

Enhancing House Rent Prediction Using A Stacking Ensemble Model for Sustainable Community Development

Abiola, Oladimeji Arowolo¹, Rotimi, Israel Taye², ODUFUWA, Jesufunminire Emmanuel³

^{1,2}Department of Computer Science, University of Ibadan, Nigeria.

³University of Nicosia, Cyprus

ARTICLE INFO	ABSTRACT
Published Online: 07 October 2026	House rent estimation remains a challenge due to inconsistent pricing methods, reliance on manual valuation, and the absence of intelligent systems for determining fair rent values. Traditional approaches used by house agents and property owners often result in price variations and inaccurate rent estimations. This study developed a House Rent Prediction System using stack ensemble to estimate annual house rent values based on housing characteristics. The dataset used for the study was obtained from house agents and secondary sources and was preprocessed through data cleaning, encoding, and feature scaling using StandardScaler. The developed system was implemented using Streamlit and included functionalities such as manual prediction, property visualization, house comparison, prediction history, market analytics, and property carousel display. Model evaluation was carried out using MAE, MSE, RMSE, R ² Score, and K-Fold Cross Validation. The results showed that the developed system was effective in predicting house rent values and can support intelligent housing valuation and decision-making in community development.
Corresponding Author: Rotimi, Israel Taye	
KEYWORDS: House Rent Prediction, Linear Regression, Machine Learning, Streamlit, Housing Valuation.	

I. INTRODUCTION

Housing gives people the necessary security and shelter, making it a basic necessity in community development. However, problems like exorbitant rent costs and a lack of transparency frequently beset the housing market. These problems present significant difficulties for both homebuyers and landlords. Some agents have a tendency to set rent prices too high, disconnecting them from the real value of the properties and the going rates in the market. Recent developments in machine learning methods, however, provide encouraging answers to this problem, as such methods have been shown to reduce the information gap between tenants, investors, and property owners by improving the accuracy and accessibility of rental price predictions, leading to more equitable pricing mechanisms (Nor et al., 2025). In order to produce accurate and insightful price estimations, this work intends to propose a house price prediction model that makes use of machine learning methods. The model intends to aid real estate professionals, homeowners, and investors make wise decisions in the dynamic real estate market by utilizing data and computational tools according to M. Sharma et al., (2024).

Therefore, accurate property value is crucial for community development. The real estate market is one of the most dynamic and economically significant sectors in many countries. Housing prices are influenced by a wide range of factors, including location, property size, structural features, and prevailing market conditions. Accurate property valuation is therefore essential for effective decision-making by buyers, sellers, investors, and policy makers. Traditionally, property valuation has depended on manual appraisal methods, expert judgment, and basic comparative analysis. This project outlines using stack ensemble to predict housing prices based on key property features, which also incorporates essential data preparation steps such as handling missing values, feature scaling, and k-fold cross validation to ensure a robust and reliable model. It improved the accuracy and diversity of the model by learning how to optimally combine the base classifiers (Akinola et.al., 2026). This project describes how to use stack ensemble to forecast home values depending on important characteristics of a property. In order to provide a solid and trustworthy model, it also includes crucial data preparation procedures including

handling missing values, feature scaling, and k-fold cross validation. Due to the intricate interactions of variables such location, property size, structural characteristics, market demand, and economic conditions, accurately estimating the price of residential properties is a significant difficulty in the real estate sector.

The scope is limited to the variables available within the datasets, which may include features such as property size, number of rooms, location indicators, and other relevant attributes. Feature engineering was done which includes data cleaning, handling missing values, feature scaling, model Training, and performance evaluation using k-fold cross-Stack ensemble which include Random Forest, Gradient Boost, Linear Regression and Ridge Regression for the Meta model was used as the predictive technique, although other advanced models are acknowledged but not implemented. Despite the contributions and practical relevance of this study, several limitations must be acknowledged to provide a balanced Deng (2025) examined the role of location and multi-source image data in housing price analysis and found that geographic coordinates and spatially-varying image features contributed significantly to predictive performance, with effects differing markedly between city-centre and suburban areas. When comparing several machine learning methods for housing prediction, Khan, et al. (2026) found that a stacked ensemble combining CatBoost, XGBoost, and LightGBM with a ridge-regression meta-learner produced markedly higher accuracy than Linear Regression and other individual base models. In their evaluation of techniques and data types utilized in home price prediction systems, Geerts et al. (2023) emphasized the increasing significance of machine learning in contemporary real estate analytics.

Hu and Tang (2023) proposed a geographically weighted ensemble learning model for residential rent prediction and observed that incorporating spatial information significantly improved prediction performance across different urban locations. Chen et al. (2024) applied machine learning algorithms to rental valuation and found that property size, location, building characteristics, and neighborhood conditions were among the most influential determinants of rental prices.

Cao et al., (2024) investigated rental price prediction using machine learning techniques and reported that data-driven models provide reliable estimates of rental values when relevant housing characteristics are properly represented in the dataset. Oluyele et al., (2024) emphasized that housing attributes such as location, number of rooms, accessibility to infrastructure, and environmental quality have substantial effects on rental prices within urban housing markets. Trinkaus and Kauermann (2023) evaluated the effectiveness of machine learning algorithms for rental valuation and concluded that modern predictive models can outperform traditional approaches when sufficient housing data are available. Zong and Song (2023) developed a machine learning-based house rent prediction model and demonstrated that predictive

understanding of the findings and to clearly define the scope within which the results should be interpreted.

II. LITERATURE REVIEW

Several recent studies have demonstrated the effectiveness of machine learning techniques in housing price prediction. Choy and Ho (2023) applied Extra Trees, k-Nearest Neighbors, and Random Forest to real estate transaction data in Hong Kong and found that all three algorithms markedly outperformed a conventional hedonic pricing model in explanatory power and error minimization. Naz et al., (2024), in a systematic literature review of machine learning, deep learning, and hybrid approaches to real estate price prediction, emphasized that data quality and preprocessing remain central to predictive accuracy, alongside the growing use of hybrid models. Ogundunmade, et al. (2023) investigated rental housing determinants in Ibadan, Nigeria.

algorithms are capable of identifying complex relationships between housing features and rental values.

III. METHODOLOGY

A. System Architecture

The architecture of the system consists of Data input, Data Preprocessing Layer, Model Training Layer, Model Evaluation Layer and Price Prediction Layer components. It describes the sequential processes involved in transforming raw housing data into an accurate rent price prediction. The architecture consists of five major components, namely: Data Input, Data Preprocessing, Model Training, Model Evaluation, and Price Prediction Output.

B. Data Flow Diagram (DFD)

- 1) **DFD Level 0 (Context Diagram):** The Level 0 Data Flow Diagram as seen below in fig. 3.1, also known as the Context Diagram, provides a high-level overview of the house rent prediction system. It illustrates the interaction between the user and the system without showing the internal processing details. The Level 0 DFD therefore provides a simplified overview of how data move between the user and the prediction system.

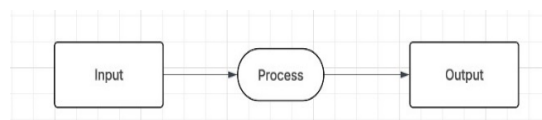


Fig. 3.1. DFD level 0

- 2) **DFD Level 1:** The Level 1 Data Flow Diagram in fig. 3.2 below provides a more detailed representation of the internal operations of the house rent prediction system. Unlike the Level 0 DFD, which treats the system as a single process, the Level 1 DFD breaks the system into multiple sub processes that describe how data are transformed step-by-step.

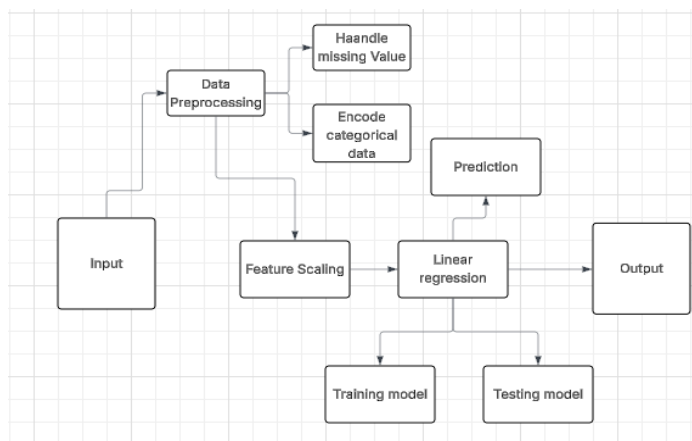


Fig. 3.2. DFD Level 1

C. Algorithm Design

Stack ensemble (using Gradient boost, Random Forest, Linear regression, and Ridge regression for the metal model) was used in the prediction Algorithm in fig. 3.3.

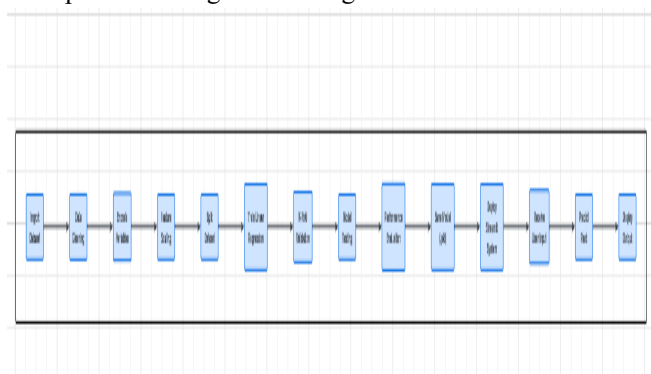


Fig. 3.3. Algorithm design

Importing the dataset, cleaning the data, handling missing values, encoding categorical variables, performing feature scaling, dividing the dataset into training and testing sets, training a linear regression model, performing K-fold cross validation, testing and evaluating the model's performance, saving model files (.pkl), deploying the streamlit application, accepting user input, preprocessing user input, predicting house rent, displaying the prediction output, and storing the prediction history

Lastly, the performance and efficacy of the constructed house rent prediction model were evaluated using evaluation metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2).

IV. RESULT DISCUSSION

This chapter presents the implementation and testing of the House Rent Prediction System developed for predicting annual house rent values using Stack Ensemble using Gradient Boost, Random Forest, Linear Regression and Ridge Regression for the metal model. The implementation stage transformed the methodology and design presented in Chapter Three into a functional predictive application capable of estimating house

rent prices using housing attributes. The developed system integrates machine learning techniques which is stack ensemble using gradient boost, linear regression, random forest and ridge regression for metal model with a user-friendly Streamlit interface to provide an interactive environment for rent estimation. The implementation process involved dataset preparation, preprocessing, model development, deployment, and interface design. The implementation also incorporated K-Fold Cross Validation to improve model reliability and prediction consistency.

The implementation of the House Rent Prediction System required suitable hardware like Computer and software resources like Jupyter, python and it libraries to support the operations of the machine learning model using stack ensemble, data preprocessing(which involves data cleaning, removing duplicates, checking for missing values), visualizations and deployment using streamlit for the interface. The dataset used in this project was obtained from two sources and the datasets as in Figure 4.1 was merged to improve coverage and diversity.

	A	B	C	D	E	F	G	H	I	J	K
1	Location	Property Type	Bedrooms	Bathrooms	House Size sqm	Lot Size sqm	Year Built	Furnishing	Water Supply	Power Supply	Annual Rent NGN
2	Agboko	Room	1	1	29	436	2019	Unfurnished	Borehole	Stable	209000
3	Apata	Mini Flat	1	2	59	935	1997	Fully Furnished	Borehole	Moderate	439000
4	Oluwole	Self Contained	1	1	31	849	2010	Unfurnished	Borehole	Poor	390000
5	Alimosho	Mini Flat	1	2	58	681	1997	Unfurnished	Public	Stable	750000

Fig. 4.1. Dataset used in training the model

The dataset was loaded into the jupyter notebook and the dataset was subjected to duplicate removal, checking for missing values, data transformation and scaling, the machine learning model implementation discusses the data encoding which label encoding was been used, followed the stages below. Feature scaling was also done using StandardScaler, model training using stack ensemble and K-fold validation was also used which was performed using 5 fold validation, label Encoding was applied to variables including location, property type, furnishing, water Supply, PHCN Availability.

Random Forest, Gradient boost, linear regression and ridge regression for the metal model was used for model training and can be viewed in was selected because house rent prediction involves continuous numerical values. Training operations was done by splitting the data into a ratio of 80% for training set and 20% testing set, feature extraction, coefficient learning, prediction fitting. Dataset partitioning are the trained model was exported as rent_model.pkl

The developed application was deployed using Streamlit and organized into multiple modules. The home page serves as the entry point of the application and shown below in Figure 4.2. The interface provides the application title which gives us information about the name of the application which is called House rent prediction system.

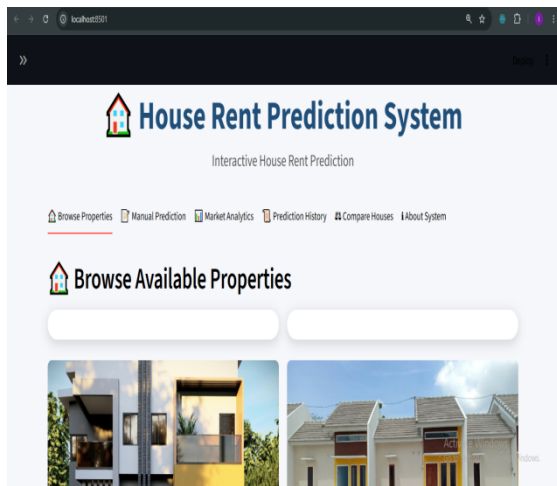


Figure 4.2 Home Interface

The prediction interface in fig 4.3. enables users to manually enter housing details and the prediction inputs.

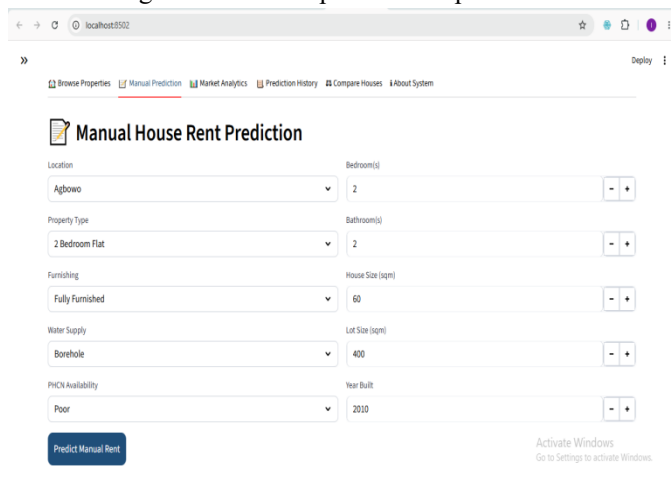


Fig. 4.3 Manual Prediction Interface

The developed system in Figure 4.4 includes a property browsing section. Properties are displayed using image cards. Each card contains house image, description and housing information. Users may select properties directly for prediction.

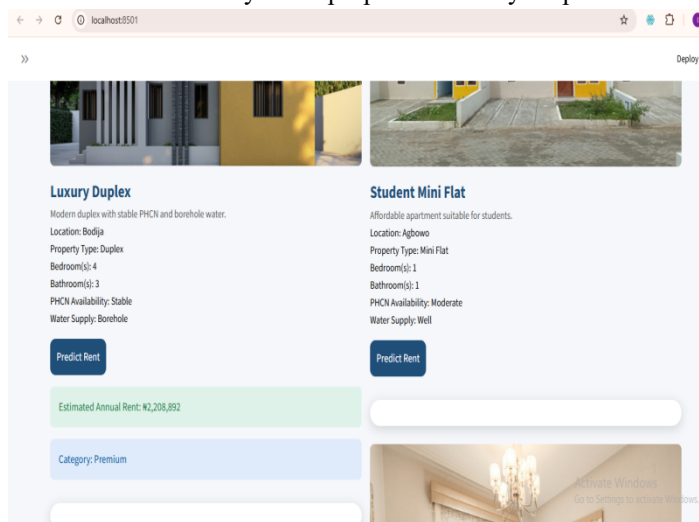


Figure 4.4 Property Gallery Interface

The house comparison interface in fig 4.5 allow users to compare two different houses, then click on the compare button after selecting the two houses to be compare and evaluate rent values. The comparison interface assists users in cost evaluation, property selection and housing decisions.

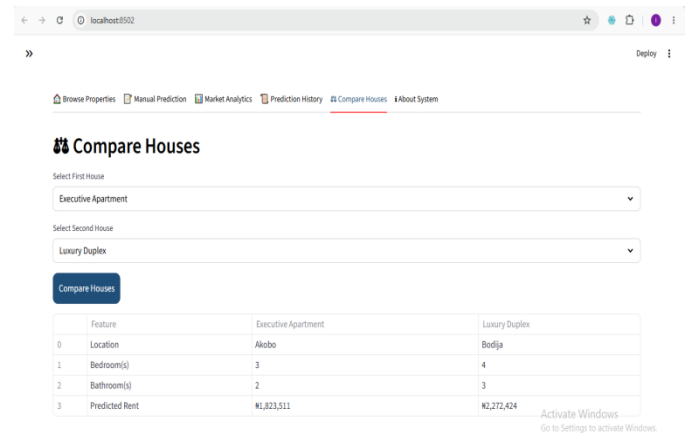


Fig. 4.5. House Comparison interface

Prediction history stores previously generated outputs of the prediction done anywhere in the interface meaning all the pages which are the browse house property, the manual prediction interface, the house comparison interface as shown in fig. 4.5 is also not left out. This module enables users to revisit earlier predictions done. Stored information in Figure 4.6 includes House details, prediction values, and time records.

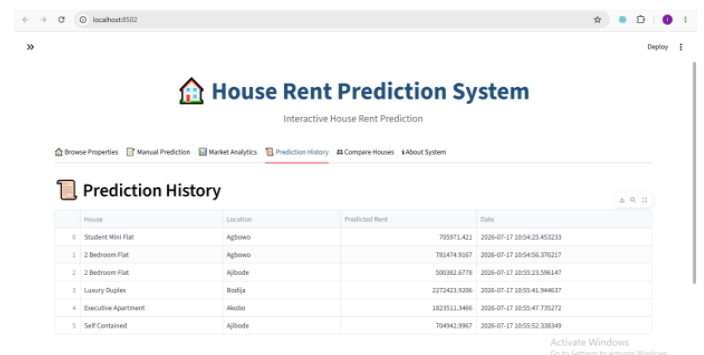


Fig. 4.6. Prediction History Interface

The analytics module provides visualization of housing information. Just as in figure 4.7, it displays information like average rent values, housing distribution and statistical summaries. This module improves market understanding.

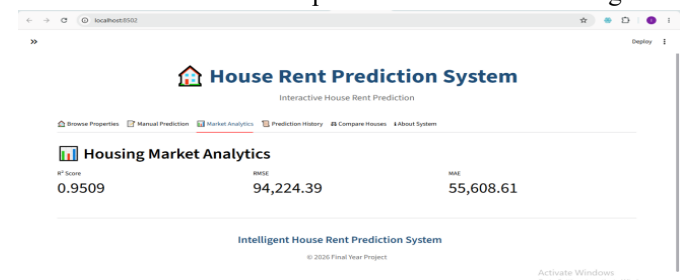


Fig. 4.7. Market Analytics Interface

“Enhancing House Rent Prediction Using A Stacking Ensemble Model for Sustainable Community Development”

The dark mode interface in Figure 4.8 allows users especially those who do not like the white mode to switch, mostly developers also like to use the dark mode and it applies to all tabs.

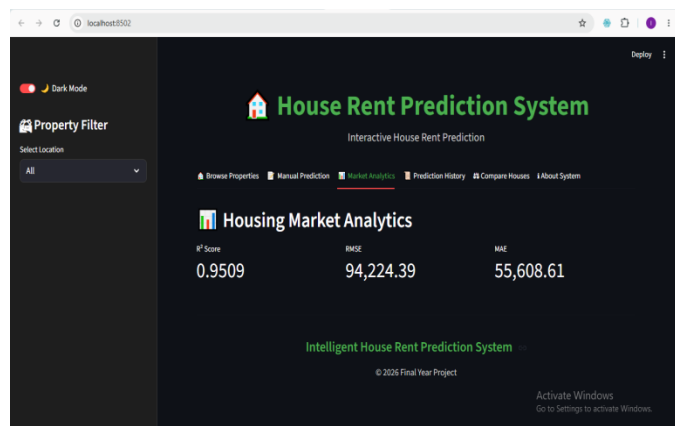


Fig. 4.8. Dark Mode Interface

The About System Interface as displayed in Figure 4.9 tells us about the system which is a machine learning-based model application developed for predicting house rent prices.

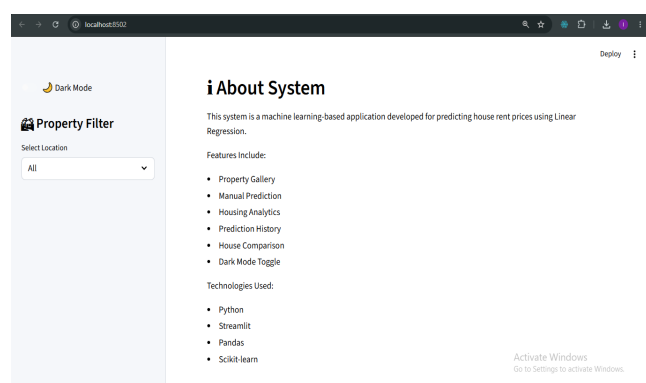


Fig. 4.9. About System

V. SUMMARY AND CONCLUSION

The increasing inconsistency in house rent pricing, especially within developing urban environments, has created challenges for house seekers, landlords, and real estate practitioners. Traditional methods of rent determination rely heavily on house agents, manual valuation, and subjective judgment, which often result in inaccurate and inconsistent pricing.

To address these challenges, this study developed a machine learning-based House Rent Prediction System using Linear Regression.

The study began with data acquisition from two major sources:

1. Housing information collected from house agents.
2. Secondary housing datasets obtained from news and social media.

The collected datasets were combined and prepared for model development through preprocessing operations such as removal of duplicate records, handling missing values, data transformation, One-Hot Encoding of categorical variables and Feature scaling using StandardScaler

The developed model utilized housing variables including location, property Type, bedroom(s), bathroom(s), House Size, Lot Size, Furnishing Status, Water Supply, PHCN Availability and Year Built. The stack ensemble algorithm was selected because it is suitable for predicting continuous numerical values such as house rent.

The developed system was deployed using Streamlit and implemented with several modules including manual prediction module, Property browsing module, House image display section, Property display, Prediction history module, House comparison module, Market analytics section and Dark mode customization. Model evaluation was carried out using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Coefficient of Determination (R^2). Additionally, K-Fold Cross Validation was incorporated to improve model reliability and minimize overfitting. Testing results showed that the developed application successfully generated house rent predictions and supported interactive housing exploration.

Stacking model and encoding method

The study employed a stacking ensemble consisting of Random Forest, Gradient Boosting, and Linear Regression as base learners, with Ridge Regression as the meta-model. Categorical variables were encoded using One-Hot Encoding, while numerical features were standardized using StandardScaler.

Model-performance results

The developed stacking ensemble achieved approximately 95% performance (R^2). Model evaluation was conducted using MAE, MSE, RMSE, R^2 , and five-fold cross-validation.

MAE: 55608.61170869439

MSE: 8878235002.393314

RMSE: 94224.3864527295

R^2 Score: 0.9509184040771506

Dataset sources and preprocessing

The dataset was obtained from Kaggle and supplemented with housing information from secondary sources. The data were merged and preprocessed through duplicate removal, missing-value treatment, One-Hot Encoding of categorical variables, and feature scaling using StandardScaler. An 80:20 training-testing split and five-fold cross-validation were applied.

Sustainability justification

The system supports sustainable community development by promoting transparent and data-driven rental valuation. It can assist tenants, landlords, and real estate practitioners in making informed housing and pricing decisions.

Property images and licensing

The property images used in the system were obtained from open-source platforms and were used for property visualization and improved user interaction. They were not used as predictive features.

Data privacy

The dataset was obtained from open-source/publicly available sources, including Kaggle and publicly available housing information. No private or confidential personal information was intentionally collected for model development.

Reproducibility

The system was implemented using Python, Jupyter Notebook, Visual Studio Code, Streamlit, Scikit-learn, Pandas, NumPy, Joblib, Matplotlib, and PIL. The study also reports the 80:20 train-test split and five-fold cross-validation, providing a basis for reproducing the experiment.

REFERENCES

1. Akinola S. O., Osunade O., Abiola O. A. and Adebisi R. O.(2026). Development of a Multimodal Artificial Intelligence Framework for Forest Fire Prediction University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR),Vol. 16 No. 1, pp. 202–212
2. Choy, L. H. T., & Ho, W. K. O. (2023). The use of machine learning in real estate research. *Land*, 12(4), 740. <https://doi.org/10.3390/land12040740>.
3. Deng, L. (2025). Real estate valuation with multi-source image fusion and enhanced machine learning pipeline. *PLOS ONE*, 20(5), e0321951. <https://doi.org/10.1371/journal.pone.0321951>.
4. Ogundunmade, T. P., Abidoye, M., & Olunfunbi, O. M. (2023). Modelling residential housing rent price using machine learning models. *Modern Economics & Management*, 2, 14. <https://doi.org/10.53964/mem.2023014>.
5. Zong, H., & Song, J. (2023). Research on the prediction model of house rent based on machine learning. . *Proceedings of the International Conference on Big Data Blockchain and Economy Management.*, <https://doi.org/10.4108/eai.19-5-2023.2334250>.
6. Trinkaus, O., & Kauermann, G. (2023). Can machine learning algorithms deliver superior models for rental guides?. *ASTA Wirtschafts-und Sozialstatistisches Archiv*, 17, 305–330.
7. Cao, S., Liao, W., & Huang, J. (2024). Research on renting price prediction based on machine learning. . *Proceedings of the 5th Management Science Informatization and Economic Innovation Development Conference.*, <https://doi.org/10.4108/eai.8-12-2023.2344718>.
8. Chen, W., Farag, S., Butt, U., & Al-Khateeb, H. . (2024). Leveraging machine learning for sophisticated rental value predictions: A case study from Munich, Germany. *Applied Sciences*, 14(20), 9528. <https://doi.org/10.3390/app14209528>.
9. Hu, G., & Tang, Y. (2023). GERPM: A geographically weighted stacking ensemble learning-based urban residential rents prediction model. *Mathematics*, 11(14), 3160. <https://doi.org/10.3390/math11143160>.
10. Oluyele, S., Akingbade, J., Akinode, V., & Idoghor, R. (2024). Prediction of urban house rental prices in Lagos, Nigeria: A machine learning approach. *ABUAD Journal of Engineering Research and Development*, 7(2), 216–228.
11. Khan, M. K., Khan, A. U., Ahmed, S., & Ali, W. (2026). A robust stacked ensemble approach for house price prediction. *Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences*, 63(2).
12. Naz, R., Jamil, B., & Ijaz, H. (2024). Machine learning, deep learning, and hybrid approaches in real estate price prediction: A comprehensive systematic literature review. *Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences*, 61(2), 129–144. [https://doi.org/10.53560/PPASA\(61-2\)863](https://doi.org/10.53560/PPASA(61-2)863).
13. Nor, M. I., Audu, B., & Mohamed, A. D. (2025). Investigating the impact of property characteristics, cost of living, and environmental factors on rental prices in Baidoa's climate-affected real estate market: A hybrid approach using hedonic regression and neural networks. *Frontiers in Sustainable Cities*, 7, Article 1511761. <https://doi.org/10.3389/frsc.2025.1511761>
14. Sharma, H., Harsora, H., & Ogunleye, B. (2024). An optimal house price prediction algorithm: XGBoost. . *Analytics*, 3(1), 30–45. <https://doi.org/10.3390/analytics3010003>.
15. Geerts, M., vanden Broucke, S., & De Weerd, J. (2023). A Survey of Methods and Input Data Types for House Price Prediction. *ISPRS International Journal of GeoInformation*, 12(5), 200. <https://doi.org/10.3390/ijgi12050200>.