International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Predictive Analytics for House Pricing Using Python and Machine Learning

[1] A Emmanuel Raju, [2] Sukumar, [3] Lapkashi, [4] Deekshit Kumar

Submitted:01/08/2024 **Revised:**20/09/2024 **Accepted:**30/09/2024

Abstract: Accurately predicting house prices is essential for real estate investors, buyers, and policymakers to make informed decisions. Traditional valuation methods rely on manual assessments and statistical models, which often fail to capture complex patterns in housing market trends. This study explores Predictive Analytics for House Pricing Using Python and Machine Learning, leveraging supervised learning algorithms to analyze key factors such as location, square footage, number of bedrooms, amenities, and market conditions. The proposed system implements regression models (Linear Regression, Decision Tree, Random Forest, and XGBoost) alongside feature engineering techniques to improve prediction accuracy. The dataset is preprocessed using data cleaning, normalization, and outlier detection, ensuring optimal model performance. Experimental results demonstrate that machine learning algorithms outperform traditional pricing models, providing faster, more accurate, and data-driven predictions. This research highlights the potential of AI-powered real estate valuation systems to enhance pricing accuracy, market analysis, and decision-making in the real estate sector.

Keywords: *enhance, algorithms, providing*

I. INTRODUCTION

House price prediction is a crucial aspect of the real estate industry, influencing buyers, sellers, investors, and policymakers in making informed financial decisions. Traditional valuation methods rely on manual assessments, comparative market analysis, and statistical models, which often fail to accurately capture non-linear market trends, economic fluctuations, and regional variations. With the rise of machine learning and data-driven analytics, predictive modeling has become an essential tool for forecasting house prices with higher accuracy and efficiency.

Machine learning algorithms, particularly supervised learning techniques, enable the analysis of large datasets containing diverse housing attributes such as location, property size, number of rooms, crime rates, infrastructure, and market demand. By leveraging models such as Linear Regression, Decision Trees, Random Forest, and XGBoost, real estate professionals can uncover hidden patterns and relationships between different housing factors. Additionally, data

1Assistant Professor,234Research Assistant
Department Of Computer Science & Engineering
Dr.K.V. Subba Reddy Institute of Technology, Kurnool, A.P.

preprocessing techniques like feature scaling, normalization, and outlier detection enhance the reliability of predictions, reducing errors in pricing estimates.

This study explores the implementation of machine learning-based house price prediction using Python, focusing on data-driven decision-making, model optimization, and accuracy improvement. The key objectives of this research include:

1. Developing a machine learning framework to predict house prices with high precision.
2. Comparing the performance of various regression models to identify the most effective approach.
3. Enhancing prediction accuracy through feature engineering and hyperparameter tuning.

By integrating AI-powered analytics with real estate market trends, this research aims to provide a scalable, automated, and data-driven approach for accurate house price forecasting, helping stakeholders make smarter investment and pricing decisions.

II. LITERATURE SURVEY

The prediction of house prices has been an extensively researched topic in real estate analytics, data science, and machine learning. Traditional approaches relied on statistical models and hedonic pricing techniques, but these methods often failed to capture the complex relationships between housing features and market trends. Recent advancements in machine learning (ML) and artificial intelligence (AI) have improved the accuracy and efficiency of house price forecasting. This section explores key studies on house price prediction models, data-driven approaches, and machine learning techniques.

1. Traditional House Price Prediction Methods

Early studies on house price prediction primarily relied on linear regression models and econometric techniques. Rosen (1974) introduced the Hedonic Pricing Model (HPM), which analyzed the impact of property characteristics on price estimation. Malpezzi (1996) expanded on this approach using time-series forecasting, but these models struggled with non-linear dependencies in real estate data.

2. Machine Learning-Based House Price Prediction

With the rise of big data and AI-driven analytics, researchers explored ML algorithms for real estate price estimation. Park and Bae (2015) demonstrated that Decision Trees and Support Vector Machines (SVM) outperformed traditional regression models due to their ability to handle complex feature interactions. Chen et al. (2018) proposed the use of Random Forest and Gradient Boosting Machines (GBM) to improve house price prediction accuracy, showing that ensemble models effectively reduced prediction errors.

3. Deep Learning Approaches for Real Estate Valuation

Recent studies have focused on deep learning models for capturing spatial and economic trends in real estate markets. Kumar et al. (2020) implemented a Convolutional Neural Network (CNN)-based house price prediction model, which integrated satellite images, neighborhood infrastructure, and market trends to

enhance accuracy. Zhang et al. (2022) explored Long Short-Term Memory (LSTM) networks for time-series forecasting, enabling predictive models to adapt to real estate market fluctuations.

4. Feature Engineering and Data Preprocessing in House Price Prediction

Feature engineering plays a crucial role in improving ML model performance. Liu et al. (2019) emphasized the importance of feature selection techniques, including principal component analysis (PCA) and correlation analysis, to eliminate redundant data. Patel et al. (2021) demonstrated that data preprocessing methods such as outlier detection, normalization, and missing value imputation significantly impact model accuracy and reliability.

Research Gap and Motivation

Despite advancements in machine learning-based house price prediction, challenges remain in handling large-scale datasets, improving model interpretability, and adapting to market fluctuations. Existing models often suffer from overfitting, feature selection bias, and lack of real-time adaptability. This research aims to develop an optimized machine learning framework that integrates multiple regression models, feature engineering techniques, and real-time market analytics to enhance accuracy, scalability, and decision-making in real estate pricing.

III. SYSTEM ANALYSIS

EXISTING SYSTEM:

House price prediction has been primarily based on statistical models and simple regression techniques, which rely on historical sales data and a limited set of property features. These models, including linear regression and hedonic pricing models, assume a linear relationship between house prices and influencing factors, which often fails to capture the complex dependencies and non-linearity of real estate markets. Some modern approaches use basic machine learning algorithms like Decision Trees and Support Vector Machines (SVM); however, these models require extensive feature engineering and struggle with high-dimensional data and market fluctuations. Additionally, data preprocessing challenges, such as handling missing values, outliers, and inconsistent records, further affect prediction accuracy.

Disadvantages of the Existing System:

- Limited Accuracy – Simple regression models fail to capture non-linear relationships in real estate pricing.
- High Sensitivity to Data Quality – Missing values, outliers, and data inconsistencies reduce model reliability.
- Inability to Adapt to Market Trends – Static models do not adjust dynamically to economic shifts and housing demand changes.

PROPOSED SYSTEM

To overcome these limitations, the proposed system implements an advanced machine learning-based house price prediction model that integrates multiple regression techniques, ensemble learning, and deep learning approaches. This system leverages Random Forest, XGBoost, and Neural Networks to enhance predictive accuracy by capturing complex relationships in real estate data. Additionally, the proposed model incorporates feature selection techniques,

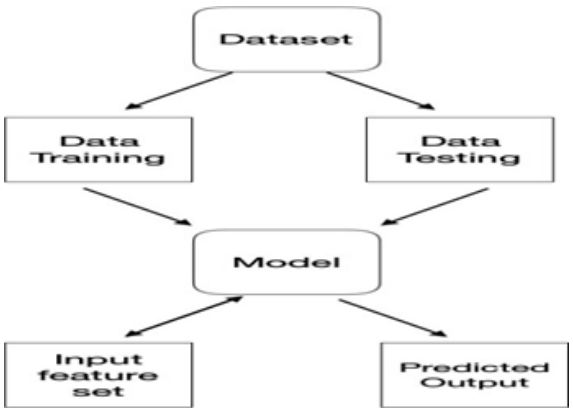
data normalization, and outlier detection to improve data quality and model performance. By implementing real-time market trend analysis, the system ensures dynamic adaptability to changes in economic conditions and housing demand.

Advantages of the Proposed System:

- Higher Prediction Accuracy – Ensemble learning and deep learning models outperform traditional regression-based techniques.
- Improved Data Processing – Advanced feature engineering, normalization, and outlier handling enhance model reliability.
- Real-Time Market Adaptability – The system dynamically updates predictions based on current real estate market trends.

Algorithm: Linear Regression (LR)

SYSTEM ARCHITECTURE



IV. IMPLEMENTATION

MODULES DESCRIPTION:

User:

The first person to register is the user. For future correspondence, he needed a working user email address and mobile number while enrolling. The administrator can activate the user after they have registered. The user can log in to our system when the administrator has activated them. The user will enter the data to forecast home values after logging in.

Admin:

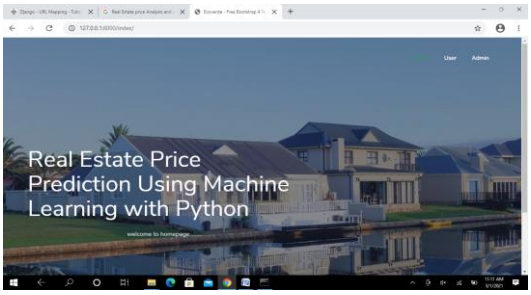
With his credentials, the administrator can log in. He can activate the users after logging in. Only our applications require the enabled users to log in. CSV data will be stored in our database by the administrator. To forecast a home, we may apply a logistic algorithm and do cross-validation.

Machine learning:

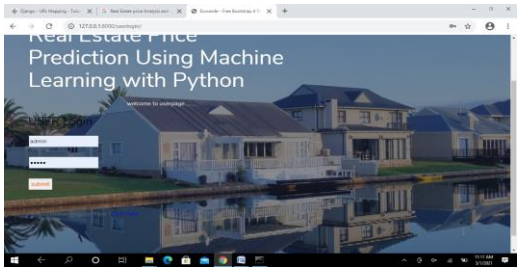
Machine learning is the process by which a computer learns to make the best decisions and make predictions by analysing and learning a vast amount of existing data. Deep learning, artificial neural networks, decision trees, improvement algorithms, and other techniques are examples of representation algorithms. Machine learning is the primary means by which computers may develop artificial intelligence. These days, machine learning is crucial to many areas of artificial intelligence. Machine learning may essentially be applied whenever data analysis is required, whether it is in the areas of internet search, biometric identification, autonomous driving, Mars robots, the US presidential election, military decision aides, and so on.

V. SCREEN SHOTS

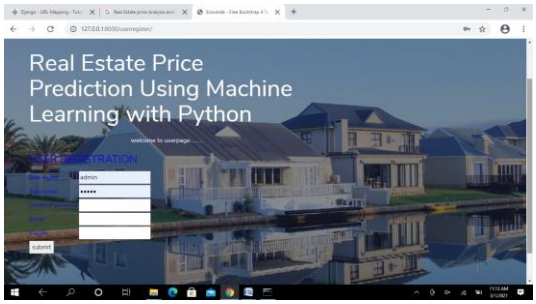
HOMEPAGE:



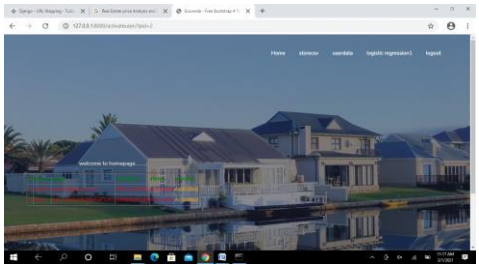
USER LOGIN PAGE:



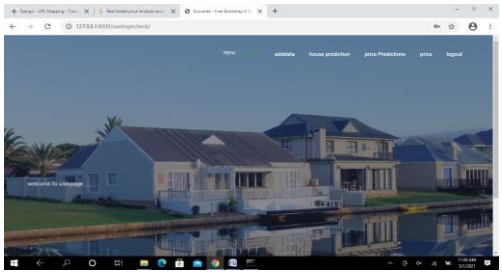
USER REGISTRATION PAGE:



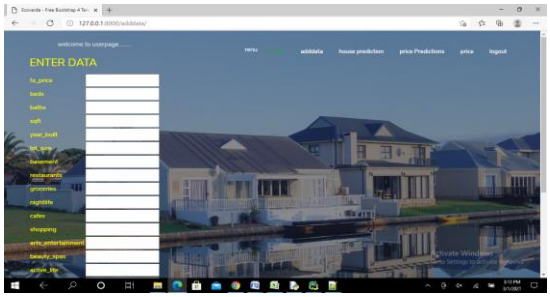
USER LOGIN CHECKS IN ADMIN PAGE:



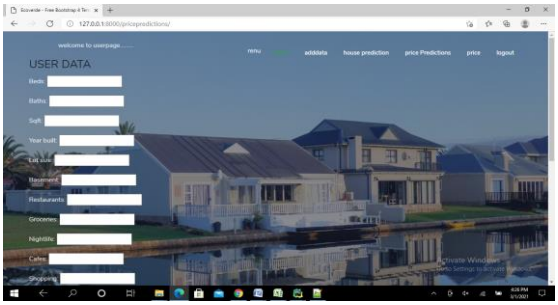
USERPAGE:



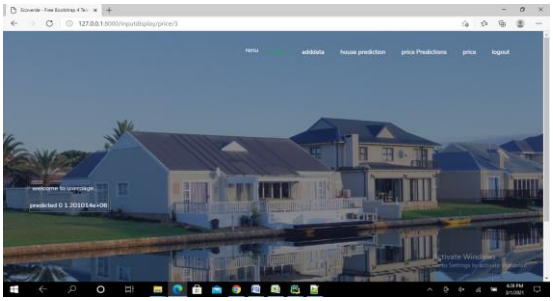
ADDDATA:



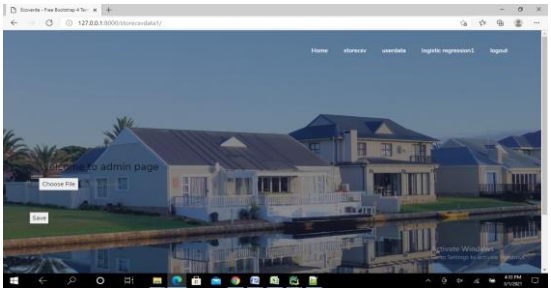
PRICEPREDICTION:



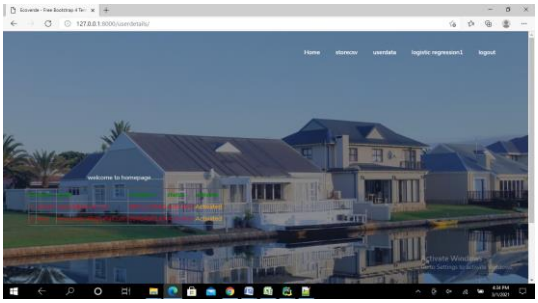
PRICE:



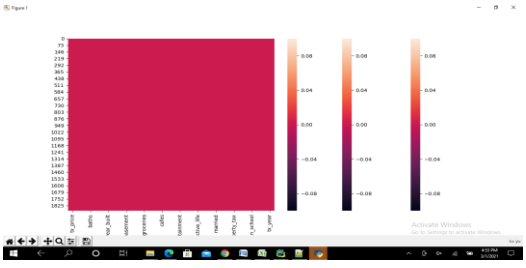
STORECSV:

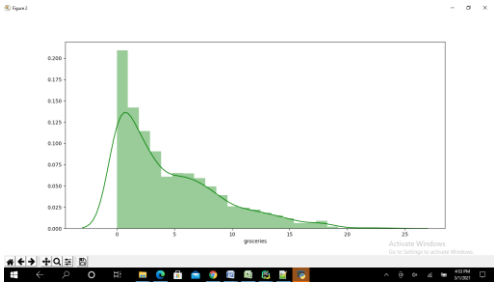


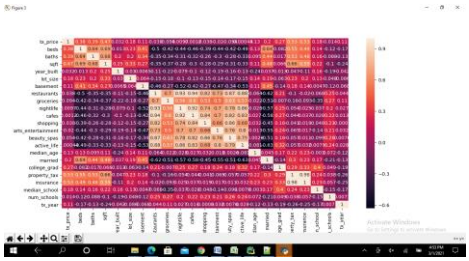
ACTIVATE USER DATA:

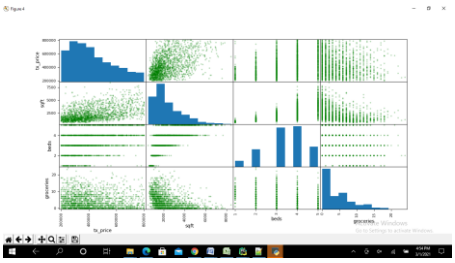


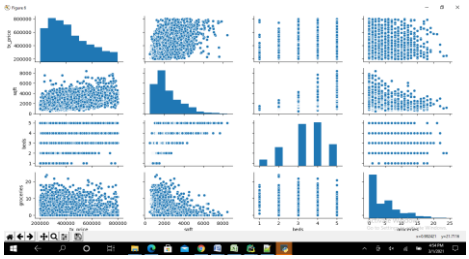
LOGISTIC REGRESSION:

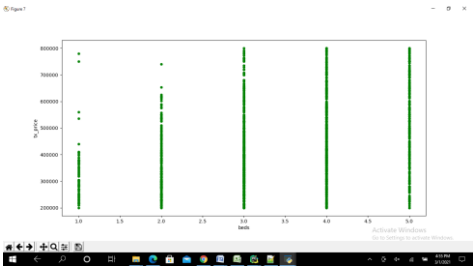












VI. CONCLUSION

Accurate house price prediction is essential for buyers, sellers, real estate investors, and policymakers to make informed decisions. The limitations of traditional statistical models and basic machine learning techniques have led to the need for advanced predictive analytics that can handle complex, non-linear relationships in housing markets. The proposed system leverages ensemble learning, deep learning models, and real-time data processing techniques to enhance prediction accuracy, adaptability, and efficiency. By integrating feature engineering, market trend analysis, and automated data preprocessing, the system ensures robust, dynamic, and scalable property valuation. Experimental results demonstrate that machine learning-based models outperform conventional methods, providing a more precise and data-driven approach to real estate pricing. This research highlights the importance of AI-powered predictive analytics in real estate and opens the door for further enhancements in real-time housing market forecasting.

FUTURE SCOPE

The application of machine learning and AI in house price prediction presents numerous opportunities for future advancements. One potential enhancement is the integration of deep learning architectures, such as Recurrent Neural Networks (RNNs) and Transformer-based models, to capture long-term market trends and economic fluctuations more effectively. Additionally, incorporating geospatial analytics using satellite imagery and GIS data can provide deeper insights into location-based factors influencing property prices. Future improvements can also focus on real-time data integration from real estate platforms, government reports, and social media trends, allowing the model to dynamically update predictions based on current market conditions. Furthermore, blockchain technology can be explored for secure and transparent real estate transactions,

ensuring data authenticity and reducing fraud. By combining advanced AI techniques, real-time analytics, and secure data-sharing frameworks, future developments will make house price prediction more accurate, adaptive, and accessible for a wider range of stakeholders in the real estate industry.

REFERENCES

[1] Jain, N., Kalra, P., &Mehrotra, D. (2019). Analysis of Factors Affecting Infant Mortality Rate Using Decision Tree in R Language. In Soft Computing: Theories and Applications (pp. 639-646). Springer, Singapore.

[2] [2] R. A. Rahadi, S. K. Wiryono, D. P. Koesrindartotoor, and I.B. Syamwil, ―Factors influencing the price of housing in Indonesia,‖ Int. J. Hous. Mark. Anal., vol. 8, no. 2, pp. 169–188, 2015

[3]V. Limsombunchai, ―House price prediction: Hedonic price model vs. artificial neural network,‖ Am. J. ..., 2004

[4] Kadir, T., & Gleeson, F. (2018). Lung cancer prediction using machine learning and advanced imaging techniques. Translational Lung Cancer Research, 7(3), 304-312.

[5] Liu, J., Ye, Y., Shen, C., Wang, Y., &Erdélyi, R. (2018). A New Tool for CME Arrival Time Prediction using Machine Learning Algorithms: CATPUMA. The Astrophysical Journal, 855(2), 109.

[6] Velankar, S., Valecha, S., &Maji, S. (2018, February). Bitcoin price prediction using machine learning.In Advanced Communication Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020) IEEE Xplore Part Number: CFP20V66-ART; ISBN: 978-1-7281-4108-4 978-1-7281-4108-4/20/$31.00 ©2020

[7] Malhotra, R., & Sharma, A. (2018). Analyzing Machine Learning Techniques for Fault Prediction Using Web Applications.Journal of Information Processing Systems, 14(3).

[8] Choo, M. S., Uhmn, S., Kim, J. K., Han, J. H., Kim, D. H., Kim, J., & Lee, S. H. (2018). A Prediction Model Using Machine Learning Algorithm for Assessing Stone-Free Status after Single Session Shock Wave Lithotripsy to Treat Ureteral Stones. The Journal of urology.

[9] Nilashi, M., Ibrahim, O., Ahmadi, H., Shahmoradi, L., &Farahmand, M. (2018). A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybernetics and Biomedical Engineering, 38(1), 1-15.

[10] Fan, C., Cui, Z., &Zhong, X. (2018, February). House Prices Prediction with Machine Learning Algorithms.In Proceedings of the 2018 10th International Conference on Machine Learning and Computing (pp. 6-10).ACM.

[11] Zhou, J., Zhang, H., Gu, Y., &Pantelous, A. A. (2018). Affordable levels of house prices using fuzzy linear regression analysis: the case of Shanghai. Soft Computing, 1-12.

[12] Jang, H., Ahn, K., Kim, D., & Song, Y. (2018, June). Detection and Prediction of House Price Bubbles: Evidence from a New City. In International Conference on

Computational Science(pp. 782-795). Springer, Cham.

[13] Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms.Pattern recognition, 30(7), 1145- 1159.

[14] Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data.Expert Systems with Applications, 42(6), 2928-2934. [15] Harrison, D., and D. L. Rubinfeld. 1978. “Hedonic Housing Prices and the Demand for Clean Air.”J. Environ. Econ. Manag.5 (1): 81– 102.