

Mass Valuation of Unimproved Land Value Case Study: Nairobi County

Edwin Kochulem¹, Dennis Mwaniki², Felix Mutua¹

¹Department of Geomatic Engineering and Geospatial Information System (GEGIS), Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya

²Data and Analytics Section, UN-Habitat, Nairobi, Kenya

Email: edwincorir@gmail.com

How to cite this paper: Kochulem, E., Mwaniki, D. and Mutua, F. (2023) Mass Valuation of Unimproved Land Value Case Study: Nairobi County. *Journal of Geographic Information System*, **15**, 122-139. <https://doi.org/10.4236/jgis.2023.151008>

Received: November 18, 2022

Accepted: February 25, 2023

Published: February 28, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The purpose of this study is to investigate mass valuation of unimproved land value using machine learning techniques. The study was conducted in Nairobi County. It is one of the 47 Kenyan Counties under the 2010 constitution. A total of 1440 geocoded data points containing the market selling price of vacant land in Nairobi were web scraped from major property listing websites. These data points were adopted as dependent variables given as unit price of vacant land per square meter. The Covariates used in this study were categorized into Accessibility, Environmental, Physical and Socio-Economic Factors. Due to multi-collinearity problem present in the covariates, PLS and PCA methods were adopted to transform the observed features using a set of vectors. These methods resulted in an uncorrelated set of components that were used in training machine learning algorithms. The dependent variable and uncorrelated components derived feature reduction methods were used as training data for training different machine learning regression models namely; Random forest, support vector regression and extreme gradient boosting regression (XGboost regression). PLS performed better than PCA because the former maximizes the covariance between dependent and independent variables while the latter maximizes variance between the independent variables only and ignores the relationship between predictors and response. The first 9 components were identified as significant both by PLS and PCA methods. The spatial distribution of vacant land value within Nairobi County was consistent for all the three machine learning models. It was also noted that the land value pattern was higher in the central business district and the pattern spread northwards and westwards relative to the CBD. A relative low vacant land value pattern was observed on the eastern side of the county and also at the extreme periphery of Nairobi County boundary. From the accuracy metrics of R-squared and MAPE, Random Forest Regression

model performed better than XGBoost and SVR models. This confirms the capability of random forest model to predict valid estimates of vacant land value for purposes of property taxation in Nairobi County.

Keywords

Machine Learning, Property Valuation, GIS, PLS, PCA

1. Introduction

The valuation for rating Act 1984 provides the framework that governs the imposition of rates on vacant land in Kenya. According to this Act, there is a need to conduct property reevaluations every 10 years. However, based on existing literature, the last re-evaluation in Kenya was conducted in 1982, showing existing gaps in the property valuation exercise. Therefore, Nairobi relies on an outdated valuation roll whose values have no relation to the current market values. Additionally, Chapter 267 of the Rating Act 1986 provides various options for property tax bases which can be adopted by the local authorities such as agricultural land, unimproved land rating, unimproved land value rating plus improvement rating. According to [1], it shall be the duty of the rating authority in adopting any method or methods of rating under this Act. This is done to ensure that the costs of the rating authority's general expenses, and of the general expenses of every local council in whose area the rating authority levies a rate are distributed equitably over all parts of the respective areas of the rating authority. The act also states that the Minister may give such directions to the rating authority as he considers necessary for the purpose of obtaining equitable distribution as aforesaid and any such authority shall comply therewith [2].

Lack of trained valuers, proper incentives and conflicting approaches adopted by valuers are some of the negative consequences of individual "single parcel" valuation processes. Most Kenyan valuers allocate their time to non-rating valuation activities owing to the relative lack of remuneration for doing valuation rolls. All these challenges, coupled with the lack of any form of mass valuation make the valuation process inadequate, inefficient and costly in terms of time and resources [3].

Over the past two decades, empirical experiments have been done to model the value of real estate properties on a geographical scale. Geographical information systems (GIS) integrated with spatial statistics tools have given the opportunity to estimate land prices at unknown locations and generate accurate land variability maps. The hedonic modelling techniques, such as ordinary least square regression, have gained momentum recently with wider adoption of GIS tools and availability of big data. This has given rise to the development of geostatistical models such as geographical weighted regression and regression-kriging which account for spatial heterogeneity in land spatial variability [4]. Recent developments in machine learning algorithms have also provided the opportunity to train a variety of machine learning models for estimating variations in land

value. These machine learning algorithms include random forest (RF), support vector machine (SVM), extreme gradient boosting, artificial neural network (ANN) and decision tree algorithm. Studies have shown that random forest models are robust and generate more accurate estimates of real estate prices [5]. This was achieved by the ability of random forest model to handle non-linear relationships by relying on decision trees to estimate values of real estate. [6] highlighted the need to have a large amount of training data as one of the major challenges of ANN model.

The purpose of this study is to provide an analytical approach to provide solutions to vacant land valuation problem. This is achieved by developing a mass valuation model that utilizes a set of variables that were identified in recent studies within Nairobi County. The mass valuation model is based on machine learning and GIS techniques to assess several characteristics simultaneously using comparable land attributes and selling prices within the same neighborhood.

2. Materials and Methods

2.1. Study Area

The study was conducted in Nairobi County (**Figure 1**). Nairobi County is one of the 47 Kenyan Counties under the 2010 constitution. It borders Kiambu County to the North and West, Machakos to the East, and Kajiado to the South. The county is home to Kenya's Capital City, with a population of 4.397 million as of 2019 [7]. The County Government of Nairobi collects property tax through the county's finance department. Property tax contributes to about 25% of the total own-source revenue [8].

The datasets used in this study were categorized into Dependent Variables, Accessibility Factors, Environmental Factors, Physical Factors and Socio-Economic Factors as shown in **Table 1**.

Table 1. Datasets.

Accessibility	Environmental	Physical	Socio-economic
Distance to tourist attraction sites	Distance to Open lakes and reservoirs	Distance to CBD	Parcel Area
Distance to Super-markets and Malls	Distance to market centers	Distance to Bus stops	Population density 2019
Distance to rivers, streams and water-ways	Distance to informal settlements	Distance to Road intersections	Nighttime light In-tensity
Distance to the piped water network	Distance to Industrial sites	Distance to Rail stops	Built up density
Distance to Sewer Network	Distance to health centers	Distance to major roads	Land use
Distance to Protected areas	Distance to open spaces	Distance to minor roads	Poverty Index
Distance to Police station	Distance to forests	Slope	
Distance to Outdoor leisure sites	Distance to University & Colleges	Elevation	
Distance to schools			

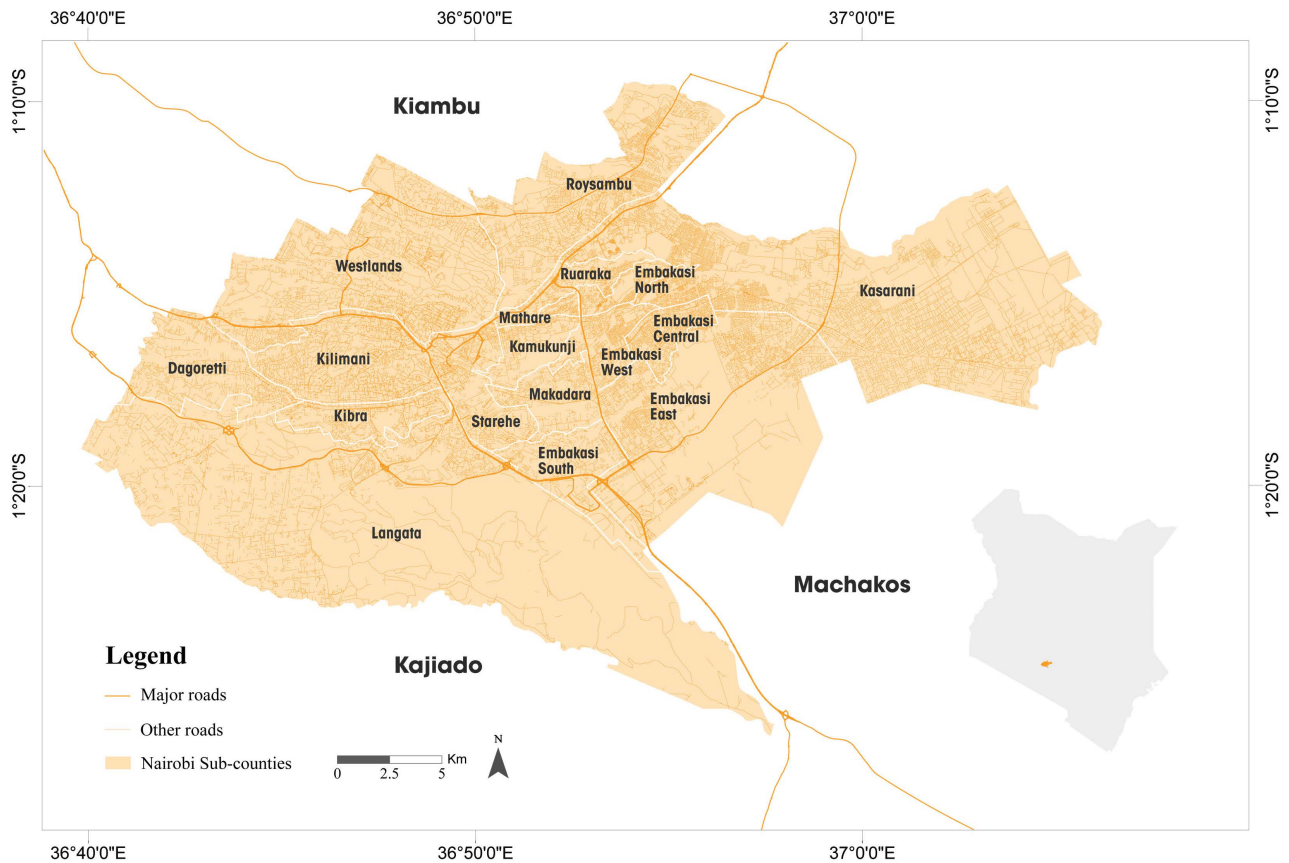


Figure 1. Study area.

Accessibility factors such as informal settlements, industrial sites, Nairobi Land Use Data, Roads, Sewer, and water data were obtained from OpenStreet-Map. Distances to points and features were obtained through the calculation of Euclidean distances.

All the environmental factors were obtained by calculating Euclidean distances to the features obtained from OpenStreetMap.

Street block shape data were obtained from Google. Elevation data was obtained from the ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) Digital Elevation Model (DEM). The slope was computed from the elevation data.

Population data were obtained from Global Human Settlement (GHS) data. Nighttime light (NTL) data was obtained from MODIS. Land use was obtained from the Ministry of Planning, Nairobi. Built-up data was obtained from Google buildings.

2.2. Methods

Significant factors refer to the most relevant variables that can be used in a prediction task. The main goal of the feature importance measure is to provide an interpretable index by identifying allowing variables that are significantly important in the prediction of the response variable (**Figure 2**).

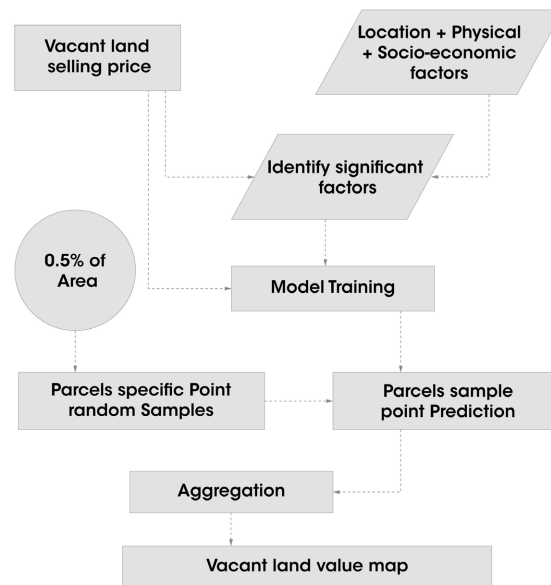


Figure 2. Methodology flow chart.

Vacant land selling prices of 1440 properties posted to the real property websites were web scraped. These websites include Buy-rent Kenya, property 24 and Jiji. The specific parcel details such as parcel selling price, parcel area, location address, land use category and the date posted were extracted.

The parcels were geocoded using google geocoding API to link the parcel address to specific XY coordinates. The location, physical and socio-economic details were then appended to each parcel based on their XY location obtained. All the parcel areas were converted to square meters for uniformity purposes. The value of the parcel was expressed as the total selling price of vacant land divided by the total parcel area which gives value in Kenya shillings per square meter (Ksh psm).

Training data here comprise the dependent variable given by the vacant land value expressed in psm (shillings in square meters) and the independent variables or covariates expressed by the physical, socio-economic, location and accessibility parcel-based factors. Outlier analysis was then conducted to remove outliers in the training data represented by too high or too low psm values contrary to other parcel values in the same neighborhood.

The dependent variable (selling price per square meter) was normalized using natural log transformer. Normalization of the dependent variable was done to meet the requirement of many machine-learning methods. The continuous independent variables (X variables) were standardized using standard scaler.

The categorical independent variables (X variables) such as land use were hot-encoded and represented as a binary set of 1 s and 0 s. The temporal characteristics indicator was given by the country's annual inflation rate depending on the posting date of the parcel details indicated on the property listing websites. This will help to account for the dynamic changes in the market values. The wards (Nairobi wards) boundaries were hot-encoded; this provided the geo-

graphical aspects of the properties.

Three methods were adopted to identify the main drivers of vacant land value variation in Nairobi are:

- Ordinary least square (OLS)
- Partial least square (PLS)
- Principal component analysis (PCA)

Ordinary Least Squares is a regression method that finds the coefficients of linear regression equations. It describes the relationship between a dependent variable and one or more independent quantitative variables. Ordinary Least Squares is a feature selection algorithm that finds the line of best fit for a dataset by minimizing the sum of the residuals. Residuals are usually squared to avoid negative distances.

The OLS method is used to identify factors that are significantly statistically based on their corresponding p-values. Factors were chosen that have a p-value less than 0.05 (5% significance level). **Table 2** provides a list of factors deemed to be significant.

Partial Least Squares (PLS) is a supervised feature reduction and selection tool that uses guided transformation approach. It maximizes the covariance between response and predictor variables by projecting both dependent and independent variables to a lower dimensional space. PLS further identifies the direction within the newly transformed X-space that explains maximum information or variance in the Y-transformed space [9].

Table 2. OLS variables.

Variables	p Value
(Intercept)	0.00
Parcel Area	0.00
Population density 2019	0.04
Land Use	0.02
Distance to rivers	0.05
Nighttime light intensity	0.00
Distance to tourist attraction sites	0.04
Distance to the sewer network	0.05
Distance to university and colleges	0.03
Distance to the police station	0.00
Distance to outdoor recreational sites	0.05
Distance to market centers	0.03
Distance to major roads	0.03
Distance to industrial sites	0.02
Distance to CBD	0.00
Distance to bus stops	0.02

PCA is a traditional multivariate statistical method commonly used to reduce the number of predictive variables and solve the multi-collinearity problem [10]. Principle component analysis identifies a few linear combinations of variables that can be generalized and summarized while keeping as much information about the variables. It is a maximization function applied on the variance in predictor variables only.

The accuracy statistics of PLS and PCA were computed and compared using k-fold cross-validation procedure.

Three methods were used to train and validate models used to estimate unimproved land.

- Random Forest regression (RF)
- Extreme gradient boosting regression (XGboost regression)
- Support vector regression (SVR)

Once the ML algorithms were trained and validated, the pre-trained models were applied in the mass valuation of all available parcels in Nairobi transformed parcel details.

Parcels were represented as polygons while prediction of land prices was conducted at the point level. Therefore, methods of data disaggregation from the polygon level to the point level were determined. Random samples were generated in each parcel and the count of samples in each parcel determined by the parcel area. This can be expressed by the equation;

Count of sample points in each parcel is given by 0.5% of the total parcel area. A parcel with an area of 450 m² will give:

$$\text{Count of Sample points} = \left(\frac{0.5}{100} \right) * 450 = 2.25 \text{ approximately } 2 \text{ points}$$

Therefore, larger parcels were given a higher number of random sample points while smaller parcels were given fewer random points. The independent variables for each parcel at each random sample point were obtained. The independent variables were then transformed using the same variable transformers applied during pre-processing of the training data; *i.e.* standardization, hot-encoding and PCA/PLS transformation steps. The pre-trained ML models were then applied to the transformed parcel-based random sample points prediction data. The output of this stage was vacant land value in Kenya shilling per square meter (psm) at each random sample point.

To obtain the representative value of the vacant parcel, the mean value of the predicted values at each of the random sample points predicted for each parcel was computed.

For instance, the 450 m² parcel with two representative random sample points, the mean value of the predicted land values obtained at each of the two random sample locations were computed.

$$\text{Aggregate parcel value estimate} = \sum_1^n y_{\text{estimate}} ;$$

where n is the number of sample points for each parcel.

3. Results

It was noted that the majority of these independent variables were correlated with each other. Variables that have a VIF value of more than 5 are deemed to be correlated with another independent variable. This is given by their Variance inflation factor (VIF) values as shown in **Figure 3**.

Figure 3 shows factors with VIF score of less than 5. The factors include parcel area, slope, distance to minor roads, births rate, distance to junior schools, distance to waterways, distance to major roads, built-up densities, night time intensities, distance to open lakes and reservoirs, distance to supermarkets, distance to road intersections, to bus stops, protected areas, poverty index, railway line and health centers.

The remaining factors generated VIF values of more than 5 thus rendering the identification of factors invalid for subsequent use in model prediction.

PCA is a method that generalizes the most important factors, by observing the flattening of the cumulative variance curve. From **Figure 4**, principal component 1 to principle component 9 were chosen to represent significant variables. The cumulative variation graph flattens at approximately PC9 with cumulative variance of 87%. Therefore, the inclusions of additional transformed components beyond PC9 are considered to contribute insignificant value.

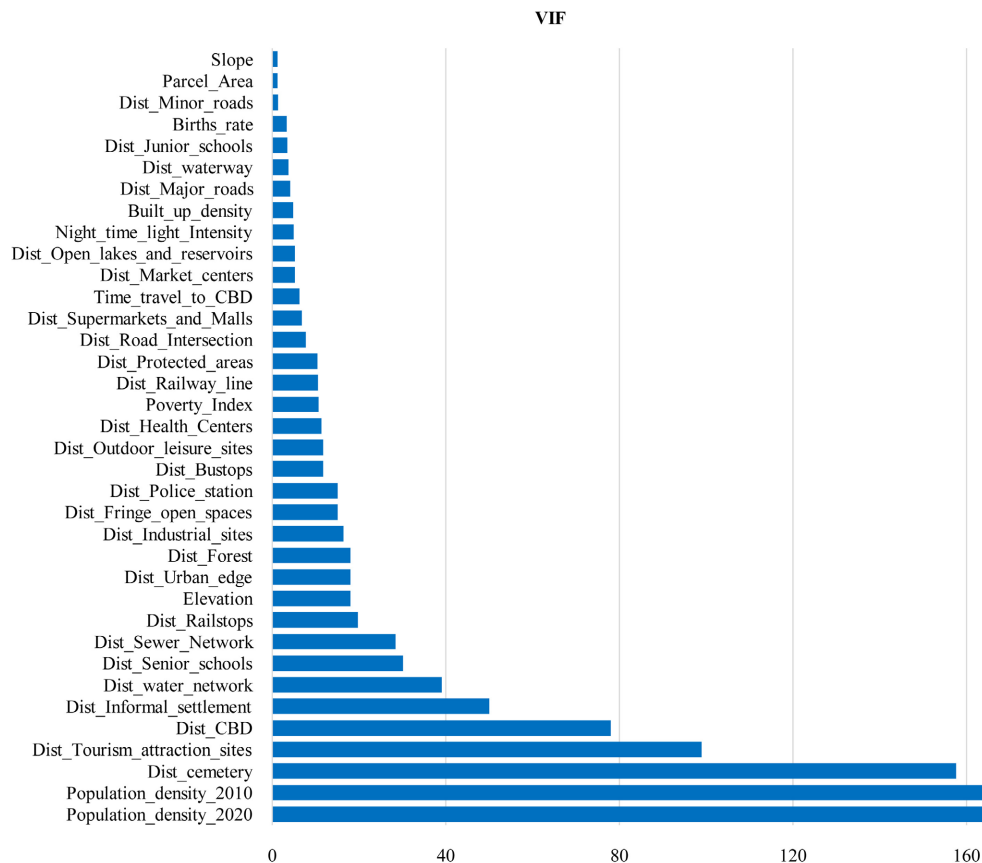


Figure 3. Variance Inflation Factor (VIF) values for the predictor variables.

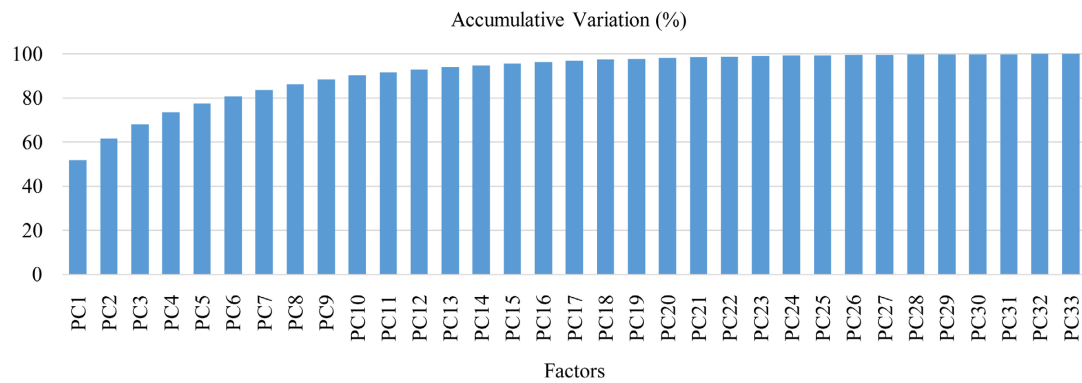


Figure 4. PCA variation in predictor variables.

From **Figure 5**, it can be noted that the significant component variables are identified from component 1 to component 9. These variables depict increasing trend in the cumulative percentage variance and flattens at approximately 70% and 80% respectively for the variance in response and factor variables approximately.

Figure 6 shows the MSE trends for the latent variables derived by PLS and PCA transformation. It is noted that PLS consistently performed better than PCA at all levels shown by the PLS trend line constantly below the PCA trend line. Moreover, the trend in MSE values is consistent for both PLS and PCA derived variables with a dip in MSE values depicted at PC9 which flattens thereafter.

Table 3 shows that PLS-based transformed variables resulted in better results than PCA-based variables. Additionally, RF model performed better than SVR and XGBoost model both in PLS and PCA transformed latent variables. The R-squared of the random forest was 0.79 while SVR had a value of 0.77 and finally XGBoost with a value of 0.73 for PLS reduced variables. On the other hand, The R-squared of the random forest was 0.77 while SVR had a value of 0.75 and finally XGBoost with a value of 0.731 for PCA derived variables.

From **Figure 7**, it is noted that the modeled land value and the actual land value depict consistent variations in land value both spatially and statistically. The actual land value shows vacant land value ranging from 818 to 109,821 psm. The modeled land value shows land values ranging between 1164 to 105,084 psm. This is in relation to the factors that were selected in modelling the land value based on random forest algorithm.

Figure 8 shows the vacant land value in Nairobi County expressed in Kenya shillings (Ksh) per square meter with the independent variables transformed using the PLS technique. The figures show the results that were depicted by the three methods using the PLS technique. Random Forest Regression generated more accurate output compared to XGBoost and SVR models.

Figure 9 shows the vacant land value in Nairobi County expressed in Kenya shillings with the independent variables transformed using PCA technique. The figures show the results that were obtained by the three methods using PCA

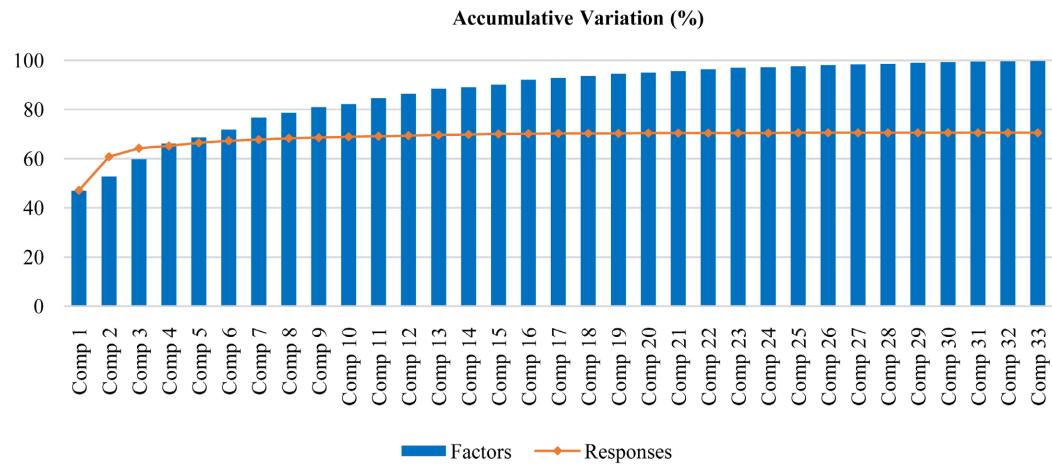


Figure 5. PLS variation in response and predictor variables.

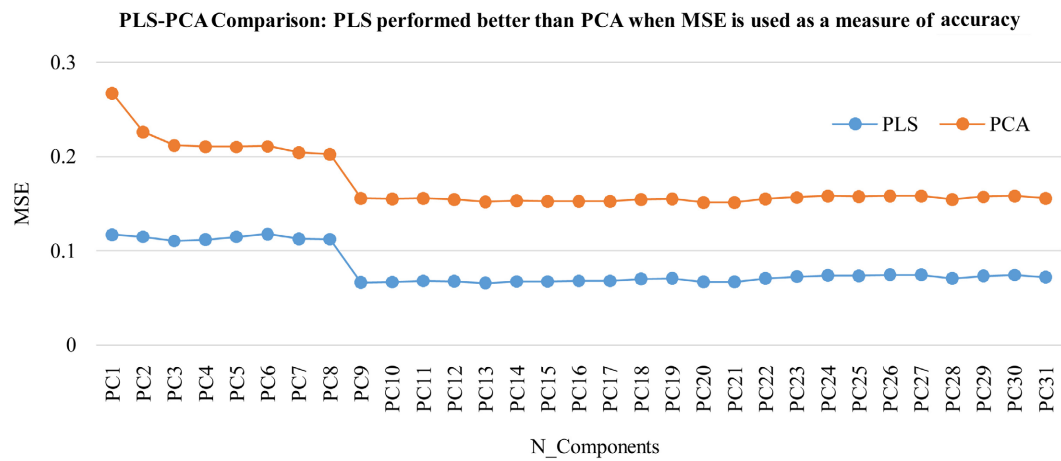


Figure 6. Mean square error statistics for the latent variables derived by PLS and PCA methods.

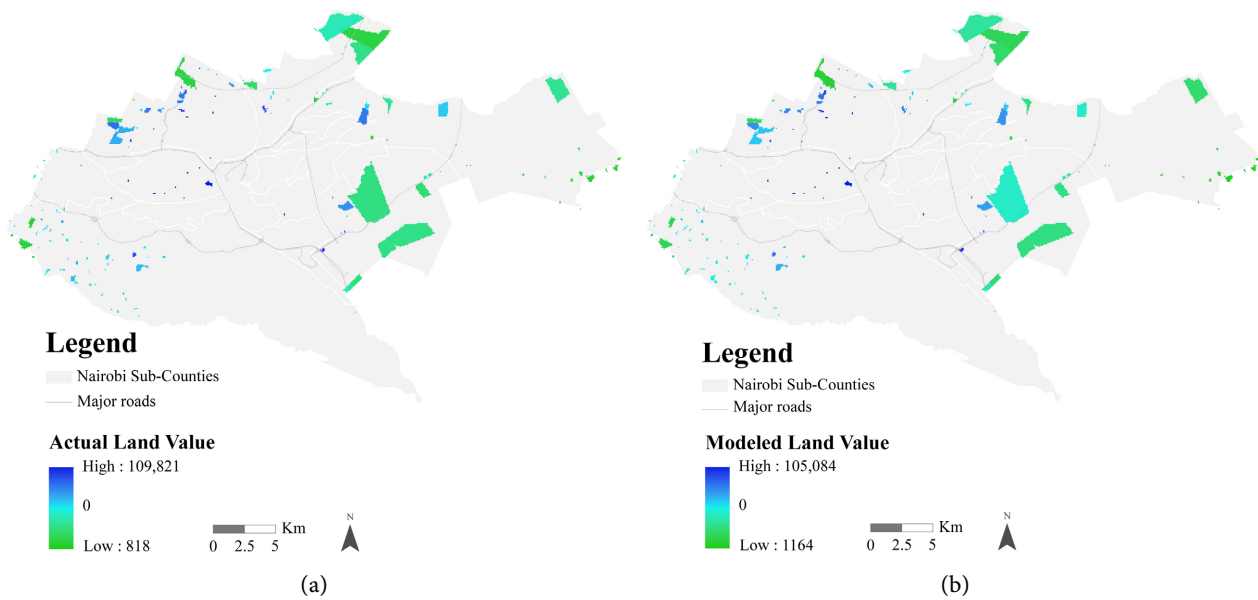
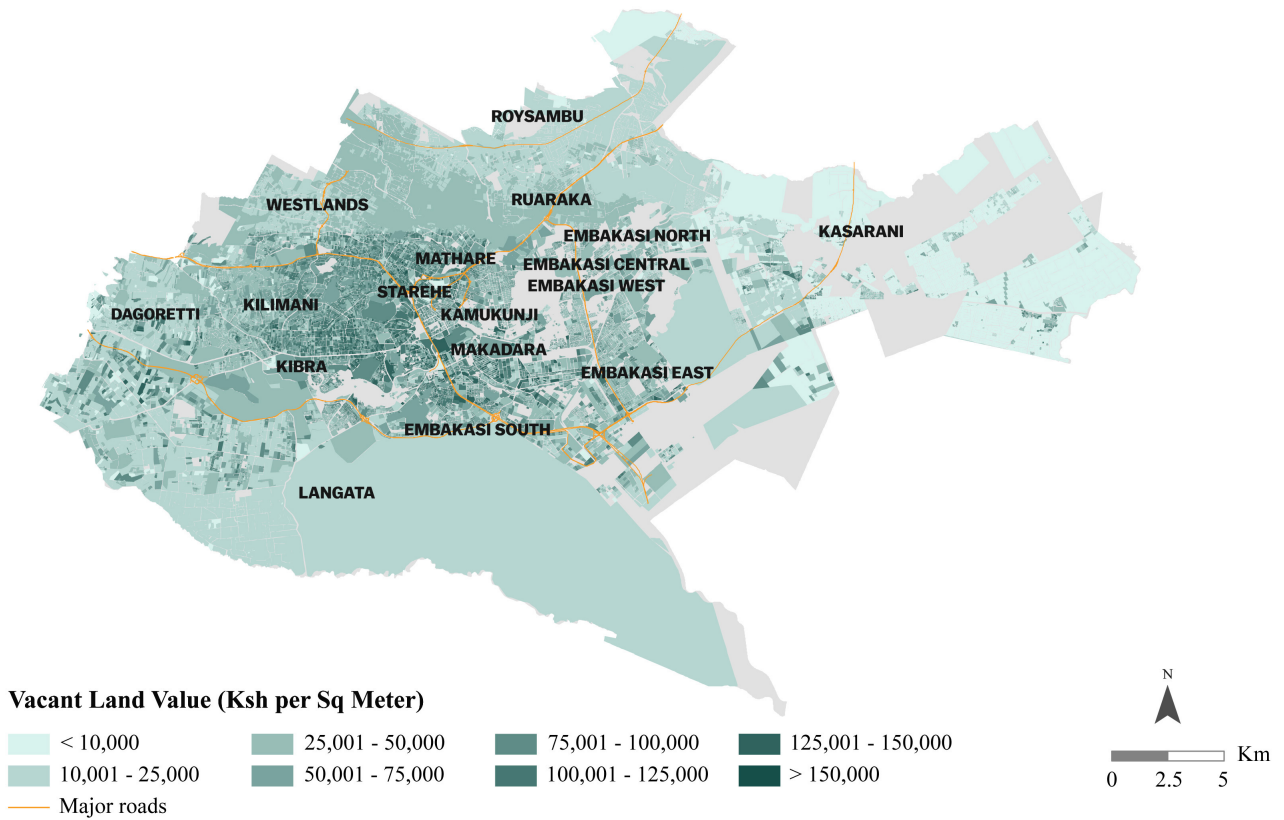
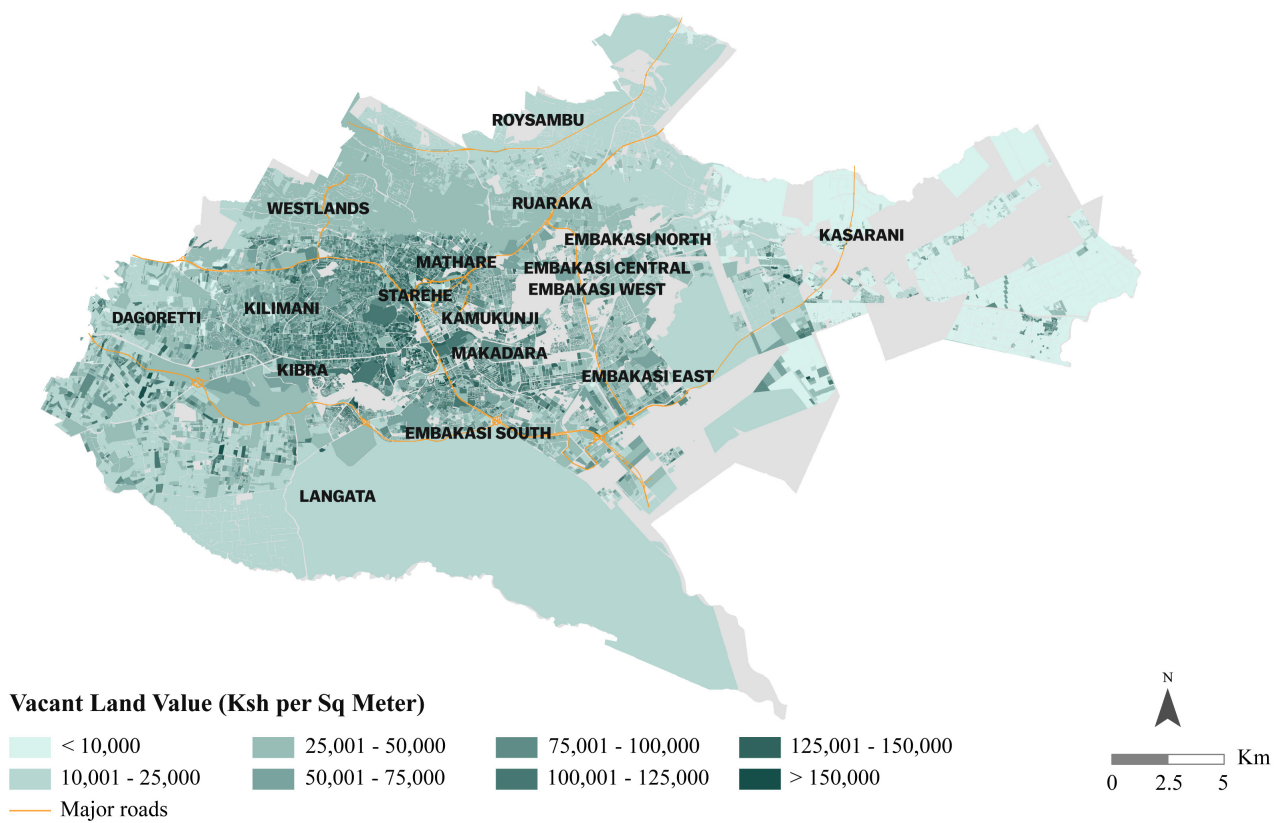


Figure 7. (a) Actual land value at sampled parcels; (b) Random forest modeled land value at sampled parcels.



(a)



(b)

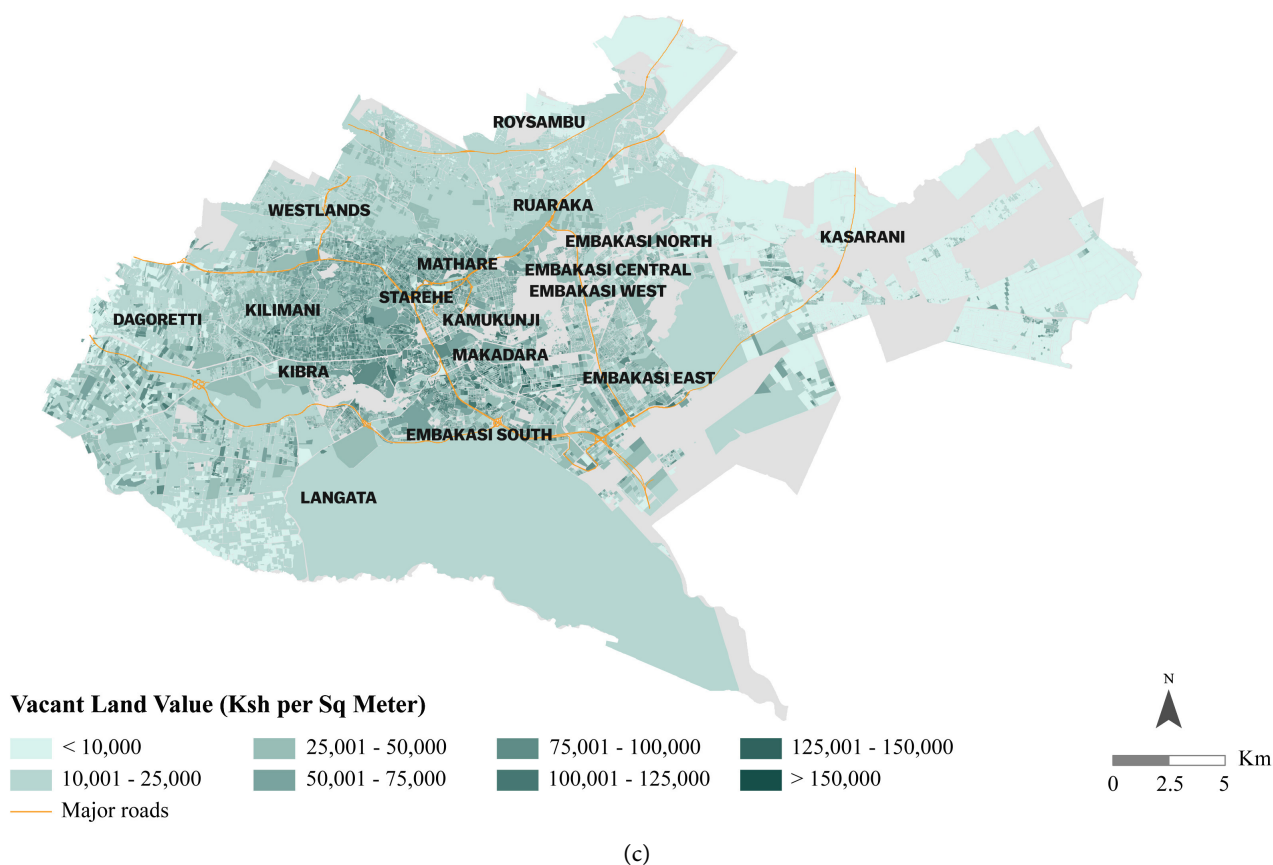


Figure 8. (a) Random Forest Regression model on variables transformed using PLS; (b) Support Vector Regression model on variables transformed using PLS; (c) XGBoost Regression model on variables transformed using PLS.

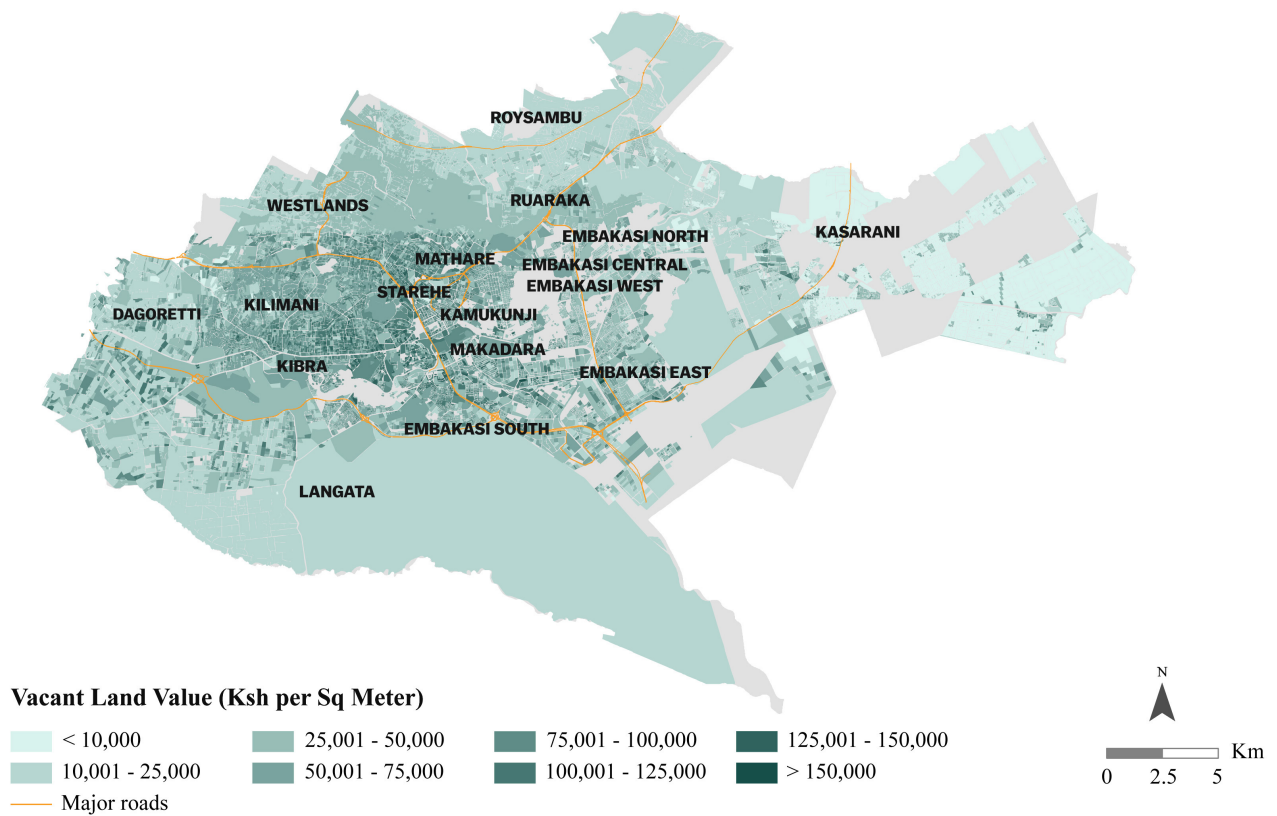
Table 3. Comparison of accuracy statistics by random forest model, Support vector regression model and XGboost regression model on PLS and PCA derived latent predictor variables.

	RF Model PLS Components	SVR Model PLS Components	XGBoost Model PLS Components	RF Model PCA Components	SVR Model PCA Components	XGBoost PCA Components
R-Squared	0.79	0.77	0.73	0.77	0.75	0.71
MAPE (%)	4.25	4.38	4.75	4.43	4.53	5.13

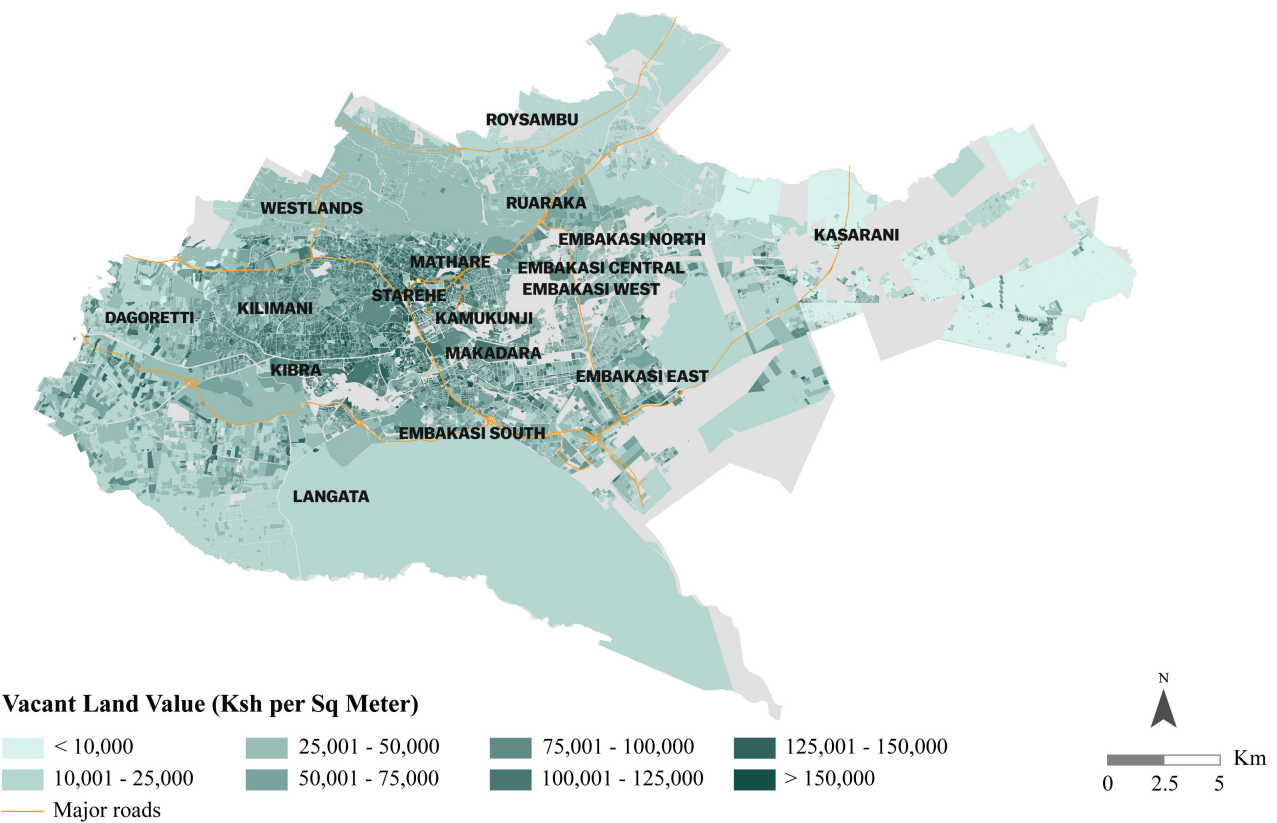
transformed input variables. Random Forest Regression generated more accurate output compared to XGBoost and SVR models similar to PLS transformed input variables.

In both scenarios, the spatial distribution of vacant land in Nairobi is generally consistent regardless of the variable reduction technique and machine learning model chosen.

Land value is seen to increase towards the central part of the county within the Central business district (CBD) and spreads westwards. These zones with high vacant land value have the constitutes important variables derived using PLS and PCA techniques. Relatively, low vacant land values are seen on the eastern side of the county.



(a)



(b)

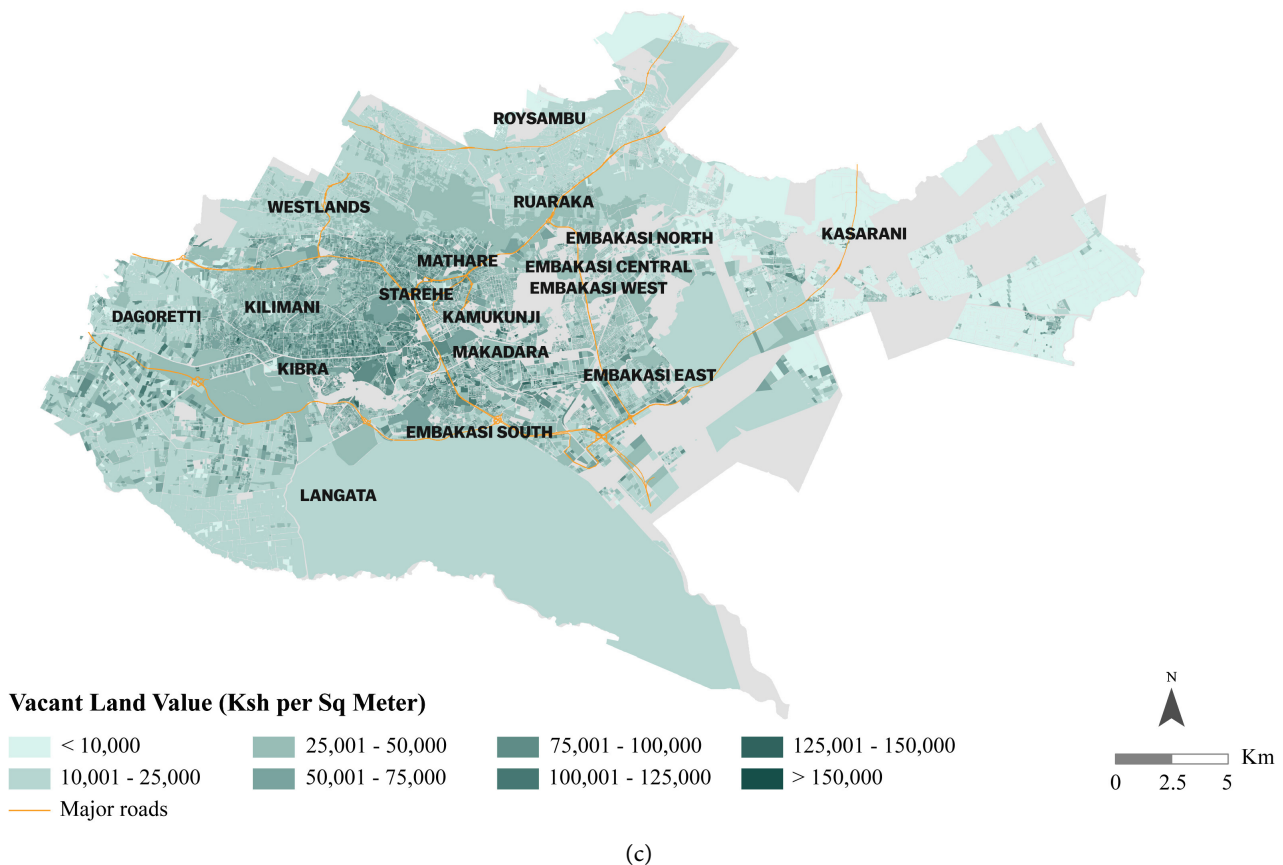


Figure 9. (a) Random Forest Regression model on variables transformed using PCA; (b) Support Vector Regression model on variables transformed using PCA; (c) XGBoost Regression model on variables transformed using PCA.

4. Discussion

During property valuation process, the identification of important environmental, accessibility, physical social and economic factors that form baseline for comparability analysis of similar properties, is vital. This is usually done by valuation surveyors. Different methods have been used to come up with the mass valuation of land in this study. First, there are factors that were considered to depict the value of the land. These factors are accessibility factors, environmental, physical and socio-economic factors. From **Figure 3**, OLS method suggests that the majority of these independent variables were correlated with each other through the VIF values obtained. Multi-collinearity introduces redundancy which can lead to high variability in the estimates of beta coefficients. [11] suggested that Variance Inflation Factor (VIF) of more than 5 results to unsuitable regression model by inflating the regression coefficient parameters. [12] discussed that multi-collinearity can be minimized by transforming the original variable using factors that results to uncorrelated set of variables that can be adopted in modelling. Moreover, multi-collinearity problem can be solved by increasing the sample size of the variables. This will give a wider range of the expected selection hence reducing the multi-collinearity of the samples. Usually adding data though is not feasible. In this study, partial least square and princip-

al component analysis techniques were used for handling multi-collinearity problem.

In both techniques as shown by **Figure 3** and **Figure 4**, transformed components 1, 2, 3, 4, 5, 6, 7, 8 and 9 were identified as significant latent variables (or components) for land valuation. The identification of significant latent variables was done by observing the trends of cumulative variation explained by the derived latent variables. The cut-off point was selected where flattening trend was noted in the cumulative variation explained by each latent variable. The optimum number of derived latent variables was also identified by the components that yield minimum predicted mean squared error by cross-validation. **Figure 5** compares mean square error (MSE) trend obtained by PLS and PCA for each of the derived variables and it was noted that PLS performed better than PCA. This is shown by relatively lower MSE values for the PLS as compared to PCA. The dip at latent variable or component 9 (PC9) indicates the transition point that yields the lowest MSE value for both PLS and PCA. Therefore, additional components or variables do not improve the MSE score in any way, rather it only introduces noise in the datasets. The latent variables 1 - 9 models the intrinsic structure of the original variables while the remaining variables named 10 - 31 models the noise component in the original variables. This was expected outcome because PLS uses supervised transformation technique where latent variables (or components) are derived by maximizing the covariance between the independent variables and dependent variable [13]. Therefore, PLS method has the ability to handle variables with highly correlated variables with complex distributions and yield outputs that best predicts the dependent variable. On the contrary, PCA is unsupervised reduction technique that ignores the relationship between dependent and independent variables. Therefore, the PCA can easily overlook independent variables that are not significant in generating the latent variables (or components) but have significant prediction power on the dependent variable [14].

The price of vacant land value is determined by the market demand and supply of land in different locations of the county. High demand than supply drive land value up while low demand than supply drive the land value down. The spatial distribution of vacant land values in Nairobi is consistent as depicted by land value distribution maps. High values are noted at the central area of the CBD and spread towards the Western side of the county such as Hurlingham, Loresho, Lavington and Westlands. This was expected to occur because areas near the CBD are areas that were considered to have the most important variables during PLS and PCA feature reduction process. High vacant land value is mainly driven by the proximity to CBD where the land provides maximum return through commercial activities. In addition to the closeness to CBD, these areas are less populated which indicates that there are fewer users using these areas either for residential, industrial or commercial purposes resulting in high unimproved land value. These areas also are characterized by large parcel area and well planned land use compared to other locations within Nairobi county.

Additionally, Government investment on road infrastructure and amenities such as recently constructed Nairobi expressway in 2021 has also contributed to high land prices in western and central part of the county.

On the other hand, relatively low vacant land values are shown in the Eastern section and near the periphery of the county which include Eastern part of Kasarani constituency. This is linked to the increase in separating distance between these locations and the CBD. Heavy traffic congestion during peak hours is normal occurrence in these regions. Therefore, the commuting time from these locations to and from CBD for work and other social economic reasons is also high. Uncontrolled subdivision and shortage of critical services such as waste management and good roads in these areas have also played a role in reducing land demand in this region thereby resulting in low land values.

Sustainable urban policy establishes the need for urban areas to meet the social, economic functions, political and environmental protection needs of the urban ecosystem. Sustainable urbanization is achieved when the urban growth and expansion is less sprawling but rather denser and compact in nature. In this manner, services and government infrastructure development can be utilized efficiently within the defined urban area with limited wastage of resources and pollution. Sustainable urban growth facilitates walkability between work place, social place, to school without the need to use motor-able means of transport. Therefore, city planners have always focused in promoting sustainable urban growth in Nairobi by focusing the service provision and infrastructure development within the planned area of the city. These areas form major share of the region with high vacant land values. Therefore, the planned zones have relatively higher vacant land value compared to the unplanned zones of the city.

5. Conclusions

In conclusion, the study explores the value of unimproved land valuation in Nairobi, and then examines the approaches in carrying out property mass valuation of vacant land. Findings of the research are that the unimproved value of any property is a function of two elements, the site derived value and accessibility value. The accuracy of the land valuation model is determined by choice of input parameters and variable reduction technique adopted to account for multicollinearity. Random forest algorithm applied on the first 9-latent variables derived using PLS technique PLS generally performed better in estimating the variations of vacant land value in Nairobi County. The PLS technique of feature reduction performed better than PCA method.

The study provides the opportunity for the Nairobi County government to generate the most updated valuation roll for Nairobi. This will have a direct impact on ensuring fairness in the amount of property tax to be paid by Nairobi property owners. The valuation roll will be based on the modelled property values that represent the actual dynamic market prices. Therefore, the county can be able to increase their revenue and adjust for the inflation appropriately. The methodology also provides an efficient and less costly property valuation alter-

native that compliments the cumbersome and expensive traditional parcel based valuation process.

A Majority of the dependent variables or predictors used in the prediction of vacant land prices are highly correlated to each other. Thus, strategies to deal with multi-collinearity need to be prioritized in future studies. The existing gap in this study is the inability to assess the effect of price speculation on the value of vacant land. The inclusion of speculation data in the modelling process of vacant land valuation may give insights into the prediction of future land value estimates.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Kenyalaw.org. (2023) The Land Rating Act. <http://kenyalaw.org:8181/exist/rest//db/kenyalex/Kenya/Legislation/English/Amendment%20Acts/No.%202020%20of%201964.pdf>
- [2] Nyabwengi, L.M. and K'Akumu, O.A. (2019) An Evaluation of Property Tax Base in Nairobi City. *Journal of Financial Management of Property and Construction*, **24**, 184-199. <https://doi.org/10.1108/jfm-pc-05-2019-0043>
- [3] Nyabwengi, L., K'Akumu, O.A. and Kimani, M. (2020) An Evaluation of the Property Valuation Process for County Government Property Taxation, Nairobi City. *Africa Habitat Review Journal*, **14**, 1731-1743.
- [4] Scherthanner, H., Asche, H., Gonschorek, J. and Scheele, L. (2017) Spatial Modeling and Geovisualization of Rental Prices for Real Estate Portals. *International Journal of Agricultural and Environmental Information Systems*, **8**, 78-91. <https://doi.org/10.4018/ijaeis.2017040106>
- [5] Derdouri, A. and Murayama, Y. (2020) A Comparative Study of Land Price Estimation and Mapping Using Regression Kriging and Machine Learning Algorithms across Fukushima Prefecture, Japan. *Journal of Geographical Sciences*, **30**, 794-822. <https://doi.org/10.1007/s11442-020-1756-1>
- [6] Abidoye, R.B. and Chan, A.P.C. (2018) Improving Property Valuation Accuracy: A Comparison of Hedonic Pricing Model and Artificial Neural Network. *Pacific Rim Property Research Journal*, **24**, 71-83. <https://doi.org/10.1080/14445921.2018.1436306>
- [7] Kenya National Bureau of Statistics (2019) Home—Kenya National Bureau of Statistics, Nairobi, Kenya. <https://www.knbs.or.ke>
- [8] McCluskey, W., Franzsen, R., Kabinga, M. and Kasese, C. (2018) The Role of Information Communication Technology to Enhance Property Tax Revenue in Africa: A Tale of Four Cities in Three Countries. <https://doi.org/10.1080/14445921.2018.1436306>
- [9] Cook, R.D. and Forzani, L. (2019) Partial Least Squares Prediction in High-Dimensional Regression. *The Annals of Statistics*, **47**, 884-908. <https://doi.org/10.1214/18-aos1681>
- [10] Koch, I. and Naito, K. (2010) Prediction of Multivariate Responses with a Selected

-
- Number of Principal Components. *Computational Statistics & Data Analysis*, **54**, 1791-1807. <https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/14153>
<https://doi.org/10.1016/j.csda.2010.01.030>
- [11] Akinwande, M.O., Dikko, H.G. and Samson, A. (2015) Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. *Open Journal of Statistics*, **5**, 754-767. <https://doi.org/10.4236/ojs.2015.57075>
- [12] Čeh, M., Kilibarda, M., Lisec, A. and Bajat, B. (2018) Estimating the Performance of Random Forest versus Multiple Regression for Predicting Prices of the Apartments. *ISPRS International Journal of Geo-Information*, **7**, Article No. 168. <https://doi.org/10.3390/ijgi7050168>
- [13] Lasalvia, M., Capozzi, V. and Perna, G. (2022) A Comparison of PCA-LDA and PLS-DA Techniques for Classification of Vibrational Spectra. *Applied Sciences*, **12**, Article 5345. <https://doi.org/10.3390/app12115345>
- [14] Sawatsky, M.L., Clyde, M. and Meek, F. (2015) Partial Least Squares Regression in the Social Sciences. *The Quantitative Methods for Psychology*, **11**, 52-62. <https://doi.org/10.20982/tqmp.11.2.p052>