



Real Estate Price Prediction Using Data Science and Machine Learning

A. Suresh Kumar¹, Nishalin M², Afsheen A³, Haruna W⁴, Ibrahim M⁵

¹ Assistant Professor, Department of CSE, Rathinam Technical Campus, Coimbatore, Tamil Nadu, India.

^{2,3,4,5} Final Year Students, Department of CSE, Rathinam Technical Campus, Coimbatore, Tamil Nadu, India.

To Cite this Article: A. Suresh Kumar¹, Nishalin M², Afsheen A³, Haruna W⁴, Ibrahim M⁵, “Real Estate Price Prediction Using Data Science and Machine Learning”, Indian Journal of Computer Science and Technology, Volume 04, Issue 03 (September-December 2025), PP: 265-268.



Copyright: ©2025 This is an open access journal, and articles are distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by-nc-nd/4.0/); Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract: The real estate sector plays a crucial role in economic development, yet property price estimation remains a complex and challenging task due to the influence of multiple factors such as location, property size, number of bedrooms, and market demand. Traditional valuation methods largely depend on manual assessment, broker expertise, and static market comparisons, which often lead to inconsistent, time-consuming, and inaccurate price estimation. These challenges create uncertainty for buyers, sellers, and real estate professionals in making informed decisions.

This project proposes a Real Estate Price Prediction System using Data Science and Machine Learning to provide accurate and real-time property price estimation. The system utilizes historical real estate data and applies machine learning techniques to analyze key property attributes including total square footage, number of bedrooms (BHK), number of bathrooms, and location. Feature engineering and data pre-processing techniques are employed to improve prediction accuracy and effectively handle location-based price variations.

The proposed system integrates a trained machine learning model with a Flask-based backend and a web-based frontend interface. Users can input property details through a simple and user-friendly interface, and the system instantly predicts the estimated property price. The backend processes user inputs, performs real-time inference using the trained model, and returns reliable prediction results with low latency. The system ensures ease of use, scalability, and efficient deployment using lightweight and open-source technologies.

I. INTRODUCTION

Real estate is an important sector that supports economic development and drives urban growth. In many developing regions, property transactions play a major role in investment decisions and livelihood security. Despite its importance, property price estimation remains a challenging task due to factors such as location, property size, and market demand. Traditional valuation methods rely on manual assessment, broker expertise, and static comparisons, which often result in inconsistent, time-consuming, and inaccurate pricing.

The gradual adoption of digital technologies presents an opportunity to transform real estate valuation through intelligent systems. Machine learning techniques can analyze historical property data and identify pricing patterns, while web-based platforms enable real-time access to predictions. Digital solutions can reduce human bias, improve transparency, and support informed decision-making.

This project proposes a Real Estate Price Prediction System, an integrated solution to address these challenges. It combines machine learning models, data-driven analysis, and a web-based interface to deliver accurate and real-time property price estimation, thereby improving efficiency, reliability, and usability.

II. SCOPE AND OBJECTIVES

The scope of this project encompasses the development of a machine learning-based system that enables accurate and real-time real estate price prediction by analyzing key property attributes and historical data. The system supports user input for property details such as location, total area, number of bedrooms, and bathrooms, and integrates a web-based interface for instant price estimation. It focuses specifically on residential property valuation within a selected city or region and is designed to operate initially on a limited dataset with potential for scalable expansion. The system also includes features such as data preprocessing, feature encoding, and model integration to ensure reliable and consistent predictions while providing an efficient, user-friendly, and technology-driven solution for property price estimation.

Objectives:

The primary objectives of this project are to develop an accurate and automated real estate price prediction system that provides users with reliable property value estimates based on key features while supporting informed decision-making. The system aims to simplify property valuation by offering real-time price prediction based on inputs such as location, total area, number of bedrooms, and bathrooms, supported by data preprocessing and trained machine learning models. It also seeks to

reduce dependency on manual estimation and broker-based pricing, minimize pricing inconsistencies, and improve transparency through data-driven analysis. Additionally, the system intends to provide a user-friendly web interface, ensure consistent prediction results, and establish a scalable framework that can be extended to support larger datasets, additional locations, and advanced predictive models in future developments.

III. RESEARCH METHODOLOGY AND WORKFLOW

Data Acquisition and Preparation:

The data acquisition and preparation process for the Real Estate Price Prediction System involves collecting accurate and relevant historical real estate data, including property location, total area, number of bedrooms (BHK), number of bathrooms, and corresponding prices. The data is obtained from publicly available datasets and verified sources to ensure reliability. Once collected, the dataset is standardized, cleaned, and validated to remove inconsistencies such as missing values, duplicate records, and incorrect entries. Categorical features such as location are processed using appropriate encoding techniques, while numerical features are normalized to improve model performance. The prepared dataset is structured into a suitable format for machine learning training and testing. This systematic data preparation ensures reliable, consistent, and high-quality input data that supports accurate model learning and effective price prediction.

Machine Learning Model Development:

The machine learning model development phase focuses on building an effective price prediction model using regression-based algorithms. The process begins with selecting suitable machine learning techniques based on the nature of structured real estate data. Feature selection is performed to identify attributes that significantly influence property prices. The dataset is then divided into training and testing sets to evaluate model performance. The model is trained iteratively to learn relationships between property features and prices, and performance metrics are analysed to improve prediction accuracy. Once validated, the trained model is saved as a reusable artefact for deployment.

Continuous evaluation and tuning are carried out to ensure the model remains accurate, reliable, and capable of handling real-world property price prediction scenarios efficiently.

Backend and Web Application Development:

The backend of the Real Estate Price Prediction System is developed using the Flask framework to handle user requests and integrate the trained machine learning model. The backend provides RESTful API endpoints that accept user inputs, process data, and perform real-time price prediction. A web-based frontend is designed to allow users to enter property details and view predicted prices instantly. The frontend communicates with the backend through HTTP requests, ensuring smooth data flow and real-time interaction. The interface is designed to be simple, intuitive, and user-friendly while maintaining accuracy and responsiveness in prediction results.

Model Integration and Prediction Workflow:

The trained machine learning model is integrated with the Flask backend to enable real-time prediction. When a user submits property details through the web interface, the input data is validated and processed by the backend. The processed data is then passed to the machine learning model, which generates the predicted property price based on learned patterns. The predicted result is returned to the backend and displayed to the user. This automated workflow ensures fast response time, low latency, and accurate price estimation without manual intervention.

Deployment:

The deployment phase involves hosting the Real Estate Price Prediction System on a secure and reliable environment where users can access it seamlessly. The application is deployed using a stable server configuration that supports the web interface, backend services, and machine learning model execution. Deployment steps include server configuration, model loading, API testing, and frontend-backend integration testing. Once deployed, the system is monitored continuously to ensure stable performance, security, and availability. This ensures users experience a reliable, efficient, and responsive real estate price prediction application.

IV. LITERATURE SURVRY

The literature survey examines existing real estate valuation methods, machine learning models, and property price prediction systems to understand their strengths and limitations. Studies indicate that traditional property valuation approaches often suffer from subjectivity, pricing inconsistencies, and limited consideration of dynamic market factors. Research on machine learning-based prediction techniques shows that regression models and data-driven approaches can significantly improve pricing accuracy and consistency. Existing real estate platforms and prediction tools provide basic price insights but often lack real-time prediction, feature-based analysis, or user-friendly deployment. These findings help identify gaps that the proposed system aims to address by offering an accurate, real-time, and accessible machine learning-based solution for property price prediction.

V. SYSTEM ARCHITECTURE

The system uses a three-tier architecture with the following components:

User Interface Layer (Web Application):

The User Interface Layer of the Real Estate Price Prediction System provides a simple and accessible platform for users through a web-based application. This layer focuses on clear layout, easy navigation, and intuitive input forms to ensure users can

easily enter property details such as location, total area, number of bedrooms, and bathrooms. Users can submit their inputs and instantly view the predicted property price with minimal interaction.

The interface is designed to be lightweight, responsive, and user-friendly, making it suitable for users with varying levels of technical knowledge.

Application Layer:

The Application Layer enables interaction between the user interface and the machine learning model through a backend server. This layer processes user requests, validates input data, and handles real-time price prediction operations. It manages communication between the frontend and the prediction model, ensuring that user inputs are correctly formatted and securely transmitted. The Application Layer also handles essential backend operations such as request routing, model execution, and response generation, making the system reliable and efficient.

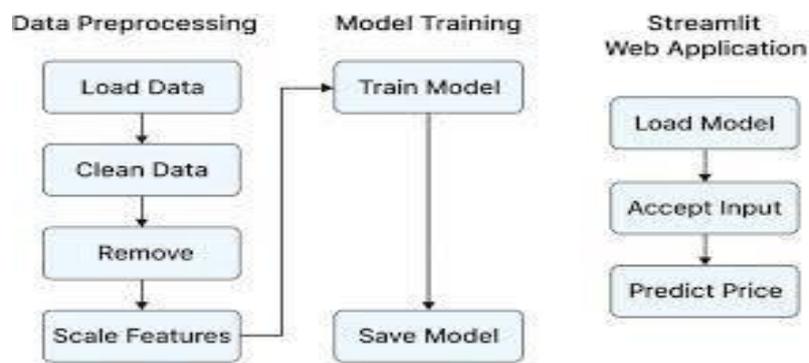
Model and Data Layer:

The Model and Data Layer stores and manages all critical components required for price prediction. It contains the trained machine learning model, feature encoding information, and structured datasets used during prediction. This layer ensures that model artefacts are securely stored and easily retrievable by the Application Layer. It acts as the backbone of the system by maintaining consistency, accuracy, and reliability of the prediction process.

Component Interconnection:

Component Interconnection refers to how different parts of the system communicate and function together. In the Real Estate Price Prediction System, the User Interface Layer communicates with the Application Layer through HTTP requests, while the Application Layer interacts with the Model and Data Layer to generate predictions.

This structured communication ensures secure data transmission, efficient processing, and seamless real-time interaction between users and the prediction system.



VI.IMPLEMENTATION

Machine Learning Model Module:

The Machine Learning Model Module forms the core of the Real Estate Price Prediction System and is responsible for generating accurate property price estimates. This module is implemented using regression-based machine learning techniques to analyze historical real estate data. Key property attributes such as location, total area, number of bedrooms (BHK), and number of bathrooms are used as input features. The dataset undergoes preprocessing, feature encoding, and normalization before training. The model is trained iteratively and evaluated to ensure reliable prediction performance. Once trained, the model is saved and reused for real-time prediction, enabling automated and data-driven property evaluation.

Prediction and Data Processing Module:

The Prediction and Data Processing Module handles the transformation of user input into a format suitable for the machine learning model. When users enter property details through the web interface, this module validates the input, applies necessary preprocessing steps, and encodes categorical features such as location. It ensures that input data matches the structure expected by the trained model. This module plays a critical role in maintaining prediction accuracy by ensuring consistent data flow between the user interface and the machine learning model.

Backend Module:

The Backend Module is implemented using the Flask framework and acts as the central control unit of the system. It manages API requests, handles communication between the frontend and the machine learning model, and performs real-time inference. The backend loads the trained model, processes incoming user requests, and returns predicted prices efficiently. It also ensures secure data handling, input validation, and smooth integration of all system components, making the application reliable and responsive.

Frontend Module:

The Frontend Module provides a simple and user-friendly web interface that allows users to interact with the system easily. It is developed using HTML, CSS, and JavaScript to ensure responsiveness and clarity. Users can enter property details, submit

prediction requests, and view results instantly. The interface is designed to minimize user effort while presenting prediction results in a clear and understandable manner, making the system accessible to users with varying technical backgrounds.

Testing and Deployment:

The Real Estate Price Prediction System undergoes systematic testing to ensure proper functionality and performance. Unit testing is conducted to validate individual components such as input handling and prediction logic, while integration testing ensures smooth communication between frontend, backend, and model modules. Once tested, the system is deployed on a reliable server environment that supports backend services and web access.

Continuous monitoring and periodic updates ensure the system remains stable, secure, and efficient for real-world usage.

VII.RECOMMENDATIONS FOR FUTURE RESEARCH

Integration of Real-Time Market Data:

Future research can focus on integrating real-time real estate market data into the prediction system, such as current property listings, recent sales records, and government valuation guidelines. This live data can be continuously fed into the system to retrain models and improve prediction accuracy. Combining historical data with real-time market trends can help generate more dynamic and up-to-date price estimates, enabling users to make better investment and purchasing decisions.

Enhancement of Machine Learning Models:

Future improvements can enhance the prediction system by incorporating advanced machine learning and deep learning techniques. Models such as ensemble learning, gradient boosting, or neural networks can be explored to capture complex relationships between property features and prices. Adaptive models that learn continuously from new data can further improve prediction reliability and handle changing market conditions more effectively.

Use of Predictive and Trend Analysis:

Future research can incorporate predictive analytics to analyse long-term price trends, demand patterns, and market growth across different regions. By studying historical price movements and economic indicators, the system can provide insights into future property value trends. This can help users identify high-growth areas, make strategic investment decisions, and reduce risks associated with property transactions.

Integration with Geographic Information Systems (GIS):

Future research can explore integrating Geographic Information Systems (GIS) into the platform to enhance location-based analysis. GIS data such as proximity to schools, hospitals, transportation, and commercial hubs can be included as additional features for prediction. This integration can improve location intelligence, provide visual insights, and enhance the accuracy of property price estimation.

Development of Mobile and Cloud- Based Platforms:

Future development can focus on deploying the system as a mobile and cloud-based application to improve accessibility and scalability. Cloud integration can support large datasets, faster computation, and multi-city deployment, while a mobile application can allow users to access price predictions anytime and anywhere. This approach ensures wider adoption, improved performance, and better user experience across different platforms.

VIII.CONCLUSION

The Real Estate Price Prediction System provides an effective digital solution for estimating property prices accurately and efficiently using data science and machine learning techniques. By integrating modules for data preprocessing, machine learning-based prediction, backend processing, and a user- friendly web interface, the system improves accuracy, transparency, and ease of use in property valuation.

Future enhancements such as real-time market data integration, advanced predictive models, geographic analysis, and mobile or cloud-based deployment can further improve system performance and accessibility. Overall, the project demonstrates the practical application of machine learning in real estate valuation and highlights the potential of intelligent technologies to support informed decision- making and modernize property pricing processes.

REFERENCES

1. Selim, H., "Determinants of House Prices in Urban Areas: A Hedonic Regression Model," *Journal of Housing Economics*, Vol. 18, No. 1, pp. 73–84, 2009.
2. Malpezzi, S., "Hedonic Pricing Models: A Selective and Applied Review," *Housing Economics*, Springer Publications, pp. 67–89, 2008.
3. Breiman, L., "Random Forests," *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001
4. Tibshirani, R., "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Series B*, Vol. 58, No. 1, pp. 267–288, 1996.
5. Zhao, Y., and Liu, H., "Machine Learning Approaches for Real Estate Market Prediction," *Journal of Real Estate Analytics*, Vol. 12, No. 3, pp. 145–156, 2021
6. Bourassa, S. C., Hoesli, M., and Peng, V. S., "Do Housing Submarkets Really Matter?" *Journal of Housing Economics*, Vol. 12, No. 1, pp. 12–28, 2003.