chinmay Kansal (Student, CSE, Inderprastha Engineering College, Uttar Pradesh, India), Ayush Katiyar (Student, CSE, Inderprastha Engineering College, Uttar Pradesh, India), Adarsh Kr. Singh (Student, CSE, Inderprastha Engineering College, Uttar Pradesh, India), ISSN / DOI: Not provided, Publisher: Likely a conference or academic institutional publication (exact source not specified in provided content)

This paper introduces an advanced House Price Prediction System that utilizes machine learning algorithms such as regression analysis, decision trees, and neural networks to estimate residential property values. By combining historical property data, geographic trends via GIS, and economic indicators, the system provides robust and real-time predictions. Data preprocessing ensures clean, reliable input, and a user-friendly interface enables users to input property details and visualize market trends instantly. Designed for buyers, sellers, and real estate agents, the platform modernizes property valuation and supports data-driven decision-making in the real estate industry.**[10]**

III. METHODOLOGY

The methodology for developing the Flask-based real estate price prediction application is a multi-stage process that integrates data collection, preprocessing, machine learning model development, and web application implementation. It begins with gathering and preparing historical housing data, followed by training and evaluating multiple machine learning models—Random Forest, XGBoost, and

LightGBM—to ensure accurate price predictions. The best-performing models are then integrated into a Flask web application, which is designed with distinct Admin and User roles, secure authentication, and a user-friendly interface. Rigorous testing and deployment procedures are followed to ensure the system's reliability, scalability, and ease of use for all stakeholders.

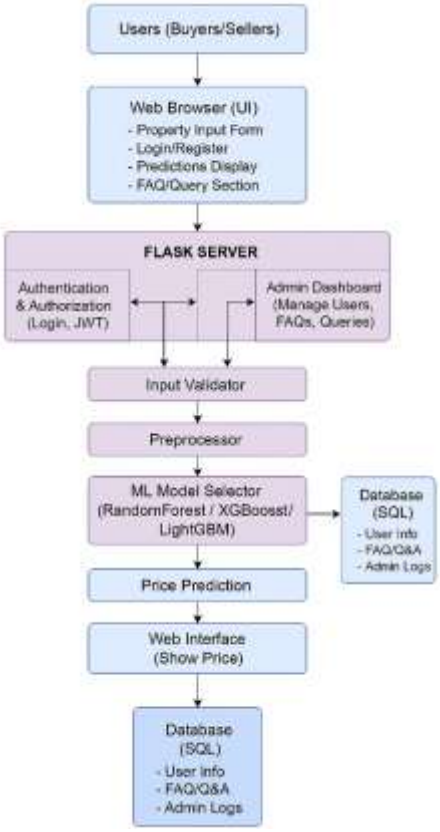

Fig 3.1. Proposed Methodology

3.1 Data Collection and Preprocessing:

The first step involves collecting comprehensive historical data on real estate properties from reliable sources. This dataset includes features such as the number of bedrooms, bathrooms, square footage, location, and other relevant property attributes. Data cleaning is performed to address missing values, remove outliers, and correct inconsistencies. Feature engineering techniques, such as encoding categorical variables and normalizing numerical features, are applied to enhance the predictive power of the models.

3.2 Machine Learning Model Development:

Three advanced regression algorithms—Random Forest, XGBoost, and LightGBM—are selected for their proven accuracy in structured data tasks. The cleaned dataset is divided into training and testing sets. Each model is trained using the training data, with hyperparameters optimized through grid search or cross-validation. The models are evaluated using metrics such as Root Mean Squared Error (RMSE) and R^2 score to ensure robust and reliable predictions. The best-performing models are saved for integration into the web application.

3.3 Web Application Development with Flask

Flask, a lightweight Python web framework, is used to build the application's backend. The trained machine learning models are loaded into the Flask environment to provide real-time predictions based on user input. The application is structured with clear routes for user registration, login, property data input, price prediction, FAQ access, and admin management. Flask's session management and decorators are utilized to implement secure authentication and role-based access control, ensuring that Admin and User functionalities remain distinct and protected.

3.4 User Interface Design

The frontend is developed using HTML, CSS, and Flask's Jinja2 templating engine to create a responsive and intuitive user experience. Users can easily submit property details, view prediction results, browse FAQs, and communicate with the Admin. The Admin dashboard provides tools for managing user accounts, FAQs, and user queries, streamlining platform administration.

3.5 Testing and Validation

Comprehensive testing is conducted at multiple levels, including unit testing, integration testing, and user acceptance testing. These tests ensure that all features function correctly, the machine learning models produce accurate predictions, and the user interface is seamless. Automated testing tools and manual verification are both employed to identify and resolve issues before deployment.

3.5 Deployment and Maintenance

The fully tested application is deployed on a production server or cloud platform, ensuring accessibility and scalability for multiple users. Security measures, such as HTTPS and secure session handling, are implemented to protect user data. Regular updates and maintenance are

planned to incorporate new data, retrain models, and add features based on user feedback, ensuring the application remains accurate and relevant over time.

IV. TECHNOLOGIES USED

The development of the real estate price prediction web application leverages a robust stack of technologies to ensure accurate machine learning predictions, a secure and scalable backend, and an intuitive user interface. Below are the primary technologies and tools utilized in this project:

4.1 Python: The core programming language for both backend development and machine learning model implementation. Python's extensive ecosystem supports rapid development and integration of machine learning workflows.

4.2 Flask: A lightweight and flexible Python web framework used to build the web application's backend. Flask handles routing, session management, and API integration, making it ideal for deploying machine learning models as web services.

4.3 Machine Learning Libraries:

4.4 Scikit-learn: Used for building and training classical machine learning models such as Random Forest, XGBoost, and LightGBM. It provides tools for data preprocessing, model evaluation, and persistence.

4.5 XGBoost and LightGBM: Specialized libraries for gradient boosting algorithms, offering high performance and scalability for structured data tasks.

4.6 TensorFlow and Keras (optional): While primarily used for deep learning, these can be integrated for advanced model experimentation or if neural network approaches are explored in future iterations.

HTML, CSS, and Jinja2: For designing the frontend user interface. HTML and CSS ensure a responsive and user-friendly layout, while Jinja2, Flask's templating engine, dynamically renders content based on user interaction and backend logic.

4.7 Flask Extensions:

Flask-SQLAlchemy: Provides ORM capabilities for database management, supporting user authentication, admin functionalities, and query storage.

Flask-Login, Flask-WTF, Flask-Principal: These extensions facilitate secure user authentication, form handling, and role-based access control, ensuring a seamless and protected user experience.

NumPy and Pandas: Essential Python libraries for data manipulation, preprocessing, and analysis. They are used extensively in preparing data for machine learning models.

Matplotlib: Used for data visualization during model development and evaluation, helping to analyze feature importance and model performance.

Joblib or Pickle: For serializing and loading trained machine learning models into the Flask application, enabling real-time predictions without retraining the models at runtime.

Database: Typically, SQLite or a similar relational database is used for storing user credentials, FAQs, and admin/user interactions, managed through Flask-SQLAlchemy.

This technology stack ensures that the application is not only capable of delivering accurate, real-time property price predictions but is also maintainable, scalable, and secure for both end-users and administrators.

V Result

Model Output

5.1

```
 Random Forest
R² Score: 0.8802
RMSE: 128,919.60

 XGBoost
R² Score: 0.8902
RMSE: 123,428.33
```

5.2

```
 LightGBM
R² Score: 0.8982
RMSE: 118,831.36

Predicted prices for the given house:
Random Forest: $240,730.70
XGBoost: $232,700.75
LightGBM: $243,136.02
```

VI. CONCLUSION

In conclusion, the integration of advanced machine learning algorithms with a user-friendly Flask-based web application offers a powerful and efficient solution for real estate price prediction. By automating and enhancing the accuracy of property valuations, this system addresses the limitations of traditional methods, streamlines the decision-making process for buyers, sellers, and professionals, and fosters greater transparency and trust in the real estate market. The role-based functionalities, secure authentication, and interactive features further ensure that both users and administrators can benefit from a seamless and reliable experience, demonstrating the significant impact of modern data-driven technologies in transforming the real estate industry.

REFERENCES

[1]. Real Estate Price Prediction on Generative Language Models, Authors: Yun Zhao, Girja Chetty, Dat Tran, DOI: 10.1109/CSDE59766.2023.10487658, Publisher: IEEE, Conference: 2023 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Nadi, Fiji, Date Added to IEEE Xplore: 05 April 2024.

[2]. Property Price and Rent Prediction System, Authors: Atharava Pratap Singh, Abhinav Rai, Devansh Karnwal, Anupam Vats, Tripti Pandey, DOI: 10.1109/SCEECS64059.2025.10940122, Publisher: IEEE, Conference: 2025 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Conference Location: Bhopal, India, Date Added to IEEE Xplore: 02 April 2025.

[3]. Prediction of House Price Using Machine Learning Algorithms, Authors: G. Kiran Kumar, D. Malathi Rani, Neeraja Koppula, Syed Ashraf, DOI: 10.1109/ICOEI51242.2021.9452820, Publisher: IEEE, Conference: 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Conference Location: Tirunelveli, India, Date Added to IEEE Xplore: 21 June 2021.

[4].Real Estate Price Prediction using Data Mining Techniques, Authors: Sayar Kumar Dey, Siddhaling Urolagin,DOI:10.1109/GUCON50781.2021.9573829, Publisher: IEEE, Conference: 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), Conference Location: Kuala Lumpur, Malaysia, Date Added to IEEE Xplore: 02 November 2021.

[5]. Data-Driven House Price Prediction, Authors: N. Shalini, A. Sai Kiran, B. Sathwik, G. Trishul Reddy, Kilaru Aswini, M. Geetha Yadav, DOI: 10.1109/ICAAIC60222.2024.10575519, Publisher: IEEE, Conference: 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Conference Location: Salem, India, Date Added to IEEE Xplore: 02 July 2024.

[6]. Real Estate Price Prediction Using PySpark MLlib, Authors: Kadam Prajwal Dharmaraj, Pradeep Kumar Gupta, Keerthana Ajith, Mahi Kolli, Sangita Khare, Niharika Panda, DOI: 10.1109/ICSPCRE62303.2024.10675142, Publisher: IEEE, Conference: 2024 IEEE International Conference on Smart Power Control and Renewable Energy (ICSPCRE), Conference Location: Rourkela, India, Date Added to IEEE Xplore: 25 November 2024.

[7]. Revolutionizing House Price Prediction: Harnessing the Power of Data Science and Augmented Reality, Authors: Biresh Kumar, Manisha Kumari, Pallab Banerjee, Pooja Jha, DOI: 10.1109/SMARTGENCON60755.2023.10442517, Publisher: IEEE, Conference: 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON),Conference Location: Bangalore, India, Date Added to IEEE Xplore: 28 February 2024.

[8].Housing Price Prediction via Improved Machine Learning Techniques, Authors: Quang Truong, Minh Nguyen, Hy Dang, Bo Mei, DOI: 10.1016/j.procs.2020.06.111, Publisher: Elsevier (Published in Procedia Computer Science, Volume 174, 2020, Pages 433–442).

[9].Real Estate Price Prediction Using Machine Learning, Authors: Anil Nahak, Deepika Yadav, Shashikant Gupta, ISSN: 2582-7421, Publisher: International Journal of Research Publication and Reviews (IJRPR), Journal Homepage: www.ijrpr.com.

[10].House Price Prediction, Authors: Harshita (Assistant Professor, CSE, Inderprastha Engineering College, Uttar Pradesh, India), Chinmay Kansal (Student, CSE, Inderprastha Engineering College, Uttar Pradesh, India), Ayush Katiyar (Student, CSE, Inderprastha Engineering College, Uttar Pradesh, India), Adarsh Kr. Singh (Student, CSE, Inderprastha Engineering College, Uttar Pradesh, India), ISSN / DOI: Not provided, Publisher: Likely a conference or academic institutional publication (exact source not specified in provided content).