



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Property Price Prediction using Machine Learning

Prof. Karpagam V¹, Pandiselvam P², Rajakirubaharan E³, Sabarikumar K B⁴

Professor and Faculty Mentor, Dept. of Computer Science and Engineering, K L N College of Engineering, Madurai,
Tamil Nadu, India

B. E Student, Dept. of Computer Science and Engineering, K L N College of Engineering, Madurai, Tamil Nadu, India

ABSTRACT: Property price estimation is an essential use of machine learning in the real estate sector, allowing for price estimation with precision using several impacting parameters. The present research recommends a predictive model for property valuation based on machine learning, taking into account essential parameters including location, bedrooms, bathrooms, year of built, furnishing status, and existing facilities. The dataset is pre-processed to address missing values, categorical representation, and feature scaling. Different machine learning models such as Linear Regression, Decision Trees, Random Forest, and Gradient Boosting are used and compared to identify the best-performing model. Performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) are utilized to measure the effectiveness of the model. The findings show that adding location and amenities greatly improves the predictive power. This research offers useful information for real estate investors, sellers, and buyers to make informed data-driven pricing decisions, enhancing market transparency and efficiency.

KEYWORDS: Text detection, Inpainting, Morphological operations, Connected component labelling.

I. INTRODUCTION

Applying machine learning to forecast property values from characteristics such as location, area, amenities, and market trends, delivering precise and data-driven insights for real estate choice. Vision: To transform real estate valuation with cutting-edge predictive analytics and technology. Conventional price estimation techniques used for property valuations are prone to inaccuracy and do not reflect the sophisticated interactions between a range of parameters, resulting in questionable predictions. Sophisticated machine learning algorithms like Random Forest, Gradient Boosting, and Deep Learning process varied property and market information to forecast prices with enhanced precision and speed. These models facilitate improved comprehension of property value patterns, equipping buyers, sellers, and investors with accurate forecasts for informed decision-making.

II. LITERATURE SURVEY

In [1] Real-Trading-Oriented Price Prediction with Explainable Multi objective Optimization in Quantitative Trading-T. Yin, X. Du, W. Zhang, Y. Zhao, B. Han and J. Yan. This Model improves financial market forecasts with machine learning and multi-objective optimization to enhance accuracy, risk management, and transparency. Through the use of deep learning, reinforcement learning, and gradient boosting, the system examines past data and market trends to predict asset prices and optimize trading strategies. In contrast to conventional models that are only concerned with returns, it optimizes multiple objectives such as reducing risk and maximizing Sharpe ratios. Explainable AI (XAI) guarantees interpretability to allow for well-informed decisions. The model dynamically real-time reacts to market change by optimizing trade-offs between transaction costs, liquidity needs, and volatility in order to realize more effective quantitative trading.

In [2] House Price Prediction: A Multi-Source Data Fusion Perspective," in Big Data Mining and Analytics-Y. Zhao, J. Zhao and E. Y. Lam. This research improves house price forecasting by incorporating multi-source data fusion, fusing structured and unstructured data such as economic indicators, infrastructure, demographics, and social media data. Conventional models are based on sparse datasets, but this method pools data from housing records, real estate websites, satellite images, and urban reports. Machine learning methods like Random Forest, Gradient Boosting, and Deep Learning enhance the accuracy of forecasting. Spatial and temporal analysis records geographic and market movements, and data fusion enhances computational efficiency. The scalable model is

useful to investors, buyers, and policymakers, enhancing market transparency, real estate appraisal, and urban planning choices.

In [3] Real-Estate Price Prediction System using Machine Learning-V. Gampala, N. Y. Sai and T. N. Sai Bhavya. This research improves machine learning-based real estate price prediction beyond statistical models. It examines parameters such as location, size of property, amenities, and market conditions using historical prices, demographics, and economics. Techniques like Linear Regression, Random Forest, and Gradient Boosting enhance precision with feature selection and preprocessing of data. Assessed using RMSE and MAE, the system serves investors and buyers with real-time data-driven insights for optimal decision-making and market trend analysis.

In [4] Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along with Joint Self-Attention Mechanism-P.-Y. Wang, C.-T. Chen, J.-W. Su, T.-Y. Wang and S.-Huang. This research improves real estate price forecasting through deep learning and a Joint Self-Attention Mechanism (JSAM) for inspection of heterogeneous information. By combining structured data (property location, size) and unstructured data (demographics, satellite images), the model captures intricate interactions for enhanced accuracy. Neural networks such as CNNs and RNNs handle spatial and temporal data dynamically altering feature significance. Tested on MAE and RMSE, the system performs better than conventional models, providing accurate, data-driven real estate pricing estimations.

III. EXISTING SYSTEM

In today's real estate market, property price estimation is typically manual, based on historical trends and expert opinion, resulting in inconsistencies and potential errors. Conventional methods are dependent on real estate agents, survey reports, and market comparisons, which can be subjective and time-consuming. Furthermore, prevailing automated valuation models tend to use limited parameters, excluding important factors like furnishing status and specific amenities. Most property listing sites give approximate prices derived from broad trends instead of individualized predictions. Furthermore, location-specific variations and market trends are not adequately considered, leading to price divergences. Lack of sophisticated machine learning approaches in most current systems limits prediction precision, leaving buyers and sellers at a disadvantage in making educated decisions. Therefore, there is a requirement for a stronger and data-based methodology that takes into account several property characteristics to improve the accuracy of price forecasts.

IV. PROPOSED SYSTEM

The suggested system utilizes machine learning methods to maximize the accuracy of property price forecasting by taking into account various impactful factors like location, the number of bedrooms, the number of bathrooms, year built, furnishing status, and accessible amenities. The system pre-processed real estate data via feature selection, normalization, and categorical encoding to optimize model performance. Multiple machine learning models such as Linear Regression, Decision Tree, Random Forest, and Gradient Boosting are utilized and compared to identify the best-performing model. The model is trained on past property data so that it can understand intricate relationships among various attributes and market trends. Performance is tested with metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) to achieve high accuracy. By incorporating a data-driven methodology, this system allows buyers, sellers, and investors to make better-informed pricing decisions, lessening uncertainty and enhancing market transparency. The system's automated nature also reduces human bias and increases efficiency in real estate valuation.

V. METHODOLOGY

5.1 Data Collection

The basis of the property price forecasting system is a detailed real estate database that has details of different properties, including attributes such as area, number of bedrooms, number of bathrooms, year of construction, furnishing status, and amenities offered, along with the sale price. The data can be obtained from open real estate websites, government databases, or property websites. It is important that the dataset is representative and includes various types of properties in different locations to make the model generalizable. This module loads, checks, and

stores the raw data for processing. Any inconsistencies or missing values detected at this point are marked for correction during preprocessing.

5.2 Data Pre-Processing

The pre-processing block makes the raw data machine learning-ready through cleaning and converting it into a working format. Missing or null values are dealt with first through effective strategies like mean/mode imputation or deletion of incomplete records. Second, categorical variables like furnishing status and location are represented in numerical form through mechanisms like label encoding or one-hot encoding. Outliers are detected and treated to prevent biased model training. Numerical features are normalized or standardized so that all attributes have an equal influence on the prediction. Data splitting into training and test sets to cross-validate the model's performance on new data is also part of this step. Proper pre-processing improves model performance and minimizes error possibilities during training.

5.3 Feature Extraction

In the feature extraction module, the appropriate features of the dataset are chosen as input to the machine learning models. At this stage, the most significant features that have an impact on the property price are chosen, including the number of bedrooms and bathrooms, area, year of build, type of furnishing, and availability of amenities (e.g., car park, gym, swimming pool). Statistical techniques and also domain expertise can be utilized to decrease dimensionality and eliminate useless or redundant features, enhancing the efficiency as well as the accuracy of model training. In a few instances, feature engineering methodologies can be used to construct new meaningful variables, like property age or a location-specific price index, to further improve predictive capability.

5.4 Training & Model Evaluation

This module deals with developing and testing the performance of different machine learning models on the pre-processed dataset. The algorithms chosen for training are Linear Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM). Each model is trained on the training set of the dataset and tested on the held-out test data. The system employs major evaluation criteria such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) to evaluate the precision and effectiveness of every algorithm. Of these, Random Forest tends to perform most optimally because it can manage non-linear relationships and prevent overfitting. The module of evaluation ensures that the model deployed is most accurate and dependable among all the tested models.

5.5 Price Prediction

The price prediction module is the last step of the system in which the trained machine learning model is utilized to forecast the price of a property using input features from the user. As a user inputs property information like location, bedrooms and bathrooms, year constructed, furnishing status, and amenities, these inputs are processed by the same pre-processing and feature extraction pipeline applied while training. The best model, usually the one with the highest scores in evaluation (e.g., Random Forest), then produces an estimated price. This module gives a quick, dependable, and easy-to-use interface that may be embedded within real estate portals, allowing stakeholders to make knowledgeable decisions using information-driven inputs.

VI. SYSTEM ARCHITECTURE

A Software Architecture Diagram gives a top-down view of a system, showing its essential elements, relationships, and data flow. It features modules that process particular functionalities, interfaces that specify communication among components, and data flow paths that outline how information is processed. It can also reflect deployment information, indicating how components are spread over servers or cloud hosts. This diagram acts as a roadmap for developers and stakeholders, providing clarity, scalability, and streamlined system design while allowing collaboration and decision making.

The property price prediction architecture diagram is as shown in Fig 1.



FIG1.ARCHITECTURE OF THE SYSTEM

The System shows the process flow of house price prediction with machine learning models. It begins with the data acquisition, which is followed by data preprocessing to deal with missing values, categorical data encoding, and feature normalization. The pre-processed data is divided into a training set and a testing set. The training set is utilized to build Random Forest and Gradient Boosting models, which are robust ensemble methods. These models are ultimately combined in the final ensemble model to increase precision and reliability. The trained model is ultimately used for predicting home prices, with the test set being utilized in order to monitor model performance.

VII. EXPERIMENTAL RESULTS

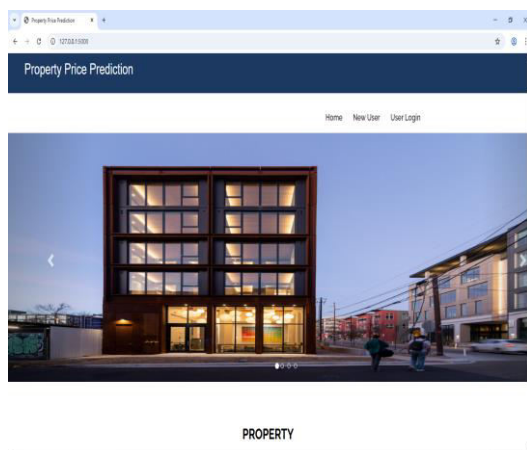


FIG2.HOMEPAGE

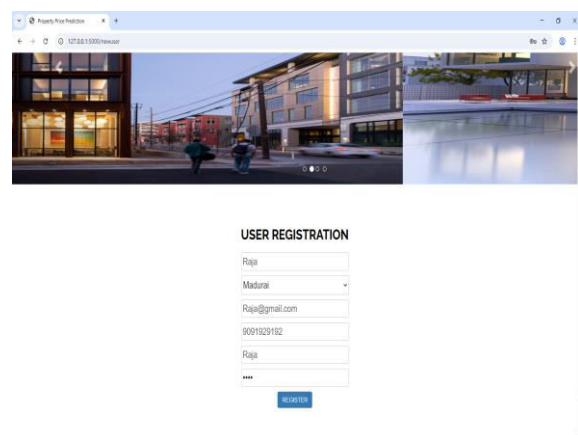


FIG3.USER REGISTRATION

In Fig2 show the homepage of a Property Price Prediction web application, featuring a navigation bar with links for Home, New User, and User Login. Fig3 shows the allows new users to register by entering their personal details. A "REGISTER" button is provided to submit the form.

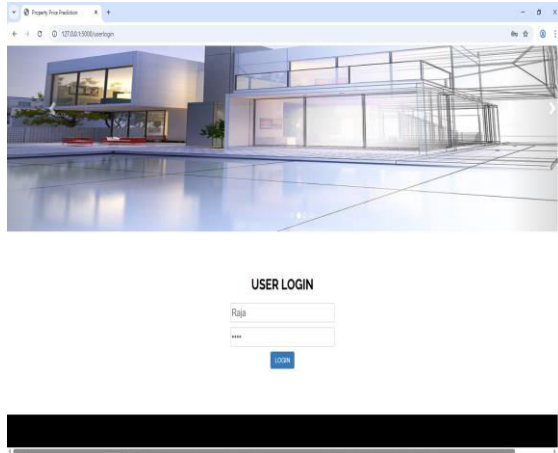


FIG4.USER LOGIN

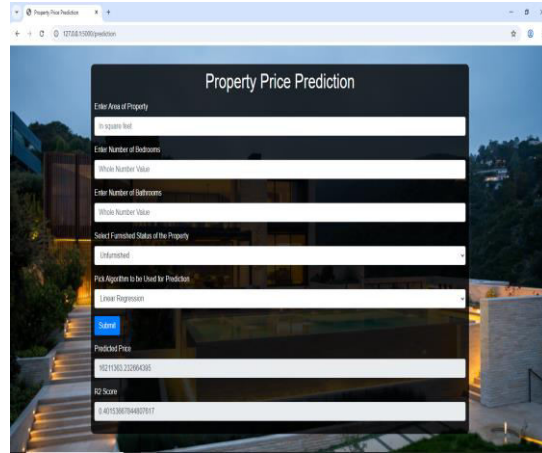


FIG5.PROPERTY PRICE PREDICTION

In Fig 4 shows the provides a login interface where users can enter their username and password. A "LOGIN" button is available to access the Property Price Prediction system. Fig 5 displays the Users enter property details and choose an algorithm to predict the price. The predicted price and R^2 score are then displayed.

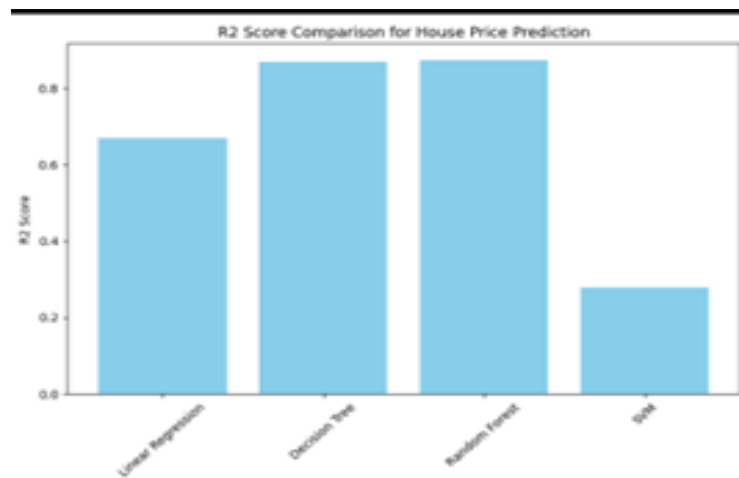


FIG 6. PERFORMANCE ANALYSIS

In Fig 6 Shows the output of the bar chart compares the R^2 scores of different models for house price prediction. Decision Tree and Random Forest show the highest accuracy, while the SVR model performs the least effectively.

VIII. CONCLUSION

The system of property price prediction based on machine learning as proposed offers data-driven, effective, and precise real estate valuation. The key factors of location, number of bedrooms, number of bathrooms, year of construction, furnishing status, and amenities are considered. The system estimates prices with significant improvement over conventional techniques. The incorporation of machine learning patterns allows the model to adjust to market patterns and offer more accurate predictions. Performance measurement employing MAE, MSE, and R^2 attests to the efficacy of the model in minimizing pricing discrepancies. This model not only Favors buyers and sellers by providing improved transparency and decision-making but also increases efficiency in the real estate market. Future enhancements may involve including other parameters like neighbourhood crime statistics, proximity to schools and

hospitals, and economic patterns to further improve accuracy. Overall, this system is a major improvement in property valuation, turning real estate transactions more data-driven and informed.

REFERENCES

- [1] T. Yin, X. Du, W. Zhang, Y. Zhao, B. Han and J. Yan, "Real-Trading-Oriented Price Prediction with Explainable Multi objective Optimization in Quantitative Trading," in IEEE Access, vol. 10, pp. 57685-57695, 2022.
- [2]. Y. Zhao, J. Zhao and E. Y. Lam, "House Price Prediction: A Multi-Source Data Fusion Perspective," in Big Data Mining and Analytics, vol. 7, no. 3, pp. 603-620, September 2024.
- [3] V. Gampala, N. Y. Sai and T. N. Sai Bhavya IEEE Access, "Real-Estate Price Prediction System using Machine Learning" IEEE ACCESS 2024, vol 7, no. 3, pp. 603-620.
- [4] P.-Y. Wang, C. -T. Chen, J. -W. Su, T. -Y. Wang and S. -H. Huang, "Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along with Joint Self-Attention Mechanism," in IEEE Access, vol. 9, pp. 55244-55259, 2021.
- [5] Petropoulos, F., & Makridakis, S. (2020). Forecasting the Real Estate Market Using Machine Learning. Journal of Real Estate Research, 45(1), 85–110.
- [6] Sanya, K. M., & Sathish, M. (2021). Housing Price Prediction Using Random Forest and Support Vector Machine. In 2021 International Conference on Intelligent Technologies (CONIT) (pp. 1–5). IEEE.
- [7] Zeng, Y., & Zhou, R. (2018). Research on Second-hand House Price Prediction Based on Machine Learning. In 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA) (pp. 357–360). IEEE.
- [8] Kumar, A., & Gupta, A. (2021). House Price Prediction Using Ensemble Learning Techniques. Journal of Emerging Technologies and Innovative Research (JETIR), 8(6), 473–479.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details