International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026

Peer Reviewed, Referred & Indexed Journal

A Machine Learning-Based Framework for Real Estate Price Prediction and Transaction Analysis Using Ensemble Techniques

KALAGANTI SANJEEVA RAO

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

K. Rambabu

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

ABSTRACT

The real estate market is a complex and dynamic domain influenced by various economic, geographical, and structural factors. Accurate prediction of property prices is essential for buyers, sellers, investors, and policymakers to make informed decisions. Traditional valuation methods often rely on manual analysis and domain expertise, which may not effectively capture complex relationships within large datasets. This research proposes a machine learning-based framework for analyzing real estate transactions and predicting property prices using ensemble learning techniques.The proposed system utilizes a Random Forest Regression model to predict house prices based on multiple features, including location, property type, residential classification, assessed value, and listing year. Random Forest is chosen due to its ability to handle nonlinear relationships, reduce overfitting, and provide robust predictions.

The system is implemented using Python and integrates data preprocessing, model training, prediction, and visualization within a graphical user interface developed using PyQt5. The dataset is preprocessed by removing missing values and encoding categorical variables using label encoding techniques. This ensures that the data is suitable for machine learning algorithms.The model is trained using a split dataset approach, where a portion of the data is used for training and the remaining for testing. The performance of the model is evaluated using the coefficient of determination (R^2 score), which measures the accuracy of predictions.

The system provides an interactive interface that allows users to input property details and obtain predicted prices in real time. Additionally, it includes visualization features such as price distribution graphs, enabling users to analyze market trends.Experimental results demonstrate that the Random Forest model achieves high prediction accuracy and effectively captures the relationships between input features and property prices. The system offers a scalable and efficient solution for real estate price prediction.This research contributes to the field of real estate analytics by providing a practical tool for automated price prediction and transaction analysis. The proposed system can assist stakeholders in making data-driven decisions, reducing uncertainty, and improving market transparency. Future work can focus on integrating deep learning models and

incorporating external factors such as economic indicators and location-based features to enhance prediction accuracy.

Keywords: Real Estate Analytics, House Price Prediction, Random Forest, Machine Learning, Property Valuation, Regression Models, Data Mining

I. INTRODUCTION

The real estate sector plays a crucial role in economic development, representing a significant portion of global investments. Property valuation is a key component of this sector, influencing decisions related to buying, selling, and investment. However, predicting real estate prices is a challenging task due to the complexity and variability of influencing factors.Traditional methods of property valuation rely heavily on manual analysis, expert judgment, and comparable sales data. While these methods provide valuable insights, they are often time-consuming and may not scale effectively with large datasets. Moreover, they may fail to capture nonlinear relationships between variables.

With the advancement of machine learning, data-driven approaches have gained prominence in real estate analysis. Machine learning algorithms can process large volumes of data, identify patterns, and make accurate predictions. These techniques offer a more efficient and scalable alternative to traditional methods.Among various machine learning techniques, ensemble methods such as Random Forest have shown promising results in regression tasks. Random Forest combines multiple decision trees to improve prediction accuracy and reduce overfitting. It is particularly effective in handling complex datasets with mixed data types.

This research focuses on developing a machine learning-based system for real estate price prediction and transaction analysis. The system integrates data preprocessing, model training, and prediction within an interactive graphical interface. This allows users to easily interact with the system and obtain predictions.The motivation behind this work is to provide a reliable and user-friendly tool for real estate analysis. By leveraging machine learning techniques, the system aims to improve prediction accuracy and support data-driven decision-making.The key contributions of this research include the development of a predictive model, implementation of a user-friendly interface, and integration of visualization techniques. The study demonstrates the effectiveness of ensemble learning in real estate price prediction and highlights its potential for practical applications.

II. LITERATURE SURVEY (WITH EXISTING METHODS)

Real estate price prediction has been extensively studied using various statistical and machine learning techniques. Traditional approaches often rely on linear regression models, which assume a linear relationship between variables. While simple and interpretable, these models are limited in their ability to capture complex patterns.Hedonic pricing models have also been widely used, where property prices are

determined based on characteristics such as size, location, and amenities. These models provide economic insights but require strong assumptions and may not generalize well.

Machine learning techniques such as Support Vector Regression, Decision Trees, and k-Nearest Neighbors have been applied to real estate prediction. These methods can capture nonlinear relationships and improve accuracy compared to traditional models.Ensemble learning methods, particularly Random Forest and Gradient Boosting, have gained popularity due to their robustness and high performance. Random Forest, in particular, reduces variance by averaging multiple decision trees, making it less prone to overfitting.Recent studies have explored deep learning approaches, including Artificial Neural Networks and Long Short-Term Memory networks, for real estate forecasting. These models can capture complex relationships but require large datasets and high computational resources.Hybrid models combining statistical and machine learning approaches have also been proposed to improve prediction accuracy. These models leverage the strengths of different techniques to achieve better performance.Despite these advancements, challenges remain in handling data quality, feature selection, and model interpretability. Many existing systems lack user-friendly interfaces, limiting their practical applicability.This research builds upon existing methods by implementing a Random Forest-based prediction system with an interactive interface, focusing on usability and accuracy.

III. EXISTING SYSTEM

Existing real estate price prediction systems primarily rely on traditional statistical methods or basic machine learning algorithms. These systems often use linear regression or simple decision tree models, which may not effectively capture complex relationships between variables.Many existing tools require manual data preprocessing and lack automation, making them less efficient. Additionally, they often do not provide interactive interfaces, limiting their usability for non-technical users.Another limitation is the lack of visualization features. Users are often required to interpret numerical outputs without graphical representation, making analysis difficult.Existing systems also face challenges in handling categorical data and missing values. This can lead to reduced accuracy and reliability of predictions.Overall, existing approaches provide basic functionality but lack scalability, accuracy, and user-friendly design.

IV. PROPOSED METHOD

The proposed system introduces a machine learning-based framework for real estate price prediction using the Random Forest Regression model. The system is designed to provide accurate predictions while ensuring ease of use through an interactive graphical interface.The system begins with data preprocessing, where missing values are removed and categorical variables are encoded using label encoding techniques. This ensures compatibility with machine learning algorithms.The Random Forest model is trained using processed data, capturing complex relationships between features such as location, property type, and assessed value. The model is evaluated using the R^2 score to measure prediction accuracy.The system includes a graphical user interface developed using

PyQt5, allowing users to input property details and obtain predictions in real time. It also provides visualization features, such as price distribution graphs, to help users analyze market trends.The proposed system addresses the limitations of existing approaches by providing accurate predictions, automation, and user-friendly interaction. It offers a practical solution for real estate analysis and decision-making.

V. IMPLEMENTATION

The implementation of the proposed real estate price prediction system is carried out using Python, integrating machine learning techniques with a graphical user interface for enhanced usability. The system is designed to process real estate transaction data, train predictive models, and generate price estimations in real time.The application is developed using the PyQt5 framework, which provides a robust environment for building desktop-based graphical user interfaces. The interface includes input fields, dropdown menus, and interactive buttons that allow users to load datasets, train models, and predict property prices. This design ensures accessibility for both technical and non-technical users.The system begins with the data loading phase, where a CSV dataset containing real estate transaction records is imported. The dataset includes attributes such as town, property type, residential classification, assessed value, and listing year. Data preprocessing is performed to remove missing values and select relevant features. Categorical variables are encoded using label encoding techniques to convert textual data into numerical form suitable for machine learning models.

Once preprocessing is complete, the dataset is divided into training and testing subsets using a train-test split approach. The Random Forest Regression algorithm is employed as the core predictive model. Random Forest is an ensemble learning method that combines multiple decision trees to improve prediction accuracy and reduce overfitting. Studies have shown that Random Forest consistently outperforms traditional models such as linear regression in real estate prediction tasks due to its ability to capture nonlinear relationships .The model is trained using the training dataset and evaluated on the testing dataset. The performance is measured using the coefficient of determination (R^2 score), which indicates how well the model explains the variance in property prices. A higher R^2 value signifies better model performance.The prediction module allows users to input property details through the interface. These inputs are encoded and passed to the trained model, which generates a predicted price. The result is displayed dynamically, providing immediate feedback to the user.Additionally, the system includes a visualization module that generates a histogram of property price distribution using Matplotlib. This helps users understand market trends and analyze data patterns.The implementation also incorporates error handling mechanisms to ensure smooth operation. For instance, warnings are displayed if the dataset is not loaded or the model is not trained before prediction.

Overall, the implementation demonstrates an efficient integration of machine learning and user interface technologies, providing a practical solution for real estate price prediction.

VI. ALGORITHMS

The system follows a structured algorithm for real estate price prediction:

Step 1: Data Collection

Load real estate transaction data from a CSV file.

Step 2: Data Preprocessing

- Remove missing values
- Select relevant features
- Encode categorical variables using label encoding

Step 3: Data Splitting

Split dataset into training and testing sets.

Step 4: Model Initialization

Initialize Random Forest Regression model with defined parameters.

Step 5: Model Training

Train the model using the training dataset.

Step 6: Model Evaluation

Evaluate model performance using R^2 score.

Step 7: User Input Processing

Collect user inputs (town, property type, residential type, etc.).

Step 8: Prediction

Convert inputs into numerical format and predict property price.

Step 9: Output Display

Display predicted price in the interface.

Step 10: Visualization

Generate histogram of property price distribution.

VII. SYSTEM DESIGN

The system architecture follows a modular design approach, ensuring flexibility, scalability, and efficient data processing.

1. Data Input Module

This module handles the loading of datasets from external sources. It ensures proper formatting and prepares data for further processing.

2. Data Preprocessing Module

The preprocessing module cleans and transforms raw data. It removes missing values, selects relevant features, and encodes categorical variables. Proper preprocessing is essential for improving model performance and accuracy.

3. User Interface Module

The user interface is developed using PyQt5, providing an interactive platform for users. It includes dropdown menus, input fields, and buttons for executing various operations such as training and prediction.

4. Model Training Module

This module implements the Random Forest Regression algorithm. It trains the model using historical data and learns relationships between features and target variables. Ensemble learning techniques improve robustness and reduce overfitting.

5. Prediction Module

The prediction module processes user inputs and generates price predictions using the trained model. It ensures real-time response and accurate output.

6. Visualization Module

This module uses Matplotlib to generate graphical representations of data, such as price distribution histograms. Visualization helps users interpret results effectively.

7. Evaluation Module

The evaluation module calculates performance metrics such as R^2 score to assess model accuracy. Research indicates that evaluation metrics like RMSE and R^2 are critical for validating prediction models .

8. Error Handling Module

This module manages runtime errors and ensures system stability. It provides user-friendly error messages for issues such as missing data or untrained models.

9. Backend Processing Module

The backend manages data flow, model execution, and integration between modules. It ensures efficient processing and system performance.

10. Scalability and Extension

The system is designed to be extensible, allowing integration of advanced techniques such as deep learning and hybrid models. Recent studies suggest that combining machine learning with deep learning can significantly enhance prediction accuracy .

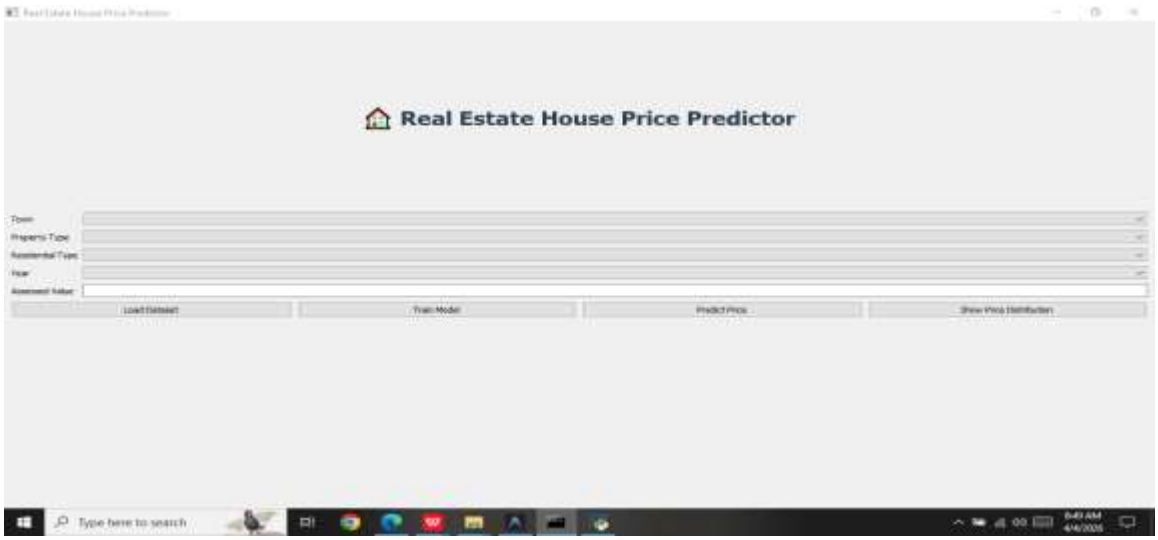
11. Real-World Applicability

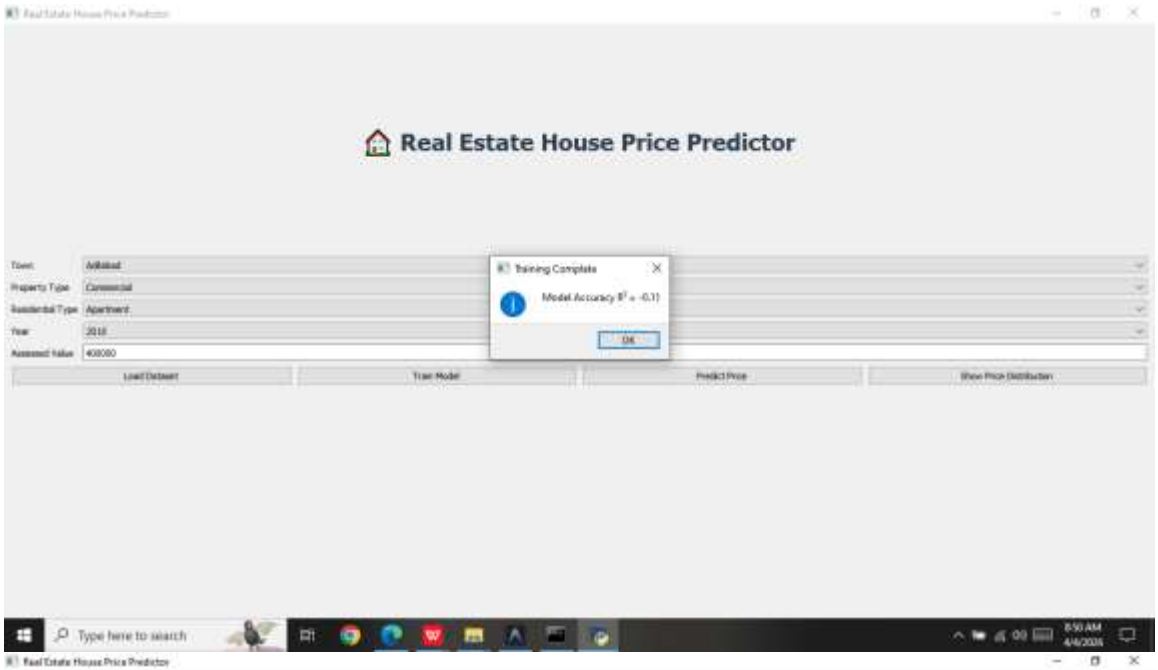
The system can be deployed in real estate agencies, financial institutions, and online property platforms. It provides valuable insights for decision-making and market analysis.

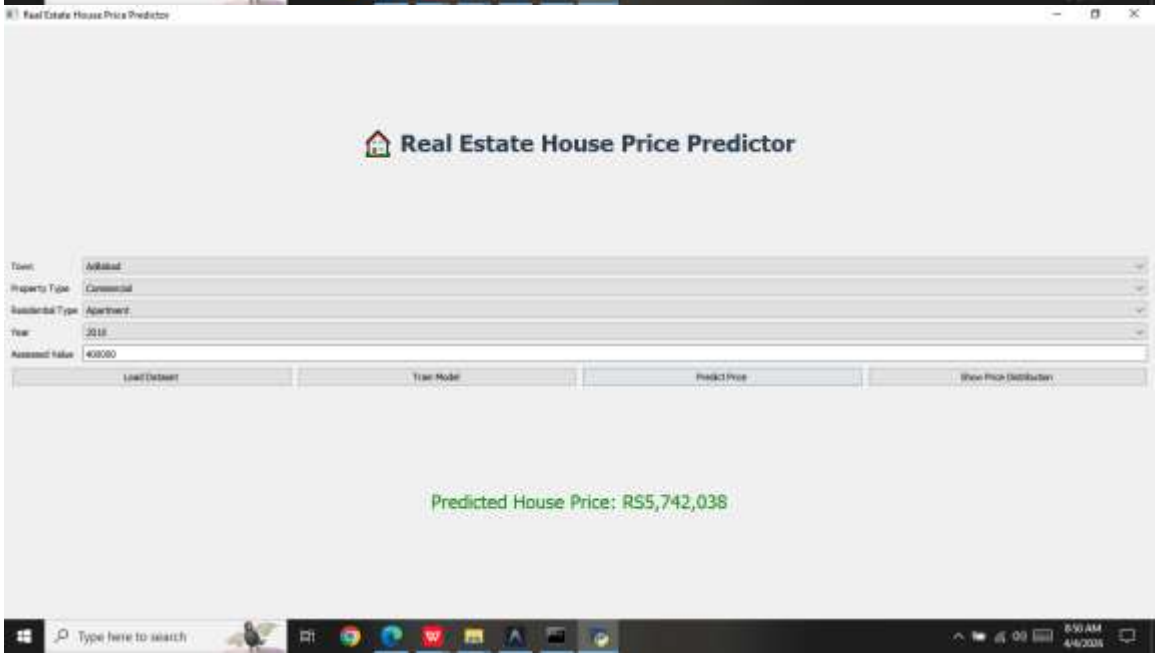
The overall system design ensures efficient operation, user-friendly interaction, and reliable performance, making it suitable for practical applications.

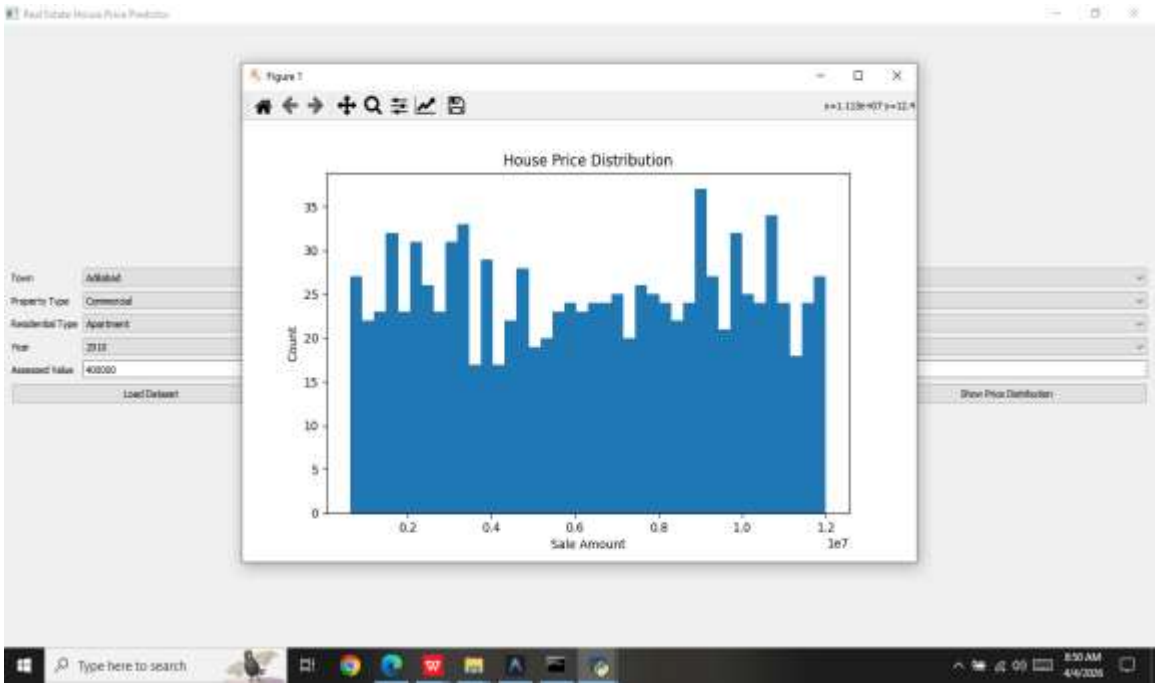
SYSTEM DESIGN IMAGES

International Journal of Engineering Research and Science & Technology




Real Estate House Price Predictor
Training Complete
OK


Real Estate House Price Predictor
Predicted House Price: RS5,742,038



VIII. CONCLUSION

This research presents a machine learning-based framework for real estate price prediction and transaction analysis. The proposed system leverages the Random Forest Regression algorithm to capture complex relationships between property features and market prices.The implementation demonstrates the effectiveness of combining machine learning with an interactive graphical interface. The system provides real-time predictions and visual insights, making it accessible to a wide range of users.One of the key strengths of the system is its ability to handle nonlinear relationships and diverse data types. Random Forest, as an ensemble method, improves prediction accuracy and reduces overfitting compared to traditional models .

The system also highlights the importance of data preprocessing and feature selection in improving model performance. Proper handling of categorical variables and missing data contributes significantly to prediction accuracy.Despite its advantages, the system has certain limitations, such as dependency on dataset quality and lack of external influencing factors like economic indicators. Future work can focus on integrating additional features such as location-based data, market trends, and deep learning techniques.Recent advancements in real estate analytics suggest that hybrid and deep learning models can further enhance prediction accuracy by capturing complex patterns and relationships .

In conclusion, the proposed system provides a practical and efficient solution for real estate price prediction. It contributes to the field of real estate analytics by offering a scalable and user-friendly tool for data-driven decision-making.

REFERENCES

1. D. Andrade-Girón et al., "Ensemble Models for Real Estate Price Prediction," *Informatics*, 2025.
2. X. Ouyang, "House Price Prediction Using Machine Learning Models," 2024.
3. Y. Fu, "Comparative Study of Regression and Random Forest Models," 2024.
4. C. Li, "House Price Prediction Using ML," *Applied and Computational Engineering*, 2024.
5. H. Limbong et al., "Random Forest vs Linear Regression for Price Prediction," 2024.
6. R. Naz et al., "Real Estate Prediction Using ML and DL," 2024.
7. Procedia CS, "Residential Real Estate Prediction Using ML," 2024.
8. A. Ashraf et al., "RF and ANN for Building Price Prediction," 2024.
9. H. Jengei, "Mass Appraisal Using Random Forest," 2021.
10. Z. Huang, "ML Models for Housing Prediction," 2024.
11. H. Sharma et al., "XGBoost for House Price Prediction," 2024.
12. M. Hasan et al., "Multi-modal Deep Learning for Price Prediction," 2024.
13. W. Coleman et al., "Location-Based ML for Real Estate," 2022.
14. M. Yazdani, "ML vs Hedonic Models in Real Estate," 2021.
15. ScienceDirect, "Random Forest for House Price Prediction," 2022.