JOURNAL OF COMPUTATIONAL INNOVATION AND ANALYTICS
https://e-journal.uum.edu.my/index.php/jcia

How to cite this article:
Obaude, A., Oyekanmi, E., Adegboye, A., & Abe, O. (2026). House price prediction in Ajah, Lagos using a random forest regression model. *Journal of Computational Innovation and Analytics*, *5*(2), 86-101.
https://doi.org/10.32890/jcia2026.5.2.5

House price prediction in Ajah, Lagos using a random forest regression model

[1]Ayodeji Obaude, [2]Ezekiel Oyekanmi, [3]Adegboyega Adegboye & [4]Oluwaseyi Abe
[1,2,3,4]Department of Computer Science, Achievers University
Owo, Ondo State, Nigeria

[1]Corresponding author: obaude-pg0011@achievers.edu.ng

Received: 12/03/2025 Revised: 01/02/2026 Accepted: 03/02/2026 Published: 31/07/2026

ABSTRACT

Housing is one of the most essential need to commoner but several factors contribute to acquiring a suitable home, factor such as low income, high material cost and most importantly the presence of middleman in acquiring suitable house posed significant challenges thereby increases the cost, this study addresses this problem by developing a random forest regressor model to predict house price in Ajah Lagos, a dataset of 24,326 instances was sourced from Kaggle and filtered to 2142 instance containing Ajah listing. The data underwent several preprocessing steps; outliers were detected using the interquartile range, and Recursive Feature Elimination was used to identify relevant features that contribute to the model while eliminating less important ones. The RF model demonstrates a robust performance with an R-squared of 68.5%, a Mean Absolute Percentage Error of 18.96%, and a Root Mean Squared Error of ₦11.1 million. Compared with existing models in the literature, the model achieved a mean absolute error of ₦8.09 million, representing a 4.82% improvement and outperforming the ridge regression models. This research contributes to the field by providing stakeholders with a data-driven tool to enhance transparency and investment decision-making in the Nigerian real estate market.

Keywords: House price, machine learning, predictive modeling, random forest, real estate

INTRODUCTION

In the complex landscape of Nigerian urban development, housing serves not only as shelter but also as a vital indicator of the standard of living and national wealth (Henilane, 2016; Guyer, 2015). However, the Nigerian real estate market is characterized by information asymmetry and a reliance on conventional, often subjective, valuation methods (Oladokun & Mooya, 2024). Historically, the transformation of housing into a financial asset was accelerated by the economic restructuring of the 1980s, which eased access to credit and turned real estate into a primary investment vehicle (Stirling et al., 2022).

Despite this importance, the market in cities like Lagos suffers from middleman inflation and a lack of standardized pricing models. Traditional valuation often relies on manual assessments and historical comparisons that fail to capture the dynamic influence of specific features like parking space or house type (LandMall.ng, 2025). The emergence of Artificial Intelligence (AI) and Machine Learning (ML) offers a transformative solution, capable of processing vast datasets to identify complex patterns and provide transparent price forecasts (Aanda Consulting, 2025). This study situates itself within this technological shift, focusing on Ajah, Lagos, to provide a data-driven framework to mitigate current inefficiencies in the local housing sector.

Research Problem

The housing sector in Nigeria suffers from a lack of standardized valuation frameworks, leading to significant price imbalances and unpredictable market behavior that dissatisfy both buyers and sellers. The lack of transparency often leads to unrealistic overpricing, as sellers ignore market indicators and buyers' purchasing power, triggering prolonged market stagnation. The pricing inefficiency not only obstructs potential homebuyers but also results in financial losses for sellers through asset depreciation and mounting renovation costs associated with unsold, deteriorating properties.

To address this systemic inefficiency and restore market confidence, this research develops a random forest regressor and evaluates its accuracy in predicting house purchase prices, with the following specific objectives.

1. Develop a machine learning model using a random forest regressor capable of accurately predicting house sales prices based on a comprehensive set of features such as the number of bedrooms, toilets, bathrooms, and parking spaces, and the type of house, to provide valuable insights for homebuyers, real estate professionals, and market analysts.
2. Evaluate the performance of the machine learning model using metrics such as R-squared, root mean squared error, mean squared error, mean absolute percentage error, and mean absolute error for predicting house sale prices.
3. Compare the model performance with existing models in the literature.

This study is important to various stakeholders, including homebuyers, real estate agents, policymakers, market analysts, and researchers. By evaluating the accuracy of the machine learning model in predicting house purchase prices, this research provides valuable insights to inform data-driven decision-making and improve the efficiency and transparency of the housing market.

LITERATURE REVIEW

Introduction to Housing and Its Importance

Housing is one of the most prominent aspects of life, providing shelter, safety, and warmth, as well as a place of rest (Henilane, 2016). Moreover, this aligns with Ernawati et al. (2016), who argue that housing is important for the sustainable development of every country and is a basic need that improves individuals' well-being. According to Henilane (2016), housing is not just a basic need but also a measure of the standard of living. It was emphasized that housing must be economically affordable, meaning that it should be within the financial reach of individuals and families. It should also be sustainable; this implies that the cost of maintaining the property is low. Housing should be environmentally friendly, minimizing negative impacts on the environment. According to Guyer (2015), housing is an important asset for households and investors, implying that housing can generate half of national wealth and can vary considerably over time. According to a study by Jiboye (2014), various types of housing are occupied by residents in the Oshogbo community. The study found that 80% of residents lived in contemporary vernacular houses known as "face-me-I-face-you," while 14.8% resided in western-style apartment houses. Additionally, 5.2% of respondents lived in duplexes, single-family homes, and traditional courtyard dwellings. This distribution underscores how affordability and availability shape housing choices in the area, as the vast majority of residents live in low-cost neighborhoods.

According to Stirling et al. (2022), the transformation of houses into financial assets can be traced back to the economic restructuring of the 1980s, particularly the removal of regulations that had restricted borrowers' access to bank loans. This transformation enabled easier access to credit by lowering loan interest rates. During this period, changes such as the removal of regulations for acquiring loans and the introduction of less rigid credit policies, such as lower interest rates, enabled financial institutions to offer more loans and mortgages to borrowers. This transformation facilitates easier access to homeownership by enabling them to acquire more capital, turning real estate into a significant financial asset and investment opportunity. According to a study conducted by Başara and Sisman (2022), housing is important for meeting people's demand for shelter. It reveals that the house plays more roles than just meeting people's needs; housing can also be viewed as a means of acquiring influential status and as a publicly acceptable investment.

Challenges in Housing Availability in Developing Countries

According to Obi et al. (2014), the average Nigerian has difficulty obtaining affordable housing. This has affected the socioeconomic well-being of the country, like other developing nations. Several factors contribute to the difficulty of acquiring affordable housing in Nigeria, including a rapid population growth rate, urbanization, high costs of building materials, the absence of indigenous technology and skilled personnel, an insufficient financial structure, and poor managerial skills among Nigerian mortgage providers. Norhasliya et al. (2018) identified the housing affordability problems faced by the government. Some of the challenges include population growth, with low- and middle-income earners being the major groups affected.

According to Awe et al. (2023), the study reveals that everyone desires a good-quality house. However, the quality of a house depends on individual earnings, ranging from land purchase costs to construction costs. Middle- and low-income earners struggle to acquire good-quality houses due to high property costs, which force them to settle in areas with intolerable transportation costs. Bons et al. (2019). These challenges in housing availability led to diverse classifications of houses, as people with lower incomes acquire houses that reflect their incomes, similar to those of the middle class. Housing prices determine the type of house everyone can afford, as their financial capacity limits their options. This results in a range of housing types and market dynamics as people seek solutions to their housing needs. In the next section, the factors that determine house prices will be identified, providing a comprehensive understanding of how they shape the housing market.

Determinants of House Prices: A Global Perspective

According to a study by Ezgi et al. (2015), when identifying factors affecting house prices in Turkey, it was noted that the availability of heating systems in a house has a direct correlation with the price; the occurrence of earthquakes negatively impacts house prices; an increase in land value increases house prices; and the rental value of a house also drives up its price. Additionally, the existing floor of a house is inversely related to property prices. As the floor's longevity increases, house quality tends to decline, which negatively affects the house's overall value.

Ernawati et al. (2016) identified several factors affecting house prices, with location as the major determinant. Houses located close to institutions such as schools, churches, and other facilities, as well as other convenient locations that provide residents with opportunities and a preferred way of life, are positively linearly correlated with house prices. The study also reveals that macroeconomic variables such as gross domestic product, population growth, and real property gain tax play significant roles in determining house prices. It was also revealed that an increase in housing demand raises prices, and industry factors, such as increases in raw material costs, can also raise house prices. Anewuoh et al. (2023) identify several factors that influence house prices in Ghana, including construction costs, financial markets, structural features, government policies, macroeconomic determinants, and population changes. In support of this, Anthony et al. (2020) highlighted that macroeconomic variables such as Gross Domestic Product (GDP), population growth, unemployment rate, and interest rates determine house price movements in Ghana. The study emphasizes that an increase in interest rates, the cost of borrowing, will reduce demand for houses and potentially lower their prices. Manganelli et al. (2015) asserted that higher employment rates increase people's purchasing power, thereby raising demand and leading to higher house prices. The study revealed that house prices have a positive linear relationship with GDP, the employment rate, the interest rate, the inflation rate, and the average wage rate. Zeni et al. (2020) identified land area, number of bedrooms, and materials as key determinants of house prices. House prices can also be influenced by regional conditions such as infrastructure, population growth, and much more.

Truong (2020) noted that the house price index (HPI) is affected by factors such as location, area, and population. Since housing prices are strongly correlated with these factors, they are used to estimate changes in house prices in many countries. This complex interplay of factors makes predicting house prices challenging. However, machine learning can offer sophisticated tools and techniques to analyze these determinants effectively and provide more accurate predictions. By understanding the challenges in housing availability and the determinants of house prices, we can better predict house prices, particularly in regions like Ajah, Lagos, Nigeria. This understanding provides the foundation for applying predictive modeling techniques to forecast housing market trends in such areas. It also leads into the next section, where we explore the fundamentals of machine learning.

Leveraging A Machine Learning Model in The Housing Market

Machine learning is a general term for a wide variety of algorithms that use computational and statistical models to make intelligent predictions from data. These algorithms expand on the traditional statistical modeling capacity by increasing the number of datasets used and improving computational power. The decision to select a suitable machine learning algorithm for a problem depends on the characteristics of the data and the outcome. For smaller datasets, a classical approach such as a statistical linear regression model or a decision tree, which splits data based on certain conditions or rules, might be suitable (Nichols et al., 2018). The application of machine learning cuts across several industries. Machine learning will not only simplify the process of accomplishing a predefined task but also provide users with the optimal benefit of data-driven decision-making.

The housing market has benefited from this pool of tools. A study by Liao (2024) employed a machine learning model to predict house prices in Seattle. Another study was conducted by Nwankwo et al. (2023). The study used ridge regression to develop a machine learning model to predict house prices in Ajah, Lagos, Nigeria. In their study, the mean absolute error was 8.5 million.

Classification of Machine Learning Model

Machine learning models can be classified into four categories**:** supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning.

Supervised Learning

In supervised learning, this type of machine learning model maps input to output using training data containing input-output pairs. Supervised machine learning models can be used to solve classification and regression problems. Examples of this type of machine learning include the Random Forest Regressor.

Random Forest Regressor

The random forest regressor is an ensemble machine learning model that combines several decision trees into a single model. This is done by randomly selecting a subset of the data for each decision tree and identifying the optimal information gain for each tree, thereby making a prediction. The predictions of several trees are aggregated, and their average is used in a regression problem. This process is called bagging.

Gap to Fill

While there are several global studies on the application of machine learning models in house price prediction, there seems to be minimal research in the Nigerian housing market. Understanding how this model works in the Nigerian context is crucial for our research.

A recent study in Nigeria, conducted by Nwankwo (2023) on the application of a machine learning model for house price prediction in Lagos, uses ridge regression, which is suitable for handling multicollinearity and modeling linear relationships in the data. Also, the study limits its evaluation metric to mean absolute error (MAE). Our study aimed to address this limitation by employing a random forest regressor, which is well-suited to modeling both linear and nonlinear relationships, enabling the model to capture complex relationships in the data. Moreover, our study aimed to expand the evaluation metric by incorporating MAPE, MSE, RMSE, and R-squared, in addition to MAE, to assess model performance.

RESEARCH METHODOLOGY

This research study uses machine learning models to derive insights from housing data in Nigeria, employing a random forest regression algorithm. The study aimed to compare the accuracy of machine learning models with existing models that best predict the dependent variable, house price.

Data Collection

The dataset used for this study will be obtained from Kaggle, a public repository capturing features related to house sales prices in Nigeria. The dataset includes features such as the number of bedrooms, bathrooms, and toilets, parking space, house price, and property location. The target variable of interest is the house price.

Data Preprocessing

Prior to the development of the model, the data will be preprocessed to ensure data integrity, and this step involves

Data Cleaning

Several data cleaning procedures will be undertaken to ensure data cleanliness, such as fixing missing and duplicate values, removing whitespace from text, fixing outliers, and addressing other data inconsistencies.

Feature selection

Further preprocessing steps would be performed to enhance data quality and improve model performance. Recursive Feature Elimination (RFE), in conjunction with the Random Forest Regressor, will be used iteratively to select the most relevant features in the dataset that predict the dependent variables. The RFE process will select the top features, as specified by the feature parameter, that exhibit the strongest relationship with the target variable, the house price. After selecting the number of features iteratively and evaluating the performance, the result will be recorded for comparative analysis to select the features that are the best predictors of the dependent variables,

Data Exploration

To understand the dataset's characteristics, insights will be gained through techniques such as correlation, logical, and descriptive analyses. Various plots will be created, such as a scatterplot to identify relationships between features and a boxplot to show the distribution of house prices across the various house types in the data.

Model Training and Evaluation

Data Splitting

The data will be split into training and testing sets, with 80% used for training and 20% reserved for testing to validate the model's performance on unseen data.

Model Training

Two machine learning models using random forest regression will be developed: the baseline model and the comprehensive model, following thorough further preprocessing using the recursive feature elimination method.

Model Representation

A random forest regressor is an ensemble machine learning method that combines the predictions of several decision tree regressors by averaging them to achieve more robust predictions. In this study, the final prediction Yi for the house, as in Equation 1, is the average of the decision tree predictions, weighted by the number of estimators in the random forest regressor.

$$y_i = \left(\frac{1}{N}\right)\sum_{i=1}^{N} F_i(x) \tag{1}$$

Where y_i is the average house price prediction from the total decision tree in the forest, N is the total number of decision trees in the random forest regressor, and $F_i(x)$ is the prediction of the decision tree

$\sum_{i=1}^{N}$ which is the total sum of all N decision trees in the random forest.

Model Evaluation

The models will be evaluated using various metrics to assess their performance in predicting house sales prices. The evaluation metrics are listed as follows.

R-squared (R^2) measures the proportion of variance in the target variable explained by the model, as given in Equation 2.

$$R2 = 1 - \frac{\sum_{i=1}^{n}(yi - \hat{y})}{\sum_{i=1}^{n}(yi - \bar{y})} \tag{2}$$

where y_i is the actual value, $\hat{y}$ is the predicted value, $\bar{y}$ is the mean of the actual value, and n represents the number of data points.

This study also evaluated Mean Squared Error (MSE), which measures the average squared difference between the model's predicted values and the actual values, given in Equation 3.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}) \tag{3}$$

where y_i is the actual value, $\hat{y}$ is the predicted value, and n represents the number of data points.

Root Mean Squared Error (RMSE), which measures the average difference between the model's predicted value and the actual result, was also assessed using Equation 4. RMSE is the square root of MSE and provides a more interpretable measure of prediction error.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y})} = \sqrt{MSE} \tag{4}$$

where y_i is the actual value, $\hat{y}$ is the predicted value, and n is the number of data points.

In addition to those metrics, this study computes the Mean Absolute Error (MAE), which measures the average absolute difference between the model's predicted values and the actual values, as given in Equation 5.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|yi - \hat{y}| \tag{5}$$

where y_i is the actual value, $\hat{y}$ is the predicted value, and n is the number of data points.

Mean Absolute Percentage Error (MAPE) was also employed for model evaluation, which calculates the average percentage difference between the model's predicted values and the actual values, as given by Equation 6.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{yi-\hat{y}}{yi}\right| \tag{6}$$

where y_i is the actual value, $\hat{y}$ is the predicted value, and n is the number of data points.

After evaluating model performance, the results were recorded for comprehensive comparison, and the optimal model with the best performance was selected for deployment.

RESULTS AND DISCUSSION

Data Collection and Description

The dataset used for this study was obtained from Kaggle. The dataset comprises features related to house prices in Nigeria, namely bedroom, bathroom, parking spaces, toilet, town, state, house type, and house price, which is the target variable.

Data preprocessing

The dataset contains no missing values. Inconsistency was identified in the "state" column, where "Anambra" was misspelled as "Anambara". Upon further scrutiny, it was observed that towns associated with the misspelled "Anambara" did not correspond to any actual towns in Anambra State. To maintain data accuracy and integrity, all entries associated with 'Anambara' were removed, reducing the dataset from 24,326 to 24,181 instances. The dataset was filtered to include only houses in Ajah, Lagos, Nigeria, our area of study. The filtered dataset comprises 2142 housing instances in Ajah, Lagos, Nigeria.

Exploratory Data Analysis

The findings reveal the average prices for each house type in Ajah, Lagos, with the Semi-Detached Duplex having the highest average price at ₦230,453,586. The Detached Duplex is the second-highest, with an average price of ₦118,674,928, as illustrated in Figure 1. This shows that duplexes have the highest property sale prices among house types, at less than one-third of the Semi-Detached Duplex's price. All listed properties with the lowest average prices are shown in Table 1.

Figure 1

Average price of houses in Ajah, Lagos by House Type

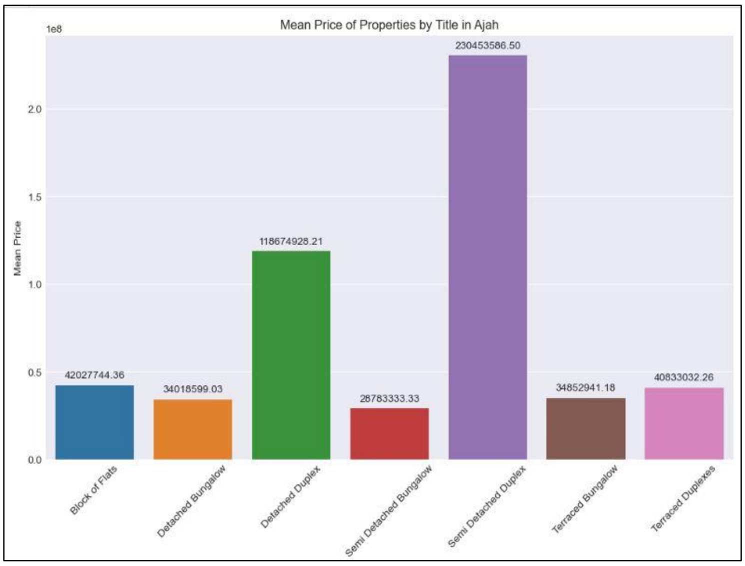


Table 1

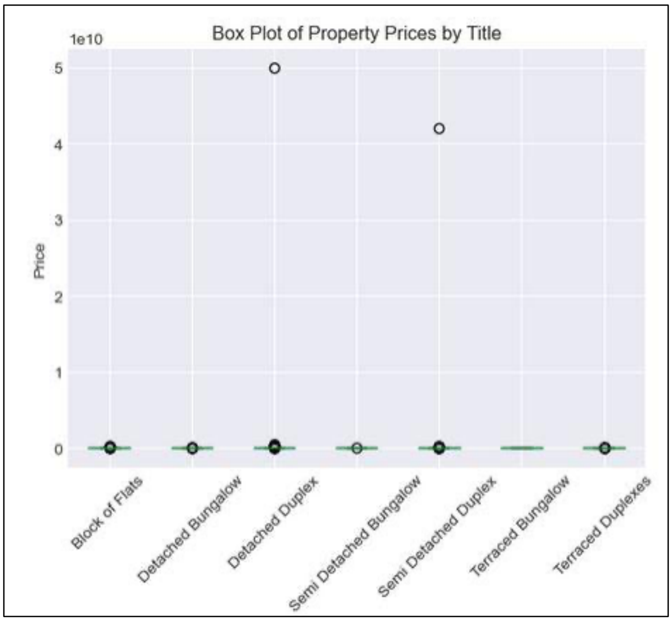
Baseline Model Performance

No	Type of properties	Average price
1	Detached Bungalow	₦34,018,599
2	Semi-Detached Bungalow	₦28,783,333
3	Terraced Bungalow	₦34,852,941
4	Block of Flats	₦42,027,744
5	Terraced Duplex	₦40,833,032

The box plots appear to be compressed for each house type, as shown in Figure 2, due to excessive outliers in the data. During data exploration, outliers were detected in the house price variable across various house types. The outlier indicates the presence of expensive houses within each house type, but it could skew the machine learning model.

Figure 2

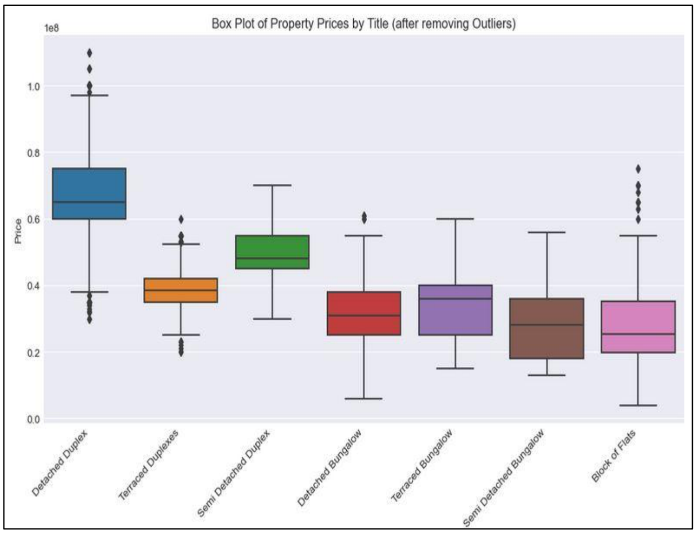
Outlier Detection Across All House Types in Ajah, Lagos



An outlier is a data point that falls outside the box-whisker. To address this, the outlier was removed from each house type's price by clipping data points that fall outside the whiskers, which represent the maximum price we can tolerate. The whiskers were calculated using the formula (1.5 * interquartile range) + Q3 for the upper whisker and (1.5 * interquartile range) – Q1 for the lower whisker. This ensures a more representative distribution of house prices across various house types and reduces the impact of extreme values. Despite the removal, outliers can still be identified in the new dataset, as shown in Figure 3.

Figure 3

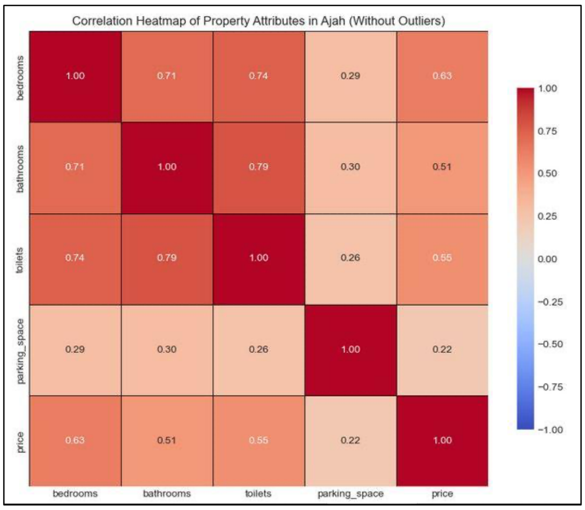
Outlier Removed Across All House Types in Ajah, Lagos.



Our findings show a strong positive linear relationship (0.71) between bedrooms and bathrooms, as indicated in Figure 4. This implies that an increase in bedrooms results in an increase in bathrooms. Additionally, the number of toilets is strongly positively correlated with both bedrooms (0.74) and bathrooms (0.79). This means that as the number of toilets increases, the number of both bedrooms and bathrooms also increases. There is a slight positive linear relationship between toilets and parking spaces (0.26). Furthermore, the price has a moderate positive correlation with bedrooms (0.63), bathrooms (0.51), and toilets (0.55), indicating that an increase in these features is associated with higher prices, as shown in Figure 4. It is important to note that correlation does not imply causality.

Figure 4

Correlation Analysis Among Variables in the Ajah Dataset



Label Encoding and Dropping Redundant Columns

In order for the data to be suitable for modeling, further data preprocessing was done, which included encoding the categorical variables into numerical data and also dropping redundant columns such as town and state, since we are working with a dataset from the same town and state, which might not provide significant benefit to the model. Categorical variables in the dataset, including house type, town, and state, were encoded as numerical values using a *LabelEncoder* from scikit-learn.

Model Training

Before training the model, the dataset was split into training and testing sets using scikit-learn's *train_test_split.* The split data contains 80% of the dataset, which was used to train the model, while the remaining 20% was reserved for testing and evaluating its performance on unseen data. Before the comprehensive model incorporating recursive feature elimination was developed, a baseline model was built using the Random Forest algorithm. This baseline model serves as a reference point for a comprehensive model to compare with and outperform. The training data was used to fit a random forest algorithm to develop the baseline model. Following the development of the baseline model, the comprehensive model was built iteratively using the recursive feature elimination method and the random forest regression algorithm.

Model Evaluation

Baseline Model Performance

The baseline model was evaluated using several metrics, including root mean squared error (RMSE), mean squared error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), as summarized in Table 2. The performance of the model is as follows: the RMSE is ₦11,141,959, the MAE is ₦8,282,709, and the MAPE is 19.29%, indicating that the predicted result, on average, is 19.29% away from the actual result. The explained variance is 67%, indicating that the independent variables account for 67% of the variance in the dependent variable, which in this case is the house price.

Table 2

Baseline Model Performance

Metric	Value
MSE	124,143 trillion
RMSE	11,141,959
MAE	8,282,709.18626
MAPE	19.28550
R2	0.67251

Comprehensive Model Performance

To improve the baseline model's performance, we performed additional data preprocessing. This includes feature selection with Recursive Feature Elimination (RFE), which identifies the top five features: bedrooms, bathrooms, toilets, parking spaces, and house type out of six features.

The comprehensive model in terms of performance achieves a lower MAE of ₦8,088,956, a lower MAPE of 18.96%, and a higher R^2 of 0.685 (see Table 3). This implies that the independent variables explained approximately 68.5% of the variance in the dependent variable. The comprehensive model outperforms the baseline model by achieving a better result. This is due to the impact of the recursive feature elimination method.

Table 3

Table Showing Comprehensive Model Performance.

Metric	Value
MSE	119,415 million
RMSE	10,927,729
MAE	8,088,956
MAPE	18.96%
R2	0.684983477

Feature Importance Analysis

To understand how the model made its decisions, we conducted a feature importance analysis, where the results are shown in Table 4. Unlike the linear regression model, which relies on a simple equation comprising components such as coefficients and intercepts, random forest takes a different approach: it assigns a score to each feature based on how much it reduces the impurity across all decision trees.

Table 4

Feature Importance analysis

Feature	Importance Score	Contribution (%)
Bedrooms	0.3909	39.09%
House Type	0.2985	29.85%
Toilets	0.1388	13.88%
Bathrooms	0.1334	13.34%
Parking Space	0.0384	3.84%
Total	1	100.00%

Comparison with the Existing Models

As indicated by Table 5, our model achieves a lower mean absolute error (MAE) of ₦8.09 million, compared to ₦8.5 million reported in existing studies, representing a 4.82% improvement. Unlike existing studies, our model incorporates explained variance (R-squared), which stands at 68.5%. This metric helps analysts understand how well the independent variables explain the variance in house prices, providing a solid foundation for further research. Additionally, the mean absolute percentage error (MAPE) of 18.96% highlights the extent to which the predicted results deviate from the actual values. The superior performance of our model can be attributed to the Recursive Feature Elimination (RFE) technique, which identified the most relevant features: bedrooms, bathrooms, toilets, parking spaces, and house type for predicting house prices in Ajah, Lagos.

Table 5

Comparison With the Existing Model

Evaluation Metric	Our Model	(Nwankwo,2023)	Improvement
Mean Absolute Error	#8.09million	#8.5million	4.82%
R-squared	68.5%	Nill	68.5%
Mean Absolute Percentage Error	18.96%	Nill	18.96%

CONCLUSION

The study proves that a machine learning model can be a valuable tool for simplifying the determination of house sale prices. This will potentially empower various stakeholders, such as home sellers, home buyers, and real estate agents, with the right tool to aid their decision-making, increase trust, and improve confidence in exploring the housing market. Additionally, the study also shows that features such as the bedroom, bathroom toilet and parking space are positively correlated with price of properties in Ajah, Lagos, Nigeria and this study proves that the random forest regressor, also referred to as one of the bagging algorithms, is a powerful algorithm that best predicts house price prediction, and this was proven by the model's ability to surpass the existing model developed with the ridge regression by 4.82%.

RECOMMENDATION

Real estate agents should adopt the random forest regression model as a primary valuation tool because it provides objective and data-driven property price estimates. This approach can improve valuation accuracy while increasing transparency and client trust. For home buyers and sellers, the model offers a practical means of supporting informed decision-making. Buyers may use its predictions to detect inflated prices introduced by intermediaries, whereas sellers may use it as a benchmark to determine competitive prices based on measurable property attributes, such as room count, rather than relying on intuition. Furthermore, policymakers and government authorities should prioritize developing a centralized digital data warehouse for property transactions. Such an initiative would provide reliable, high-quality data for researchers and analysts, improve the predictive accuracy of valuation models, and promote greater transparency in the property market.

LIMITATION OF THE STUDY

While the model explains 68.5% of the variance, its explanatory power remains incomplete, suggesting that performance in real-world deployment may be constrained. The study is further limited by the exclusion of potentially relevant exogenous variables, including proximity to key urban amenities and infrastructure such as schools, seaports, and neighborhood facilities. In addition, the analysis is restricted to random forest regression as an ensemble machine learning method. It does not include a comparative evaluation against other emerging machine learning or deep learning approaches, such as neural network-based models.

FURTHER RESEARCH

While the machine learning model has a high predictive power of 68.5%, additional data can be collected to improve its performance and increase its accuracy in predicting house prices in Ajah, Lagos. Data such as locations, amenities, and building materials could improve the model's performance. The superior performance of the random forest regressor model, outperforming the ridge regressor model, indicates that each algorithm has its weaknesses and strengths; more machine learning algorithms, such as the XGBoost algorithm, the CatBoost algorithm, and deep learning algorithms such as the recurrent neural network, can be explored to develop machine learning models for predicting house prices in Ajah. More datasets from other regions can be collected to develop a sophisticated machine learning model for house price prediction. This will enable the model to generalize to other regions.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to the Department of Computer Science at Achievers University for their academic support, and we would also appreciate the editor and anonymous reviewers of JCIA for their constructive feedback and insightful suggestions, which have significantly improved the quality of this manuscript. This research receives no funding from any public, commercial, or non-profit agency. The authors declared no potential conflicts of interest regarding financial, authorship, or publication that may influence the work reported in this project.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

DECLARATION OF GENERATIVE AI USE

Generative AI tools were not used for the analysis or in the preparation of this manuscript.

DATA AVAILABILITY

All data are available under the following repository:
https://www.kaggle.com/datasets/michaelanietie/nigerian-house-price-dataset

REFERENCES

Aanda Consulting. (2025, March 3). *Real estate in Lagos: AI adoption – Revolutionizing the property market*. Aanda. https://www.aanda.com.ng/ai-adoption-in-lagos-real-estate-revolutionizing-the-property-market/

Anewuoh, F., Gyadu-Asiedu, W., & Appiadu-Boakye, K. (2023). The real estate market in Ghana: The driving factors and housing prices. *International Journal of Technology and Management Research, 8*(1), 36–55.

Awe, F. C., Adeboye, A. B., Akinluyi, M. L., Okeke, F. O., Yakubu, S. U., & Awe, F. D. (2023). Assessment of the relationship between housing quality and income in the urbanizing city of Ado-Ekiti, Nigeria. *World Journal of Advanced Research and Reviews, 18*(2), 969–978.

Başara, A. C., & Sisman, Y. (2022). Importance analysis of criteria affecting house selection using the analytic hierarchy process. *Advanced Land Management, 2*(1), 29–39.

Bons, O., Onochie, A., & Nzewi, N. (2019). Where is home for the urban poor in Abuja, Nigeria? *International Journal of Trend in Research and Development, 3*(1), 45–56.

Candas, E., Kalkan, S. B., & Yomralioglu, T. (2015, May 17–21). Determining the factors affecting housing prices [Paper presentation]. FIG Working Week 2015, Sofia, Bulgaria.

Guyer, J. I. (2015). Housing as "capital." HAU: *Journal of Ethnographic Theory, 5*(1), 24–28.

Henilane, I. (2016). Housing concept and analysis of housing classification. *Baltic Journal of Real Estate Economics and Construction Management, 4*(1), 168–179.

Jiboye, A. D. (2014). Significance of house-type as a determinant of residential quality in Osogbo, Southwest Nigeria. *Frontiers of Architectural Research, 3*(1), 20–27.

Kamal, E. M., Hassan, H., & Osmadi, A. (2016). Factors influencing the housing price: Developers' perspective. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, 10(*5), 1676–1682.

LandMall Technology. (2025). *The rise of AI-powered "land value forecasting" in real estate*. LandMall.ng. https://www.landmall.ng/blog/rise-ai-powered-land-value-forecasting-real-estate

Liao, J. (2024). House price prediction using machine learning: A case study in Seattle, U.S. *Applied and Computational Engineering, 32*(1), 259–269.

Manganelli, B., & Tajani, F. (2015). Macroeconomic variables and real estate in Italy and the USA. *Scienze Regionali, 14(*2), 25–46.

Nichols, J. A., Chan, H. W. H., & Baker, M. A. B. (2019). Machine learning: Applications of artificial intelligence to imaging and diagnosis. *Biophysical Reviews, 11*(1), 111–118.

Nwankwo, M. P., Onyeizu, M., Asogwa, E. C., & Chukwuogo, O. E. (2023). Prediction of house prices in Lagos, Nigeria, using machine learning models. *European Journal of Theoretical and Applied Sciences, 1*(5), 313–326.

Oladokun, S. O., & Mooya, M. M. (2024). Another look at data challenges in property valuation practice: A case of the Lagos property market. *Journal of Property Investment & Finance*, *42*(4), 325–347.

Oloke, O., Olawale, Y. A., & Oni, A. S. (2017). Price determination for residential properties in Lagos State, Nigeria: The principal-agent dilemma. *International Journal of Property Studies, 6*(3), 28–32.

Owusu-Ansah, A., Anim-Odame, W. K., & Azasu, S. (2021). Examination of the dynamics of house prices in urban Ghana. *African Geographical Review, 40*(1), 76–91.

Stirling, P., Gallent, N., & Purves, A. (2022). The assetisation of housing: A macroeconomic resource. *European Urban and Regional Studies, 30*(1), 3–20.

Truong, Q., Nguyen, M., Dang, H., & Mei, B. (2020). Housing price prediction via improved machine learning techniques. *Procedia Computer Science, 174*, 433–442.

Wickramaarachchi, N. C. (2016). Determinants of rental value for residential properties: A landowner's perspective on boarding homes. *Built-Environment Sri Lanka, 12*(1), 10–24.