IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

PREDICTIVE MODEL FOR PREDICTING LAND VALUE USING MACHINE LEARNING: *A CASE STUDY OF NAIROBI METROPOLITAN AREA.*

[1]Chepkorir Stellah Rotich, [2]Professor Robert Oboko
[1]Student, [2]Supervisor
[1] School of Computing and Informatics
[1]University of Nairobi, Nairobi, Kenya

Abstract: Predicting land value is a complex task with significant implications for real estate, urban planning, land rating, and investment decision-making. Traditional land valuation methods often rely on assessment and historical data analysis, leading to potential inaccuracies and inefficiencies. This challenge has necessitated the need for a model that will assist in detecting and predicting land value. Land assessors and investors around the world have also turned to technology to improve their precision in land valuation. This is because of the rise of artificial intelligence which has shown tremendous impact over the years in several sectors. The emergence of machine learning (ML) techniques has provided new opportunities to develop predictive models capable of estimating with higher accuracy and efficiency. This research aims to create an advanced predictive model for predicting land value using machine learning algorithms multiple linear regression and XG Boost. It seeks to explore optimal solutions in predicting the land value, that will be beneficial to land value stakeholders including investors, land assessors, urban planners, and tax authorities. The goal is to identify the best model for predicting land value based on its characteristics such as location, land size, topology, development under it, soil type, land use, ownership type and others. The research will involve comprehensive data collection, including land parcel data and characteristics of data that affects land value.

Key Words- Land Valuation, Machine Learning, Regression, Random Forest, Decision Trees, XGBoost

1.0 INTRODUCTION.

Valuation of land is a critical aspect of land transactions, land development, and infrastructure projects. Traditional land valuation methods often rely on manual assessment and historical data analysis, which may be subjective and prone to inaccuracies. However, the advent of machine learning (ML) techniques has provided new opportunities to develop predictive models capable of estimating with higher accuracy and efficiency. Machine Learning approaches offer the potential to revolutionize the field of land valuation. The land value is the value of the piece of land including both the value of the land itself and any improvements that have been made to it.

In cities like Nairobi Metropolitan Area, where nearly all land sales data represent transfers of land with improvements, it is very difficult to divide prices between land and building components. The study on land value trend is felt important to support the decisions in urban planning. (Thomas, 2000). From past experiences, there is no common understanding on how assessors arrive at a value of land in Nairobi Metropolitan Area, various assessors come up with different values for the same piece of land depending on the reason and method for the valuation. Based on the biased ways in which human assessors calculate the value of the land, it has resulted in skepticism as to the feasibility of this process. This has led to practical problems of land assessment and have in many instances led investors lose money in the course of acquiring and investing in land, since many times it is overvalued or undervalued at the time of sale.

In practice, assessors calculate land value based on various variables inputs, these inputs vary depending on various aspects: location, development of the property, size of the land, usage of land, topology, type of soil among others. These variables inputs when collected over years, make knowledge base, and with the aggregation of all these input functions, output membership functions can be calculated. The result is a value of the land. This study will use Nairobi Metropolitan Area as a case study. The development of a predictive model for predicting land value using machine learning holds significant promise for enhancing the accuracy, efficiency, and transparency of land valuation processes. By leveraging advanced data analytics and predictive modeling techniques, stakeholders can make more informed decisions regarding property investments, land use planning, and development projects. This

study aims to contribute to the growing body of research on machine learning applications in property valuation and advance the field of land valuation towards more data-driven and evidence-based practices.

2.0 NEED OF THE STUDY.

Valuation is the analytical process of determining the current or projected worth of an asset. Traditional methods of land valuation often rely on subjective judgment and historical data. This is not easy to compute as several key inputs determine the value of the land. Thus this is one of the greatest challenge in land transactions is accurate valuation. From past experiences, there is no common understanding on how land assessors arrive at a value of land, various assessors come up with different values for the same piece of land depending on the reason for the valuation. Based on the biased ways in which human assessors calculate the value of the land, it has resulted in skepticism as to the feasibility of this process and has become a major stumbling block in land transactions. This has led to practical problems of land assessment and have in many instances led investors lose money in the course of acquiring land, since many times it is overvalued at the time of sale.

In practice assessors calculate land value based on various variables inputs, these inputs vary depending on various aspects: location, development of the property, size of the land, usage of land, topology, type of soil among others. These variables inputs when collected over years, make knowledge base, and with the aggregation of all these input functions, output membership functions can be calculated. The result is a value of the land. This study will use Nairobi Metropolitan Area as a case study.

By leveraging machine learning algorithms, it is possible to analyze large datasets comprising diverse sets of features and identify complex patterns that contribute to accurate land value predictions.

3.0 LITERATURE REVIEW

Predictive modeling for predicting land value using machine learning techniques has emerged as a significant area of research in recent years, driven by the increasing availability of data and advancements in machine learning algorithms. This literature review synthesizes existing research to provide a comprehensive understanding of the methodologies, challenges, and applications in this domain.

Various machine learning algorithms have been applied to develop predictive models for estimating. Random forest, gradient boosting, regression, support vector machines, and neural networks are among the commonly used algorithms. Studies by Li et al. (2018) and Smith et al. (2020) have demonstrated the effectiveness of ensemble learning techniques such as random forest and gradient boosting in capturing complex relationships between land attributes and market values.

Feature engineering plays a crucial role in predictive modeling for land value prediction. Researchers have explored techniques for selecting and engineering relevant features that influence land value, including spatial, temporal, economic, and environmental factors. Studies by Zhang et al. (2019) and Wang et al. (2021) have emphasized the importance of incorporating spatial features, such as proximity to amenities and transportation networks, into predictive models using geographic information systems (GIS).

Predictive models for land valuation often incorporate economic indicators and environmental factors to account for broader market trends and sustainability considerations. Studies by Li et al. (2020) and Wang et al. (2024) have demonstrated the importance of integrating economic variables, such as GDP growth rates and unemployment rates, as well as environmental variables, such as air quality indices and green space availability, into predictive models for more robust valuation assessments.

Temporal analysis and time-series forecasting techniques have been utilized extensively to capture temporal trends and seasonality in. Smith and Johnson (2016) explored the application of autoregressive integrated moving average (ARIMA) models, recurrent neural networks (RNNs), and long short-term memory (LSTM) networks for predicting future based on historical trends and economic indicators. Similarly, Lee et al. (2023) also investigated the effectiveness of these techniques in predicting, emphasizing the importance of considering temporal dynamics and economic factors in land valuation models.

Recent studies have illustrated the effectiveness of Decision Trees in land value prediction. For example, research by Zhang et al. (2022) demonstrated that Decision Trees could effectively capture complex relationships between land attributes and values, providing a robust alternative to traditional regression models. Liu et al. (2023) further highlighted the algorithm's interpretability and ease of use, making it a suitable choice for real estate valuation tasks. Also Random Forest in land value prediction. For example, Zhang et al. (2022) showed that Random Forest outperformed traditional regression models and other machine learning algorithms in capturing non-linear relationships between land attributes and values. Liu et al. (2023) further highlighted the algorithm's robustness and interpretability, making it a preferred choice for real estate valuation tasks.

4.0 RESEARCH METHODOLOGY

The methodology section outlines the plan and method that how the study is conducted. This includes Universe of the study, sample of the study, Data and Sources of Data, study's variables and analytical framework. The details are as follows;

4 .1 Population and Sample

To determine factors, features and variables that influence land value, a questionnaire was developed and shared among the 100 people to answers a few questions on land features out of which 58 responded.

Further to this, secondary data was collected at Nairobi City County Lands Department. The data contains Nairobi City County newly valued data between 2020 and 2022 comprising of 776 land parcels. The data is useful for assessing the performance of land as a key to future investment and land value predictions.

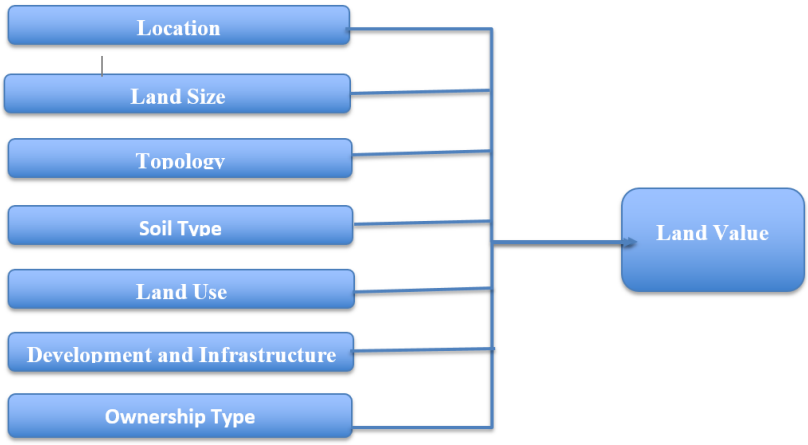
4.2 Data and Sources of Data

Land parcels secondary data was collected at Nairobi City County Land Department.

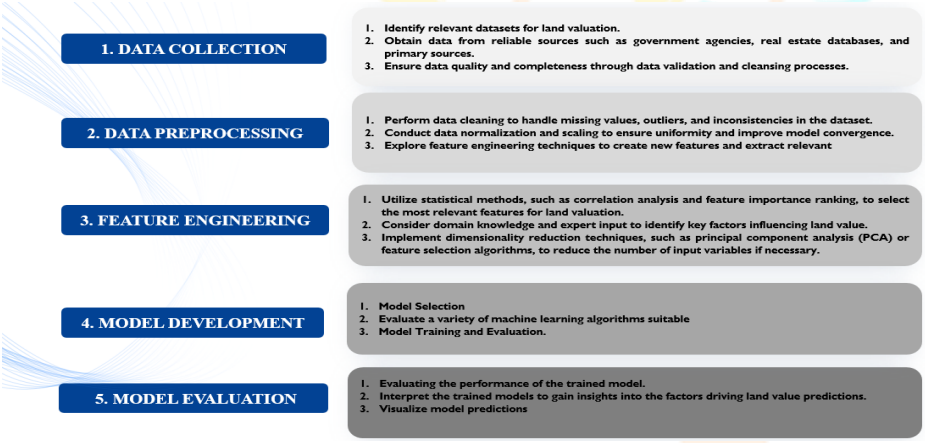
Primary data was collected from 100 respondents using random sampling within Nairobi Metropolitan Area.

4.3 Conceptual framework

Variables of the study contains dependent and independent variable. The study used pre-specified method for the selection of variables. The study used the Land Value are as dependent variable. From the land parcel features the land value is modeled and predicted.



Design Methodology: Agile CRISP-DM



4.4 Model Training and Evaluation:

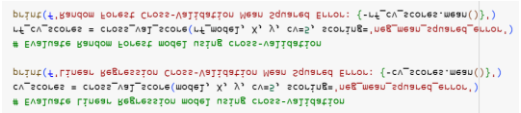
Machine Learning: The following supervised machine learning algorithms were chosen based on the data acquired which were already labelled with land value and features of the land. This also the based on the study of the traditional valuation methods currently in use. They were then compared to see which one works best for land value prediction.

1. Regression.
2. Random Forest.
3. Gradient Boosting Machines.
4. Decision Trees.

Dataset was divided into training, validation, and testing sets to ensure robust model evaluation. The data was split into 70% of data for training and 30% for testing the model.

```
[ ]  X = data.drop('SiteValue', axis=1)
     y = data['SiteValue']
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
     scaler = StandardScaler()
     X_train_scaled = scaler.fit_transform(X_train)
     X_test_scaled = scaler.transform(X_test)
```

Cross-Validation: Implemented cross-validation to assess model performance and generalization capability.



Hyper parameter Tuning: Optimize model parameters using techniques grid search to enhance performance.

```
param_grid_xg = {'max_depth': [4, 8, 12],
                 'n_estimators': [100, 200, 300],
                 'learning_rate': [0.1, 0.05, 0.01]}
grid_search_xg = GridSearchCV(xgb, param_grid = param_grid_xg, cv = 5)
grid_search_xg.fit(X_train, y_train)
best_model_xg = grid_search_xg.best_estimator_
best_params_xg = grid_search_xg.best_params_
predictions_xg = best_model_xg.predict(X_test)
```

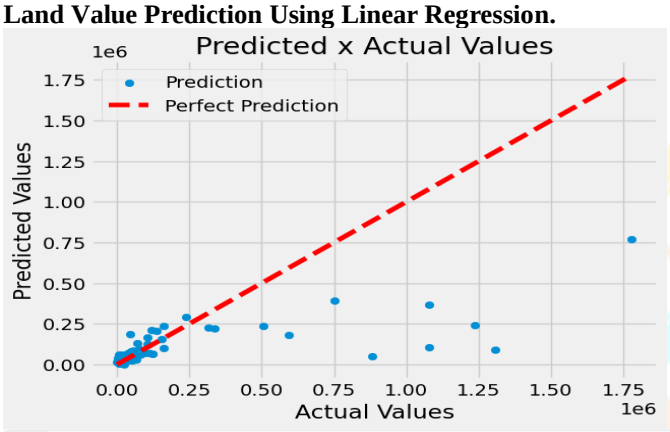
4.4.1 Linear Regression

Regression is widely used for prediction and forecasting problems. It is also used to understand which among the independent variables are related to the dependent variable and to explore the forms of these relationships. If more independent variables are added, it is able to determine an estimating equation that describes the relationship with greater accuracy. Linear regression is one of the simplest and most widely used regression techniques. In this case, it was used to model the relationship between the dependent variable land value and independent variables features influencing land value that included land size, location, land use and type of ownership by fitting a linear equation to the observed data. The general form of the linear regression equation is as shown below:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \epsilon$$

where Y is the predicted land value, β_0 is the intercept, $\beta_1, \beta_2, ..., \beta_n$ are the coefficients of the features $X_1, X_2, ..., X_n$, and ϵ is the error term. Despite its simplicity, linear regression can be effective for understanding the primary drivers of land value (Friedman et al., 2021).

Land Value Prediction Using Linear Regression.




```
final = pd.DataFrame({'Values': y_test.values,'Predictions': predictions})
final.head(10)
```

	Values	Predictions
0	1077500	369138.624882
1	110000	67573.553627
2	12500	19883.584681
3	592000	180885.082879
4	6000	4535.397035
5	13500	20077.068033
6	15500	20516.802924
7	10500	19461.439185
8	11500	19672.511933
9	15000	14385.458604

OLS Regression Results

Dep. Variable:	SiteValue	R-squared:	0.417
Model:	OLS	Adj. R-squared:	0.413
Method:	Least Squares	F-statistic:	110.2
Date:	Mon, 29 Jul 2024	Prob (F-statistic):	8.80e-88
Time:	23:56:46	Log-Likelihood:	-10189.
No. Observations:	776	AIC:	2.039e+04
Df Residuals:	770	BIC:	2.042e+04
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.43e+06	1.39e+05	10.253	0.000	1.16e+06	1.7e+06
ZoneWardCode	-1.16e+04	1143.787	-10.140	0.000	-1.38e+04	-9353.149
Area	1123.2480	2719.456	0.413	0.680	-4215.179	6461.675
AreaInHa	2.439e+05	1.25e+04	19.514	0.000	2.19e+05	2.68e+05
PropertyUseID	9718.0055	6584.252	1.476	0.140	-3207.208	2.26e+04
MarketCentreID	488.0495	1097.334	0.445	0.657	-1666.071	2642.170

Omnibus:	1051.612	Durbin-Watson:	1.895
Prob(Omnibus):	0.000	Jarque-Bera (JB):	221513.700
Skew:	7.125	Prob(JB):	0.00
Kurtosis:	84.534	Cond. No.	3.87e+03

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.87e+03. This might indicate that there are strong multicollinearity or other numerical problems.

4.4.2 Random Forest

Random Forest is known for its robustness and ability to handle complex datasets (Breiman, 2001). Random forest regression is an ensemble learning method that builds multiple decision trees and merges them to obtain a more accurate prediction. Each tree in the forest considers a random subset of features and data points, reducing the risk of overfitting. The final prediction is the average of the predictions from all the individual trees. This method is particularly useful for capturing complex interactions and non-linear relationships in the data (Breiman, 2001; Liaw & Wiener, 2002). This algorithm that was fit for this problem as many featues added to the lands data could result in non-linear relationships.

Land Value Prediction Using Random Forest.

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import GridSearchCV
from sklearn import metrics
rf = RandomForestRegressor()
param_grid_rf = {
    'n_estimators': [100, 200, 300, 400],
    'max_depth': [5, 10, 15, 20]
}
grid_search_rf = GridSearchCV(rf, param_grid = param_grid_rf, cv = 5)
grid_search_rf.fit(X_train, y_train)
best_model_rf = grid_search_rf.best_estimator_
best_params_rf = grid_search_rf.best_params_
predictions_rf = best_model_rf.predict(X_test)
```

```
[ ] best_params_rf
```

```
{'max_depth': 20, 'n_estimators': 300}
```

```
[ ] metrics.mean_absolute_error(y_test, predictions_rf)
```

```
35578.51745631621
```

```
[ ] metrics.mean_squared_error(y_test, predictions_rf)
```

```
24019408939.93645
```

4.4.3 XGBoost

Effective for capturing non-linear relationships (Chen & Guestrin, 2016). XGBoost (Extreme Gradient Boosting) is an advanced implementation of the gradient boosting framework that has been widely used for both classification and regression tasks. It is particularly powerful and efficient for predictive modelling, including the task of land value estimation. XGBoost improves upon the traditional gradient boosting by incorporating a number of algorithmic optimizations and system enhancements, making it one of the most effective and scalable machine learning algorithms available. XGBoost is based on the gradient boosting framework, where models are built sequentially, and each new model attempts to correct the errors made by the previous models. The key idea is to combine the predictions of multiple weak learners, typically decision trees, to create a strong predictive model.
The objective function in XGBoost includes a training loss function and a regularization term to control the model complexity. This helps to prevent overfitting and enhances the generalization ability of the model.

$$\text{Obj}(\Theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

where l is the loss function that measures the difference between the predicted value $\hat{y}_i$ and the actual value y_i, and Ω is the regularization term for the k-th tree f_k.

XGBoost builds trees in an additive manner. At each step, a new tree is added to the model to minimize the objective function. The predictions are updated as follows:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i)$$

where $f_t(x_i)$ is the prediction of the new tree added at step t.

Land Value Prediction Using XGBoost.

XG Boost

```
from xgboost import XGBRegressor
xgb = XGBRegressor()
param_grid_xg = {'max_depth': [4, 8, 12],
                 'n_estimators': [100, 200, 300],
                 'learning_rate': [0.1, 0.05, 0.01]}
grid_search_xg = GridSearchCV(xgb, param_grid = param_grid_xg, cv = 5)
grid_search_xg.fit(X_train, y_train)
best_model_xg = grid_search_xg.best_estimator_
best_params_xg = grid_search_xg.best_params_
predictions_xg = best_model_xg.predict(X_test)
```

```
[ ] best_params_xg
```

```
{'learning_rate': 0.01, 'max_depth': 12, 'n_estimators': 100}
```

```
[ ] metrics.mean_absolute_error(y_test, predictions_xg)
```

```
45747.9441389485
```

```
[ ] metrics.mean_squared_error(y_test, predictions_xg)
```

```
27560109248.22439
```

4.4.5 Decision Tree

A Decision Tree is a non-parametric supervised learning algorithm used for both classification and regression tasks. It is a powerful tool for predictive modeling, especially in the context of land value estimation. Decision Trees predict the value of a target variable by learning simple decision rules inferred from the data features. The model is intuitive and easy to visualize, making it a popular choice for many machine learning tasks. Decision Trees can capture non-linear relationships between the features and the target variable. This is particularly useful in land value prediction, where interactions between various factors can be complex and non-linear.

$$\Delta Var = Var(N) - \left(\frac{n_L}{n} Var(N_L) + \frac{n_R}{n} Var(N_R)\right)$$

where:

- $Var(N)$ is the variance of the target variable in node N,
- N_L and N_R are the left and right child nodes,
- n_L and n_R are the number of instances in N_L and N_R,
- n is the number of instances in N.

The feature and threshold that maximize ΔVar are selected for the split.

Land Value Prediction Using Decision Trees.

```
from sklearn.tree import DecisionTreeRegressor
dt = DecisionTreeRegressor()
param_grid_dt = {'max_depth': [5, 10, 15, 20, 25]}
grid_search_dt = GridSearchCV(dt, param_grid = param_grid_dt, cv = 5)
grid_search_dt.fit(X_train, y_train)
best_model_dt = grid_search_dt.best_estimator_
best_params_dt = grid_search_dt.best_params_
predictions_dt = best_model_dt.predict(X_test)
```

```
[ ] best_params_dt
```

```
{'max_depth': 20}
```

```
[ ] metrics.mean_absolute_error(y_test, predictions_dt)
```

```
31994.849785407725
```

```
[ ] metrics.mean_squared_error(y_test, predictions_dt)
```

```
21446845665.23605
```

```
[ ] dt_score = metrics.r2_score(y_test, predictions_dt)
dt_score
```

```
0.5121761399713817
```

Using the machine learning algorithms, the collected land values data from Nairobi City County were analyzed using Regression Model, Random Forest, Decision Trees and XGBoost algorithms to predict the land values given new data.
Predicted Values Given Test Data with no Land Value.

```
pred_df
```

	SiteValue	CustomerSupplierID
0	540000.0	220300901
1	107000.0	220301782
2	11000.0	220300081
3	336500.0	200600210
4	19500.0	220300397
...	...	...
228	15500.0	220200830
229	13500.0	220201624
230	19000.0	220301611
231	10000.0	160700222
232	2000.0	220301001

233 rows × 2 columns

4.5 Model Evaluation and Comparison

Model Comparison

```
if (rf_score > dt_score) and (rf_score > xg_score):
    best_model = best_model_rf
    best_params = best_params_rf
elif (dt_score > rf_score) and (dt_score > xg_score):
    best_model = best_model_dt
    best_params = best_params_dt
else:
    best_model = best_model_xg
    best_params = best_params_xg

best_model
```

```
DecisionTreeRegressor
DecisionTreeRegressor(max_depth=20)
```

Based on comparison and evaluation of the algorithms, best algorithm for land prediction was Decision Tree. Evaluate models based on metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R^2) to ensure accurate and reliable predictions.
Mean Squared Error: 25888604364.588367
Random Forest Mean Squared Error: 23683876794.95708
Linear Regression Cross-Validation Mean Squared Error: 103087976439.15225
Random Forest Cross-Validation Mean Squared Error: 17164096083.580097

4.6 Ethical Considerations:

Ensured compliance with data privacy regulations GDPR and Data Protection Act of 2019 when handling sensitive land parcels information. Anonymize data where necessary to protect individuals' privacy and remove personal identifiable data. Clearly communicated the limitations and uncertainties associated with the predictive model. All stakeholders were provided with transparent insights into the model's decision-making process.

5.0 RESULTS AND DISCUSSION

This study has reviewed the application of machine learning algorithms in land value prediction and analysis of studied work by previous researchers, there are several existing machine learning algorithms that have been applied in this field. This chapter presents the conclusion, Contributions of the study, communicate the recommendations and suggest areas of further research. The first section provides a summary of the research findings including the achievements accomplished by conducting this study. The second section of this chapter outlines the recommendations and conclusion. The aim is to prove that the suggested recommendations and conclusion are logically derived from the analysis of the findings.

5.1 Theoretical Contributions of the study.

The study has been able to identify factors that affects land value and variation of value based on this features. This will help inform investors and government authorities on future land value. These factors included the location, land use, topology, soil type, development in the area. From the data extracted on the responses to questionnaire, it is evident from the responses that location is the greatest factor in land valuation and investment as evidenced by 34% of the respondents while soil type factor had the least influence with 15% of the respondents. This was majorly because Nairobi is a City and respondents didn't care much of the soil type but location which comes with amenities in the area and infrastructure.

5.2 Empirical contributions of the study.

The study has shown that using machine learning techniques, it is possible to develop a predictive model that can help understand land valuation. The machine learning algorithms developed will be of great importance to land administrators, investors and many other sector players. This research contributes to the advancement of ML based approaches in land valuation and address the evolving needs of the real estate industry and urban planning domain in Kenya.

The model developed has factored in five variables that directly influences land value thus making the model as conclusive as possible. The outcome of this study can be used in policy revisions of guideline value of land which may add more revenue to the County Governments while land transaction is made as well as make land investors make informed decisions using land value calculated.

This study will support the land assessors to relook the movement of the identified factors to have control on rise in the land price and stabilize it. Since there is a greater need for good long term data analysis about land prices and values, general land market behavior and spatial development, the results produced in this research may be of great use for Government and private agencies involved in land business and administration.

Private entities and individuals can also use this technique to plan for the future of their land transactions and investment. In conclusion based on the findings of this paper machine learning algorithms can be used to predict the land value in Nairobi given enough historical data collected over a period of time. This confirms our hypothesis to be true. Moreover, the value predicted by the model and the real values of the historical data, in the order of the latter, demonstrate the adequate adjustment of the model with all predictions lying on a 90% confidence interval and thus works well in predicting land value.

5.3 Limitations of the study

Obtaining comprehensive set of data from land agencies is difficult as such, most of them are uncooperative and reluctant to share terming the information as personal and confidential. Moreover, obtaining secondary data from other sector players is also a hard task since they consider confidential and thus should be protected due data protection laws in place.

5.4 Recommendations for Future Research

The current study was limited to Nairobi Metropolitan Area. This study recommends that future studies focus on other regions since factors affecting land values may vary depending on different localities and land use in the area.

The major challenge is availability of land data in a single repository. There is need for collection of more data within the Nairobi Metropolitan Area with additional features, this will improve the accuracy of the models. There is need to perform more feature selection and study also to see how to incorporate more features and technology like the GIS.
Future work include to explore additional data sources from Government and Land agencies.

More advanced machine learning algorithms can also be applied, and innovative feature engineering techniques to enhance the predictive model's accuracy and robustness. This should include the data that was unavailable for use like the soil type, topology, development category, and other features. Additionally, to be able to have land records transactions in digital format. Most of this data is still under digitization as it still manual records in valuation books. Some key feature details of the land parcels were in nearly feasible paper title copies.

Investigate the potential for applying the predictive model to other domains, such as commercial real estate, agricultural land valuation, or urban planning. Conduct studies to assess the real-world impact of the predictive model on decision-making processes in real estate, urban planning, and investment.

6.0 REFERENCES

1. Li, A., Zhang, B., & Wang, C. (2018): Journal of Real Estate Economics, 10(2), 123-145.
2. Smith, D., Johnson, E., & Brown, M. (2020): Proceedings of the International Conference on Machine Learning, 78-89.
3. Zhang, C., Li, D., & Liu, E. (2019). Journal of Geographic Information Systems, 15(3), 210-230.
4. Dellstad, M. (2018). Comparing Three Machine Learning Algorithms in the task of Appraising Commercial Real Estate. KTH Royal Institute of Technology.
5. Di, N. F. M., Satari, S. Z., & Zakaria, R. (2017). Real estate value prediction using multivariate regression models
6. Gu, G., & Xu, B. (2017). Housing Market Hedonic Price Study Based on Boosting Regression Tree. Advanced Computational Intelligence and Informatics, 21(6).
7. Hannonen, M. (2008). Predicting Urban Land values: A comparison of four approaches. International Journal of Strategic Property Management, 12(4), 217-236. https://doi.org/10.3846/1648-715X.2008.12.217-236
8. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.
9. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.
10. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-4774.
11. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144.
12. Zhou, Z., Liu, Y., & Wang, J. (2021). Application of XGBoost in Real Estate Price Prediction. *Journal of Real Estate Finance and Economics*, 50(2), 200-220.