International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 • Website: www.ijfmr.com • Email: editor@ijfmr.com

A Machine Learning Framework for Land Price Estimation in Huye District

Ntambara Etienne[1], Nkurunziza Egide[2], Niyirora Didace[3], Akimana Aline[4], Hagenimana Jean Bosco[5], Ndayizeye Jean Baptiste[6], Nsengiyumva Elyse[7], Nzayisenga Cyprien[8]

[1,3]Assistant Lecturer, Department of ICT, Rwanda Polytechnic – Huye College
[2]Lecturer, Department of ICT, Rwanda Polytechnic – Huye College
[4,5,6,7,8]Bachelor of Technology Student, Department of ICT, Polytechnic – Huye College

Abstract
Accurate land valuation is essential for effective land governance, taxation, urban planning, and investment decision-making, particularly in rapidly urbanizing developing countries such as Rwanda. However, existing land valuation practices in Rwanda remain largely manual, subjective, and inconsistent, limiting efficiency, scalability, and transparency. This study presents a machine learning-based framework for land price estimation in Huye District.

A multi-source dataset integrating spatial and administrative land attributes was constructed from land registry databases, official valuation schedules, Geographic Information Systems (GIS), and administrative land records. The initial dataset contained 51,955 records, which were reduced to 35,511 records after preprocessing and outlier removal. Land price estimation was formulated as a supervised regression problem, and five machine learning models, including Random Forest, Decision Tree, Linear Regression, K-Nearest Neighbors (KNN), and Artificial Neural Network / Multilayer Perceptron (ANN/MLP), were implemented and evaluated.

Experimental results demonstrated substantial differences in predictive performance among the evaluated models. Random Forest achieved the best overall performance with a Mean Absolute Percentage Error (MAPE) of 0.76% and an R^2 score of 0.96, indicating strong predictive capability and generalization performance. The trained model was integrated into a Flask-based web application that enables users to retrieve land information, estimate land prices, and calculate land taxes in real time using Parcel Unique Identification (UPI).

The findings confirm that machine learning techniques can significantly improve valuation accuracy, consistency, transparency, and scalability in land administration systems. The proposed framework contributes toward the modernization of land governance in Rwanda by supporting intelligent and data-driven valuation processes.

Keywords: Machine Learning, Land Price Estimation, Automated Valuation Model, Geographic Information Systems

1. Introduction

Rapid urbanization and population growth have increased the demand for accurate, transparent, and scalable land valuation systems across developing countries [1], [2]. Land valuation plays a critical role in taxation, land administration, urban planning, compensation, infrastructure development, and real estate market regulation [3], [4]. In many developing economies, however, land valuation systems remain largely manual and dependent on expert judgment, resulting in inconsistencies, inefficiencies, limited scalability, and reduced transparency [1], [5].

In Rwanda, substantial progress has been achieved in digitizing land records through initiatives such as the Land Tenure Regularization Program and the implementation of digital land administration systems [5], [6]. Despite these developments, land valuation practices still rely heavily on traditional appraisal methods and administrative valuation schedules. These approaches are often subjective and unable to effectively adapt to rapidly changing market conditions, infrastructural development, and spatial variations in land value [3], [7].

Traditional valuation approaches, including comparative market analysis and hedonic pricing models, generally assume linear relationships between land price and explanatory variables [4], [8]. However, land markets are inherently complex and influenced by multiple interacting factors such as parcel size, land use, accessibility, infrastructure, location, and socioeconomic conditions [9], [10]. Consequently, conventional statistical models often fail to capture nonlinear relationships and complex feature interactions within heterogeneous urban environments [11], [12].

Recent advances in Machine Learning have introduced powerful predictive techniques capable of modeling complex and nonlinear relationships in large datasets [12], [13]. Machine learning models such as Random Forest, Decision Trees, K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANNs) have demonstrated strong predictive performance in automated valuation systems because of their ability to learn feature interactions and improve prediction accuracy [14], [15], [16].

Among these techniques, tree-based ensemble methods have shown superior performance in structured tabular datasets because of their robustness, resistance to overfitting, and ability to handle heterogeneous data [13], [17]. These capabilities make machine learning highly suitable for land valuation tasks where pricing patterns are influenced by diverse spatial and administrative factors [15], [18].

Despite these advancements, the application of machine learning in land valuation within Rwanda remains limited. Existing land valuation systems do not fully utilize the increasing availability of spatial and administrative land data [5], [6]. Furthermore, limited research has integrated machine learning models, spatial analysis, and practical deployment frameworks into a unified system applicable to real-world land administration processes.

This study proposes a machine learning-based framework for land price estimation in Huye District. The framework integrates data preprocessing, feature engineering, predictive modeling, evaluation, and deployment within a Flask-based web application capable of supporting real-time land price estimation and tax calculation.

1.1. Motivation

The motivation for this study arises from several limitations observed in existing land valuation systems in Rwanda. Current valuation processes are labor-intensive, time-consuming, and highly dependent on human expertise. This often introduces subjectivity, inconsistency, and delays in land price estimation [1], [7].

In addition, rapid urbanization and infrastructural development have made land markets increasingly dynamic and complex [2], [10]. Traditional valuation techniques are unable to effectively model the nonlinear relationships among factors influencing land prices [11], [12]. At the same time, the growing availability of digital land records and GIS data creates an opportunity to develop intelligent and automated valuation systems capable of improving valuation efficiency and reliability [18].

Therefore, there is a need for a robust and scalable framework capable of: improving prediction accuracy, reducing subjectivity and human bias, enhancing transparency and consistency, supporting real-time land valuation, facilitating data-driven land administration.

1.2. Research Objectives

The main objective of this study is to develop a machine learning-based framework for land price estimation in Huye District.

The specific objectives are:

- To collect and preprocess land-related spatial and administrative data.
- To develop predictive machine learning models for land price estimation.
- To compare the performance of multiple machine learning algorithms.
- To identify the most influential factors affecting land prices.
- To deploy the best-performing model in a web-based system for real-time prediction and tax estimation.

1.3. Contribution of the Study

This study makes several important contributions to the field of automated land valuation.

First, a comprehensive machine learning framework for land price estimation tailored to the Rwandan context is proposed.

Second, five machine learning models, including Random Forest, Decision Tree, K-Nearest Neighbors, Artificial Neural Network / Multilayer Perceptron, and Linear Regression, are implemented and evaluated using real-world land administration data.

Third, the study demonstrates the effectiveness of tree-based machine learning techniques, particularly Random Forest, in improving prediction accuracy and reducing valuation errors, consistent with findings from previous ensemble learning studies [15], [17].

Fourth, spatial and administrative land attributes are integrated into the predictive framework to improve model reliability and capture geographic variations in land value [9], [18].

Fifth, a Flask-based web application is developed to operationalize the predictive model, enabling users to retrieve land information, estimate land prices, and calculate taxes in real time using Parcel Unique Identification (UPI).

Finally, the study provides empirical insights into land market dynamics in Huye District and contributes toward the modernization of land administration systems in Rwanda.

1.4. Organization of the Paper

The remainder of this paper is organized as follows. Section 2 presents related work on land valuation and machine learning approaches. Section 3 describes the research methodology, including data collection, preprocessing, feature engineering, model development, and system architecture. Section 4 presents the experimental results and discussion. Section 5 concludes the paper and outlines recommendations for future research.

2. Related Work

Land valuation has long been an important component of land administration, urban planning, taxation, and real estate management [1], [3]. Traditional valuation approaches, including comparative market analysis and hedonic pricing models, have been widely adopted because of their interpretability and foundation in economic theory [4], [8]. However, these methods rely heavily on manual assessment and linear assumptions, making them less effective in modeling the complex relationships that influence land prices in dynamic urban environments [9], [11].

Several studies have investigated the application of statistical and machine learning methods for Automated Valuation Models (AVMs). Kamathe et al. (2026) applied multiple linear regression techniques to residential property valuation and found that traditional structural characteristics exhibited very weak predictive power, suggesting that external market and location-specific factors may play a more significant role in determining property values [11]. However, their approach assumed linear relationships between land attributes and market value, limiting its ability to capture nonlinear interactions among geographic, economic, and infrastructural factors. Similar limitations were observed in hedonic pricing studies, where valuation accuracy decreased significantly in heterogeneous urban regions [8], [10].

Borowiecki (2009) investigated the determinants of house prices and construction activity in the Swiss housing market using econometric and vector autoregressive (VAR) models. The study found that demographic factors, population growth, construction costs, and housing supply dynamics significantly influence property values, while GDP exhibited limited explanatory power. However, the methodology relied on traditional econometric techniques, which may be less effective in capturing complex nonlinear relationships and large-scale market dynamics [8].

Recent advancements in Machine Learning have significantly improved predictive performance in land and property valuation tasks [14], [15], [16]. Nagula (2025) compared multiple machine learning models for real estate price prediction and reported that Random Forest Regression achieved superior performance over traditional regression methods. The findings highlighted the effectiveness of ensemble learning in modeling nonlinear feature interactions and reducing prediction errors [15]. However, the research focused mainly on developed urban environments with highly structured datasets, limiting its applicability to developing countries where data quality and institutional contexts differ substantially.

Chen and Guestrin (2016) introduced XGBoost, a scalable and regularized gradient boosting framework designed to improve predictive accuracy and computational efficiency. The study demonstrated that XGBoost effectively handles sparse and large-scale datasets through optimized tree learning, parallel processing, cache-aware computation, and regularization techniques, making it a powerful tool for complex prediction tasks [19]. However, the research primarily focused on algorithmic performance and scalability rather than the integration of deployment frameworks for practical applications such as land administration and property valuation systems.

Deng and Zhang (2025) investigated automated machine learning for residential property valuation and demonstrated that integrating domain-specific property characteristics with geographic and environmental information significantly improves valuation accuracy. Their proposed AutoML4RPV framework incorporated spatial datasets, feature engineering, model generation, interpretation, and deployment capabilities, achieving superior predictive performance across multiple housing markets [16]. However, the study focused primarily on residential property valuation and did not specifically address land administration functions such as land taxation, cadastral valuation, and regulatory zoning applications.

The integration of Geographic Information Systems (GIS) with machine learning has also received increasing attention in valuation research. GIS-based studies demonstrated that spatial features such as accessibility, proximity to roads, neighborhood development, and land use patterns substantially influence land prices [10], [18]. Although GIS integration improves prediction capability, many existing systems remain research prototypes and are not operationalized into real-time applications for public use.

In addition, several automated valuation studies based on Artificial Neural Networks (ANNs) reported strong predictive performance compared to traditional statistical approaches [20], [21], [22]. Despite achieving lower prediction errors, ANN-based models often suffer from reduced interpretability, high computational complexity, and sensitivity to training data quality. These challenges limit their practical adoption in government land administration systems where transparency and explainability are important.

Tree-based machine learning methods such as Random Forest and Decision Trees have also demonstrated strong performance in structured valuation datasets because of their robustness and ability to minimize overfitting [17], [23]. Breiman (2001) demonstrated that Random Forest models improve predictive accuracy and stability by aggregating multiple decision trees generated through random feature selection and bagging [17]. Quinlan (1986) showed that decision trees are effective classification and predictive learning tools that construct interpretable models using information gain [23]. These machine learning techniques are well suited for heterogeneous land administration datasets containing mixed spatial, categorical, and numerical variables.

Most existing studies share several common limitations. First, a large proportion of valuation research has been conducted in developed countries with mature real estate databases and advanced institutional infrastructures [15], [16]. As a result, the findings may not generalize effectively to developing countries such as Rwanda, where land markets, valuation practices, and data availability differ significantly. Second, many studies focus exclusively on predictive accuracy while neglecting deployment considerations necessary for real-world implementation. Third, limited research integrates machine learning models with administrative land management systems capable of supporting taxation and decision-making processes in real time.

Furthermore, existing valuation systems in Rwanda remain predominantly manual and expert-driven despite the availability of digitized land records through national land administration initiatives [5], [6]. This creates challenges related to subjectivity, inconsistency, scalability, and transparency in land valuation processes. The absence of intelligent automated valuation systems represents a major research and practical gap within the Rwandan land administration sector.

To address these limitations, this study proposes a machine learning-based framework for land price estimation tailored to the Rwandan context, with a case study in Huye District. Unlike previous studies, the proposed framework:

- integrates spatial and administrative land data,
- evaluates multiple machine learning and ensemble learning models,
- incorporates feature engineering and preprocessing pipelines,
- and deploys the best-performing model within a Flask-based web application for real-time land price prediction and tax estimation.

2.1. Research Gap and Comparative Analysis

Several previous studies have applied machine learning techniques in property and land valuation systems. However, most existing studies were conducted in developed countries with highly structured real estate datasets and mature market infrastructures [15], [16]. In contrast, limited research has focused on developing-country contexts such as Rwanda, where land administration systems are still transitioning toward digital transformation [5], [6].

Nagula (2025) investigated machine learning techniques for real estate price prediction and reported that ensemble learning methods, particularly Random Forest Regression, outperform traditional regression approaches due to their ability to capture complex nonlinear relationships and improve generalization performance [15]. The study demonstrated that Random Forest achieved the highest predictive accuracy and the lowest prediction error among the evaluated models. However, the research focused primarily on residential housing datasets and did not incorporate administrative land management information or specialized frameworks for land administration applications.

Kamathe et al. (2026) applied statistical regression techniques to residential property valuation and found that traditional structural property characteristics exhibited very weak predictive power, with the multiple regression model achieving an R^2 value of -0.0001. Their findings suggested that housing prices are influenced by external market conditions, location-specific factors, and other unmeasured variables that are not adequately captured by conventional regression models [11]. Similar observations were made in the present study, where Linear Regression achieved the lowest predictive performance among the evaluated models, recording an R^2 score of 0.23 and a MAPE value of 10.06%, indicating limited capability in modeling complex interactions among land-related variables.

Several studies involving Artificial Neural Networks reported improved prediction accuracy compared to traditional statistical methods [20], [21]. Nevertheless, neural network models often require extensive parameter tuning, larger datasets, and higher computational resources. In this study, the ANN/MLP model achieved an R^2 score of 0.78 and a MAPE value of 5.16%, which remained lower than the performance achieved by tree-based models.

Tree-based machine learning approaches, particularly Random Forest, demonstrated superior performance in previous valuation studies because of their robustness and ability to reduce overfitting [13], [17]. Similar findings were obtained in this study, where Random Forest achieved the best overall predictive performance with an R^2 score of 0.96 and a MAPE value of 0.76%. despite the progress achieved in previous research, several important limitations remain: many studies focus only on prediction accuracy without practical deployment, limited integration of GIS and administrative land records exists, few systems support real-time land valuation and taxation, and limited research has been conducted using Rwandan land administration data.

This study addresses these gaps by:

- integrating spatial and administrative land data;
- evaluating multiple machine learning models in the Rwandan context;
- implementing a real-time Flask-based valuation system;
- and demonstrating the effectiveness of Random Forest for automated land valuation in Huye District.

The study therefore contributes both methodological and practical advancements toward intelligent land administration and automated valuation systems in Rwanda.

3. Methodology

3.1. Study Area

The study was conducted in Huye District, one of the rapidly developing districts in the Southern Province of Rwanda. Huye District is characterized by continuous urban expansion, increasing population growth, infrastructural development, and diverse land use activities including residential, commercial, agricultural, and institutional land utilization.

The district was selected as the study area because of: the availability of digitized land administration records [5], increasing land market activities, ongoing urbanization processes [6], and the growing need for transparent and efficient land valuation systems.

In addition, Huye District provides a suitable environment for evaluating machine learning-based valuation models because of the diversity of land parcels and spatial characteristics across sectors and administrative zones.

3.2. Data Collection and Integration

A multi-source dataset was developed by integrating data obtained from different land administration and spatial information sources. The dataset combined both spatial and administrative land attributes relevant to land price estimation.

The primary data sources included: official land valuation schedules, land registry databases, geographic information systems (GIS), administrative land records.

The collected features included: parcel size, sector location, cell location, land use category, valuation zone, accessibility indicators, infrastructure proximity, administrative identifiers.

The integration of multiple datasets was necessary to ensure that both spatial and regulatory factors influencing land prices were adequately represented within the predictive framework [10], [18].

Since the data originated from heterogeneous sources, harmonization procedures were applied to standardize formats, attribute names, measurement units, and coding structures. This process improved dataset consistency and compatibility during machine learning model development.

3.3. Data Preprocessing

Data preprocessing was performed to improve data quality, consistency, and model reliability. Raw datasets often contain missing values, inconsistencies, duplicates, and outliers that can negatively affect predictive performance if not properly handled [12], [13].

The following preprocessing procedures were applied:

Data Cleaning: duplicate and inconsistent records were identified and removed to ensure dataset integrity.

Missing Value Handling: records containing incomplete or invalid values were either corrected, imputed, or removed depending on the severity and nature of the missing information.

Outlier Detection and Removal: extreme values that could distort model learning were identified using statistical analysis and removed from the dataset.

Categorical Data Encoding: categorical variables such as land use type, sector, and valuation zone were transformed into numerical representations suitable for machine learning algorithms [12].

Feature Scaling and Normalization: numerical variables were normalized to improve model performance and ensure balanced feature contribution during training.

The initial dataset contained 51,955 records. After preprocessing and outlier removal, 35,511 records were retained for model development and evaluation.

Table 1: Dataset Before and After Preprocessing

Stage	**Records**
Raw Dataset	51,955
Cleaned Dataset	35,511
Removed Records	16,444

The final processed dataset was divided into input features (X) and target variables (Y) before model training.

3.4. Feature Engineering

Feature engineering was conducted to identify and construct variables that significantly influence land prices. The selection of features was guided by: literature review [15], [16], domain expertise in land administration, and data availability.

The engineered features were categorized into three major groups:

Spatial Features: these features describe the geographic location and spatial characteristics of land parcels. Examples include: sector location, cell location, accessibility, and infrastructure proximity

Physical Features: these features describe the physical properties of land parcels. Examples include: parcel size and parcel dimensions.

Regulatory Features: these features describe administrative and land-use regulations affecting land value. Examples include: land use type, valuation zone, administrative classification.

Important predictive features included: parcel size, sector location, land use type, accessibility, and valuation zone.

Feature encoding techniques were applied to convert categorical variables into numerical forms compatible with machine learning algorithms.

Feature engineering improved the ability of models to capture complex relationships among land characteristics and market prices [17], [18].

3.5. Mathematical Formulation

Land price estimation was formulated as a supervised machine learning regression problem.

The feature vector is represented as:

$$X = (x_1, x_2, x_3, \dots, x_n) \tag{1}$$

Where:

- X represents the input feature vector
- $x_1, x_2, x_3 \dots x_n$ represent individual land attributes

The target variable is represented as: Y

Where:

- Y represents the land price to be predicted

The predictive relationship between the input features and land price is defined as:

$Y = f(X) + \varepsilon$

Where:

- f(X) represents the predictive machine learning function
- ε represents the random error term

Random Forest Formulation

Since Random Forest achieved the best performance in this study, its mathematical representation is defined as:

$$\hat{f}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{2}$$

Where:

- B represents the total number of decision trees
- $T_\beta(x)$ represents the prediction from the b-th decision tree
- ŷ represents the final predicted land price

Random Forest improves prediction performance by combining multiple decision trees and averaging their predictions to reduce overfitting and improve generalization [17].

3.6. Machine Learning Models

Five machine learning models were implemented and evaluated in this study to compare predictive performance across different learning approaches.

Linear Model

- Linear Regression

Linear Regression was used as a baseline statistical model to evaluate the effectiveness of traditional linear approaches in land valuation [11].

Instance-Based Model

- K-Nearest Neighbors (KNN)

KNN predicts land prices based on similarities between neighboring parcels in feature space [24].

Tree-Based Models

- Decision Tree
- Random Forest

Decision Tree models learn hierarchical relationships among features, while Random Forest improves prediction accuracy by combining multiple decision trees through ensemble learning [17], [23].

Neural Network Model

- Artificial Neural Network / Multilayer Perceptron (ANN/MLP)

The ANN/MLP model was implemented to evaluate the capability of neural networks in modeling nonlinear relationships within land valuation data [21], [22].

These models were selected to compare: traditional statistical approaches, distance-based learning, tree-based learning, ensemble learning, and neural network approaches.

3.7. Model Training and Evaluation

The processed dataset was divided into 80% training data, and 20% testing data

The training dataset was used to train machine learning models, while the testing dataset was used to evaluate predictive performance on unseen data.

Cross-validation techniques and hyperparameter tuning were applied to improve model generalization and reduce overfitting [12], [13].

The performance of the models was evaluated using two regression evaluation metrics.

Mean Absolute Percentage Error (MAPE)

MAPE measures the average percentage difference between predicted and actual values.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{3}$$

Lower MAPE values indicate better prediction accuracy.

Coefficient of Determination (R^2)

R^2 measures how well the model explains variations in land prices.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{4}$$

Higher R^2 values indicate stronger predictive performance and better model fit.

3.8. System Architecture

The proposed framework follows a layered system architecture consisting of five major layers.

1. Data Layer

This layer contains land registry databases, GIS datasets, and administrative records used for model traiing and prediction.

2. Processing Layer

This layer performs data cleaning, preprocessing, feature engineering, and transformation operations.

3. Model Layer

This layer contains the machine learning algorithms responsible for training and prediction.

4. Application Layer

The application layer was implemented using Flask to provide user interaction and system functionality.

5. Output Layer

This layer generates land information retrieval, land price estimation, and tax calculation results.

The developed system allows users to input a Parcel Unique Identification (UPI) and receive real-time land information, estimated land price, and estimated land tax values.

The system demonstrates the practical applicability of machine learning in intelligent land administration and automated valuation systems.

4. Results and Discussion

4.1. Model Performance Evaluation

The performance of the implemented machine learning models was evaluated using two regression evaluation metrics: Mean Absolute Percentage Error (MAPE) and the Coefficient of Determination (R^2). These metrics were selected because they provide complementary insights into model accuracy and predictive capability [12], [13].

Mean Absolute Percentage Error (MAPE) measures the average percentage difference between predicted and actual land prices. Lower MAPE values indicate higher prediction accuracy.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{5}$$

The Coefficient of Determination (R^2) measures how well the model explains the variance in land prices. Higher R^2 values indicate stronger predictive performance and better model fit.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{6}$$

Five machine learning models were evaluated:

- Random Forest
- Decision Tree
- K-Nearest Neighbors (KNN)
- Artificial Neural Network / Multilayer Perceptron (ANN/MLP)
- Linear Regression

The experimental results revealed substantial differences in predictive performance among the evaluated models.

Table 1: Model Performance Comparison

Model	MAPE (%)	R^2 Score
Random Forest	0.76	0.96
Decision Tree	0.69	0.93
K-Nearest Neighbors (KNN)	2.06	0.88

ANN / MLP	5.16	0.78
Linear Regression	10.06	0.23

The results demonstrate that tree-based machine learning models significantly outperform traditional statistical methods in land price estimation tasks.

Among all evaluated models, the Random Forest model achieved the highest overall predictive performance with an R^2 score of 0.96 and a very low MAPE value of 0.76%. This indicates that Random Forest was highly effective in capturing the complex and nonlinear relationships between land characteristics and market price. The strong performance of Random Forest can be attributed to its ensemble learning mechanism, which combines multiple decision trees to improve prediction accuracy and reduce overfitting.

The Decision Tree model also produced strong performance results with a MAPE value of 0.69% and an R^2 score of 0.93. The low prediction error demonstrates that Decision Trees can effectively model nonlinear patterns in land valuation data. However, compared to Random Forest, Decision Trees are more sensitive to overfitting because predictions rely on a single tree structure.

The K-Nearest Neighbors (KNN) model achieved moderate predictive performance with an R^2 score of 0.88 and a MAPE value of 2.06%. The model performed reasonably well in identifying similarities among neighboring land parcels but showed limitations when handling high-dimensional datasets and complex feature interactions.

The ANN/MLP model achieved an R^2 score of 0.78 and a MAPE value of 5.16%. Although neural networks are capable of learning complex relationships, their performance in this study was lower than that of tree-based ensemble methods. This may be attributed to dataset characteristics, feature distribution, parameter tuning limitations, and the relatively structured nature of tabular land valuation data.

Linear Regression produced the weakest performance among all evaluated models, with a MAPE value of 10.06% and an R^2 score of 0.23. The low R^2 score indicates that the model failed to adequately explain variations in land prices. This poor performance confirms that linear models are insufficient for capturing the nonlinear relationships and feature interactions that characterize land valuation problems.

The findings demonstrate that machine learning techniques, particularly tree-based models, are highly effective for land price estimation. The superior performance of Random Forest and Decision Tree models confirms that nonlinear learning approaches are better suited for modeling complex land market dynamics than traditional regression methods.

4.2. Comparative Analysis of Models

The comparative analysis highlights clear differences between traditional statistical models and advanced machine learning approaches. Linear Regression performed poorly because of its assumption of linear relationships between explanatory variables and land prices. In contrast, machine learning models were capable of learning nonlinear interactions among features such as location, parcel size, land use type, and accessibility.

Tree-based models consistently outperformed other techniques because they effectively partition feature spaces and capture hierarchical decision patterns. Random Forest achieved the best balance between prediction accuracy and model generalization because of its ensemble structure, which reduces variance and improves robustness.

Decision Tree models achieved very low prediction errors but are more prone to overfitting when compared to ensemble methods. Random Forest overcame this limitation by aggregating predictions from multiple trees, leading to improved stability and higher predictive reliability.

The ANN/MLP model showed lower performance compared to tree-based approaches despite its ability to model nonlinear relationships. Neural networks generally require larger datasets, extensive parameter optimization, and computational resources to achieve optimal performance. In structured tabular datasets such as land valuation records, tree-based models often outperform neural networks because they better handle categorical and heterogeneous features.

KNN achieved moderate performance but exhibited sensitivity to feature scaling and data dimensionality. The algorithm depends heavily on distance-based similarity calculations, which may become less effective in complex datasets with multiple interacting variables.

The findings obtained in this study are consistent with previous research demonstrating that ensemble and tree-based machine learning models outperform traditional regression techniques in valuation tasks. However, compared to many previous studies that focused mainly on prediction experiments, this research additionally integrates administrative land data, GIS-related attributes, and a deployable web-based application for real-time land price estimation and tax calculation. This practical integration strengthens the applicability of the proposed framework within real-world land administration systems in Rwanda.

4.3. Implications of the Findings

The findings of this study have important implications for land administration and valuation systems in Rwanda. The results demonstrate that machine learning can significantly improve valuation accuracy, consistency, and transparency compared to manual valuation approaches.

The high predictive performance achieved by Random Forest indicates that automated valuation systems can support:

- Real-time land valuation
- Improved taxation systems
- Faster land transaction processing
- Data-driven urban planning
- Reduced human bias and subjectivity

The integration of machine learning into land administration systems can also improve scalability by enabling large volumes of land records to be analyzed efficiently and consistently.

Furthermore, the study demonstrates the practical applicability of machine learning in developing-country contexts where valuation systems are still transitioning from manual to digital processes.

4.4. Limitations of the Study

Although the proposed framework achieved strong predictive performance, several limitations remain.

First, the study relied primarily on administrative and spatial datasets, while real-time market transaction data was limited. The inclusion of live transaction records could further improve prediction accuracy and responsiveness to market fluctuations.

Second, the study focused exclusively on Huye District, which may limit the generalizability of the findings to other geographic regions with different market characteristics.

Third, some potentially influential socioeconomic and environmental variables were unavailable during data collection. Additional variables such as income levels, commercial activity, and demographic indicators may improve predictive performance in future studies.

Finally, although Random Forest achieved excellent results, further optimization and comparison with advanced deep learning architectures could provide additional insights into model performance.

4.5. Summary of Findings

The experimental results confirm that machine learning approaches significantly improve land price estimation accuracy compared to traditional regression methods. Among the evaluated models, Random Forest achieved the best overall performance with an R^2 score of 0.96 and a MAPE value of 0.76%.

The study demonstrates that:

- Tree-based models outperform traditional regression approaches.
- Ensemble learning techniques provide strong predictive accuracy and robustness.
- Machine learning effectively captures nonlinear land market relationships.
- Automated valuation systems are feasible within the Rwandan land administration context.
- Spatial and administrative features play a critical role in land valuation prediction.

5. Conclusion

This study developed and evaluated a machine learning-based framework for land price estimation in Huye District. By integrating spatial and administrative land data with predictive machine learning techniques, the study demonstrated that automated valuation models can significantly improve land price estimation accuracy, consistency, transparency, and efficiency compared to traditional manual valuation approaches.

Five machine learning models, including Random Forest, Decision Tree, K-Nearest Neighbors (KNN), Artificial Neural Network / Multilayer Perceptron (ANN/MLP), and Linear Regression, were implemented and evaluated using real-world land administration data. The experimental results revealed substantial performance differences among the models. Among the evaluated algorithms, Random Forest achieved the best predictive performance with a Mean Absolute Percentage Error (MAPE) of 0.76% and an R^2 score of 0.96, confirming the effectiveness of ensemble tree-based learning techniques for structured land valuation datasets.

The findings further demonstrated that machine learning models are capable of capturing complex nonlinear relationships among land attributes such as parcel size, location, accessibility, land use type, and valuation zones. The superior performance of Random Forest indicates that ensemble learning approaches are highly suitable for automated land valuation systems within developing-country contexts where land markets are heterogeneous and dynamic.

In addition, the implementation of a Flask-based web application demonstrated the practical applicability of the proposed framework in real-world land administration processes. The developed system enables users to retrieve land information, estimate land prices, and calculate land taxes in real time using Parcel Unique Identification (UPI), thereby improving accessibility and operational efficiency.

The study contributes toward the modernization of land administration systems in Rwanda by providing an intelligent, scalable, and data-driven framework capable of supporting transparent and consistent valuation processes. The proposed framework also supports broader digital land governance initiatives aimed at improving decision-making, taxation, and urban planning.

Despite the strong predictive performance achieved in this study, several limitations remain. The framework relied primarily on administrative and spatial datasets, while real-time market transaction data and certain socioeconomic indicators were limited. Furthermore, the study focused specifically on Huye District, which may limit direct generalization to other geographic regions with different land market characteristics.

Future research should extend the framework to a national scale by incorporating datasets from multiple districts across Rwanda. Additional variables such as environmental indicators, demographic factors, economic activities, and real-time market transaction data should also be integrated to further improve predictive performance. Moreover, future studies may explore advanced deep learning architectures and hybrid machine learning approaches to enhance valuation accuracy and system adaptability.

6. References

1. The World Bank, "Land Governance Assessment Framework: Implementation Manual," The World Bank, Washington, DC, USA, Implementation Manual Version: October 2013, Oct. 2013.
2. The World Bank, "Open Knowledge Repository," Open Knowledge Repository. Accessed: May 29, 2026. [Online]. Available: https://openknowledge.worldbank.org/entities/publication/c7bc435a-d635-5136-aacf-7cf0f5f3c6cf
3. Australian Property Institute, "International Valuation Standards 2022-2024," Australian Property Institute. Accessed: May 29, 2026. [Online]. Available: https://www.api.org.au/standards/ivs-2022-2024/
4. A. Institute, *Appraisal of Real Estate*, 15th Edition. Chicago: Appraisal Institute, 2020. [Online]. Available: https://www.amazon.com/Appraisal-Real-Estate-15th/dp/1935328786
5. National Land Authority (NLA), "Land Administration / Registrar of Land Titles," Land Administration / Registrar of Land Titles. Accessed: May 29, 2026. [Online]. Available: https://www.lands.rw/land-administration-registrar-of-land-titles
6. Ministry of Environment, "National Land Policy," Republic of Rwanda, Ministry of Environment, Kigali, Rwanda, Government Policy Report, Jun. 2019. [Online]. Available: https://www.environment.gov.rw/fileadmin/user_upload/Moe/Publications/Policies/Revised_National_Land_Policy-Final_Version_2019.pdf
7. International Association of Assessing Officers (IAAO), "Standard on Automated Valuation Models (AVMs)," nternational Association of Assessing Officers, Kansas City, Missouri, USA, Standard, Jul. 2018. [Online]. Available: https://www.iaao.org/wp-content/uploads/Standard_on_Automated_Valuation_Models.pdf
8. Karol Jan Borowiecki, "The Determinants of House Prices and Construction: An Empirical Investigation of the Swiss Housing Economy," *Int. Real Estate Rev.*, vol. 12, no. 3, pp. 193–220, 2009.
9. Alonso, William, *Location and Land Use: Toward a General Theory of Land Rent*. Cambridge, Massachusetts, USA: Harvard University Press. doi: 10.4159/harvard.9780674730854.
10. "Edwin Mills (economist)," *Wikipedia*. Mar. 08, 2026. Accessed: May 29, 2026. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Edwin_Mills_(economist)&oldid=1342432240
11. Rupali Kamathe, Akshata Lembhe, Seema Dokrimare, and Vishal Dubey, "Statistical Analysis of Residential Property Valuation: A Data-Driven Study of Housing Market Determinants," *Int. J. Comput. Sci. Eng.*, vol. 14, no. 4, pp. 139–143, Apr. 2026, doi: 10.26438/ijcse.v14i4.7363.
12. G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning: with applications in R*, Second edition. in Springer texts in statistics. New York, NY: Springer, 2021.

13. T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed. in Springer series in statistics. New York, NY: Springer, 2009.
14. Nur Shahirah Ja'afar, Junainah Mohamad, and Suriatini Ismail, "Machine Learning for Property Price Prediction and Price Valuation: A Systematic Literature Review," *Plan. Malays. J. Malays. Inst. Plan.*, vol. 19, no. 3, pp. 411–422, 2021, doi: 10.21837/pm.v19i17.1018.
15. Sairaj Nagula, "Real Estate Price Prediction Using Machine Learning Models," vol. 157–163, no. VII (or 7), Jul. 2025, doi: 10.22214/ijraset.2025.72962.
16. [Lin Deng and Xueqing Zhang, "Automated Machine Learning for Residential Property Valuation," *Eng. Appl. Artif. Intell.*, vol. 160, no. Part C, p. 112035, 2025, doi: 10.1016/j.engappai.2025.112035.
17. L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
18. Gabriela Droj, Anita Kwartnik-Pruc, and Laurențiu Droj, "A Comprehensive Overview Regarding the Impact of GIS on Property Valuation," *ISPRS Int. J. Geo-Inf.*, vol. 13, no. 6, p. 175, May 2024, doi: https://www.mdpi.com/2220-9964/13/6/175.
19. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
20. Abidoye, Rotimi Boluwatife and Chan, Albert P. C., "Artificial Neural Network in Property Valuation: Application Framework and Research Trend," *Prop. Manag.*, vol. 35, no. 5, pp. 554–571, 2017, doi: tps://doi.org/10.1108/PM-06-2016-0027.
21. C. M. Bishop, *Pattern Recognition and Machine Learning*, Softcover reprint of the original 1st edition 2006 (corrected at 8th printing 2009). in Information science and statistics. New York, NY: Springer New York, 2016.
22. Goodfellow, Ian, Bengio, Yoshua, and Courville, Aaron, *Deep Learning*. Cambridge, Massachusetts, USA: MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org/
23. J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1007/BF00116251.
24. T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967, doi: 10.1109/TIT.1967.1053964.