THE INTERNATIONAL JOURNAL OF
BUSINESS & MANAGEMENT

Comparative Performance of Machine Learning Classifiers in Detecting Performing and Non-performing Residential Property Renters

Dr. A. O Adewusi
Senior Lecturer, Department of Estate Management,
Federal University of Technology, Akure, Nigeria
I.A.Oguntokun
Senior Lecturer, Department of Estate Management,
Federal University of Technology, Akure, Nigeria

Abstract:
In the recent decades, the use of artificial intelligence (AI) technology in decision making has continued to gain popularity in many disciplines including finance, marketing, insurance, engineering, and medicine to mention but a few, however, their applications have been very limited in the residential rental property market. The aim of the current paper is to compare the performance of four selected ML classifiers including Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR) with a view to choosing the best classifier appropriate for rental applications screening. A total of 724 data samples of the residential rental applications were obtained from the databases of 53 Estate Surveyors and Valuers (property managers), licensed to practice in the Lagos Metropolitan property market, Nigeria. The collected data were subdivided into training and testing datasets representing 70% and 30% respectively, and were analyzed using Python 3. The data were also used in determining the respective classification power of the classifiers using eight performance metrics such as recall/sensitivity, specificity, Type I error, Type II errors, and precision, others include F1 score, F1 adjusted measure, Mathew's Correlation Coefficient and area under the curve (AUC). The results reveal among others that the performance of all the four selected classifiers was good and satisfactory. However, DT outperformed the other classifiers in detecting true positive and false positive, while SVM achieved a better result than other classifiers in detecting false negative (Type II). As revealed in the study, the empirical comparison among different algorithms suggests that no single classifier is best for all learning tasks. The models provide cost and time-saving inputs for property investors, property management professionals, and policymakers.

Keywords: *Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, Renter screening.*

1. Introduction

The increase in tenants-related challenges such as litigation arising from landlord/tenants' disputes, eviction leading to homelessness, and rent default have become worrisome in the rental property markets as these challenges impact negatively on the expected property investment returns. Studies have shown that one of the major causes of these challenges is the subjective rental application screening approach usually adopted by property managers during tenants' selection (Louis et al. 2019 and Onyejiaka et al. 2020).

The goal of every property investor is to enhance the capital invested by a way of a regular stream of income in form of returns, and good/performing tenants are very key to realizing such a goal. A good/performing tenant is the one that pays the agreed rent regularly, cares for the property, adheres to all the obligations arising from the tenancy, and lives peaceably with surrounding neighbors in the property. While bad/non-performing tenants behave to the contrary. It thus means that a wrong choice of tenants into the vacant properties is capable of frustrating the expected property returns (Gbadegesin and Ojo, 2013 and Thambu, 2019).

Generally, property managers employ a number of tenant screening models which include personal judgment, traditional statistical method (Linear Discriminant Analysis and Logistic Regression), and machine learning algorithm. It is pertinent to note that the use of personal judgment in screening prospective tenants is widely practiced in the Nigerian property market with the consequence of populating vacant premises with non – performing tenants leading to incessant crisis (Olawale, 2011). The use of machine learning algorithms has been popularized in other disciplines such as finance, engineering, science etc. but its application is limited in real estate parlance especially in the rental property sub-markets (Adewusi et al., 2016; Yacim and Boshoff, 2017). The classifiers in ML include random forest (RF), k – nearest neighbor (KNN), naive Bayes (NB), artificial neural networks (ANN), fuzzy logic (FL), support vector machine (SWM), decision trees (DT) and a host of other ensembles. ML has continued to play crucial roles in many sectors due to its efficient cost-saving and less human psychological bias (Mutupe et al, 2017). The research focusing on ML in the real estate market is still very limited or sometimes not in existence especially in developing countries (Olawale, 2011 and Yacim and Boshoff, 2017).

Aside from this, choosing a particular classifier as being uniformly best in performance may be a bit difficult given the strengths and weaknesses of each of the methods, thus, searching for the best predictive model/classifier is still in progress (Li, 2019). It is therefore imperative to combine multiple models to gain a comprehensive understanding of tenants' selection in the rental property markets. Therefore, this research assesses the application of random forest, decision tree, support vector machine, and logistic regression to residential rental application selection in the Nigerian property markets with a view to choosing the best classifier with high level of predictive power in differentiating the performing rental applicants from the non – performing rental applicants. The rest of the paper is structured as follows; the second section focuses on the review of the literature followed by the methodology, results, discussion, and conclusion.

2. Literature Review

As earlier noted, ML has continued to be in use in many fields of endeavors and for many different tasks. ML methods have been applied for classification purposes as an alternative to human judgment and statistical methods in decision making (Ghori et al., 2019). ML is used in the classification tasks to differentiate group membership based on certain characteristics (good or bad, accept or reject, fraudulent or non-fraudulent, etc.). Much work has been done in the area of classification task using ML classifiers in Finance (Pandey et al., 2017; Li,2019), security (Bahnsen et al., 2013; Yee et al., 2019), land cover image (Noi and Kappas, 2017), fraud detection and software defective (Ll et al., 2019; Iqbal et al., 2019), Medicine (Kim et al, 2017 and Taylor et al, 2018). However, ML application is limited in rental application screening (Yacim and Boshoff, 2017). The decision to select or not to select a prospective renter is a classification concern just like the ones found in finance decision, security, and land cover classification tasks. Since there is limited or no specific application of ML classifiers in the Nigerian property management sub-market, discussion on previous works in finance, security, medicine, and engineering will perhaps aid in the application of ML in residential rental application screening.

There are many studies that compare different ML methods to find the most suitable for certain kinds of data. Some of them are stated below;Galindo andTamayo (2000) investigated home mortgage loan risk using ML classifiers. The study conclude that the Decision Trees classifier outperformed other ML classifiers. The study also showed that ML methods obtained higher accuracy to predict home mortgage loan risk than the statistical method of probit analysis. However, the study only focused on mortgage loan risk assessment. In fraud detection, Bahnsen et al (2013) examined the performance of logistic regression, decision tree, and random forest in fraud detection, the results revealed that RF outperformed other ML classifiers more efficiently. Li et al. (2017) compared the performance of K-Nearest Neighbor, SVM, and ANN in driver drowsiness detection. The study showed that SVM and ANN achieved superior prediction of driver drowsiness over KNN. Noi and Kappas (2017) also compared the performance of random, KNN, and SVM in land cover image classification, the result of the study shows that SVM outperformed other ML classifiers having achieved the highest overall accuracy. Furthermore, Iqba et al (2019) compared the performance of naïve Bayes, multilayer perceptron, radial basics, SVM, KNN, k-star, decision tree, and random forest in software defection prediction. The study concluded that the selected classifiers performed very well in different problem areas, however, none of the classifiers could claim superiority in the result attained over the others across the classification decision areas. In another study by Yee et al (2019) where the performance of NB, LR, and DT was investigated, the result revealed that NB performed more efficiently than other classifiers, Ghori et al (2019) also examined the performance of ML classifiers for non-technical loss detection in electricity distribution, the results show that random forest improves the efficiency of non-technical detection prediction better than SVM and KNN.

Dornadula and Geetha (2019) examined Credit Card Fraud Detection using Machine Learning Algorithms, the results show that while other classifiers obtained equally efficiently good accuracy, random forest obtained the best result cutting across the selected performance metrics. Also, Stolfo et al (2000) investigated a credit card fraud detection system using iterative dichotonuser (ID3), regression tree, ripper, and Bayes, the result shows that Bayes had the highest predictive ability in detecting credit card fraud.

In the medical parlance, Kim et al. (2017) compared the performance of four machine learning algorithms including decision tree, random forest (RF), support vector machine (SVM), and *k*-nearest neighbor (KNN) in diagnosis of glaucoma. The random forest model shows best performance and decision tree, SVM, and KNN models show similar accuracy. In the random forest model, the classification accuracy is 0.98, sensitivity is 0.983, specificity is 0.975, and AUC is 0.979. The developed prediction models show high accuracy, sensitivity, specificity, and AUC in classifying among glaucoma and healthy eyes.

Also, Taylor et al. (2018) investigated the application of machine learning algorithms to urinary tract infections prediction. The machine learning models had an area under the curve ranging from 0.826±0.904, with extreme gradient boosting (XGBoost) the top performing algorithm for both full and reduced models. The XGBoost full and reduced models demonstrated greatly improved specificity when compared to the provider judgment proxy of UTI diagnosis OR antibiotic administration with specificity differences of 33.3 (31.3±34.3) and 29.6 (28.5±30.6), while also demonstrating superior sensitivity when compared to documentation of UTI diagnosis with sensitivity differences of 38.7 (38.1±39.4) and 33.2 (32.5±33.9). In the admission and discharge cohorts using the full XGboost model, approximately 1 in 4 patients would be re-categorized from a false positive to a true negative and approximately 1 in 11 patients would be re-categorized from a false negative to a true positive.

Although few studies in ML application have been reported in real estate decisions, however, such studies are focused on real estate valuation. For example, Yacim and Boschoff (2017) investigated the application of different ANN training algorithms to property valuation in South Africa. The study revealed that Bayesian ANN training algorithm obtained a better result than other training algorithms. The paper examined the application of ML to property valuation only but work was not extended to property management. The only known study on ML in relation to property management is Furick (2006) which examined a rental screening model using artificial neural network, the paper concluded that ANN as one of the ML classifiers could be of significant help in sorting rental applicants into groups based on their characteristics (good or bad). However, the paper only considered ANN as one of the ML classifiers but did not compare its result with other classifiers. Comparing the performance of different classifiers will perhaps be of interest and help in choosing the best classifier for residential rental application selection. Hence, the need for this study

3. Methodology

3.1. Data Collection and Variable Description

Data on 800 residential rental applications were obtained from the database of 53 property firms practicing in the Lagos metropolitan property market, Nigeria. Data on 14 renter attributes were extracted from the pre-assessment form usually given to prospective tenants, these attributes are also contained in property management related literature (Olawale, 2011). A total of 76 data samples were excluded from the analyses due to incompleteness (Karim et al, 2018). The remainder of 724 data samples was subdivided into two with 507 datasets used in training the network while 217 were set aside for testing the network. The training data serve the purpose of building the network while the testing data determine the predictive and generalization power of the built model using previously unknown rental applications (Viejo et al., 2019).

The study selected one dependent and 14 independent variables for the study. The former is tenant selection which is measured as a selected group (representing the positive class) and non – selected group (representing the minority class) with values of 1 and 0 respectively. Tenant selection is hypothesized to be a function of the 14 tenant-related attributes which include income, occupation, gender, religion, sex, and marital status others include a reference from the previous landlord/manager, tenant history, default history, family size, tenant relationship with the property manager, age, education, and ethnic background.

Table 1 shows the variable description and apriori hypothesis. Demographic factors such as age, gender, marital status, education, ethnic group, and income were hypothesized to influence the tenant selection outcome (Gan et al., 2005). The tenants' attributes/characteristics are selected from the literature and are also contained in the pre-selection and assessment forms usually given by the property manager to an intending residential renter.

The inclusion of gender, ethnic, and religious characteristics background of the rental applicants is not meant to discriminate against any applicant but to serve as a means of understanding the general nature of the prospective tenants. Especially in areas of hostility and crisis management, the background of prospective tenants may play useful roles in allocating vacant spaces (Dabara et al., 2017). Therefore, the authors retain these variables in order to build the models, more so that the intention of the property manager is to not discriminate against individuals or groups of people.

Variable	Definition	Measurement
Dependent Variable		
SLST	**Selection Status**	**Dummy (1if tenant is selected, 0 if otherwise)**
Independent Variables		
GEND (+)	Gender	Dummy (1, if an applicant is male, 0 if otherwise)
MRTS(+)	Marital Status	Dummy (1, if an applicant is married, 0 if otherwise)
RELG (+)	Religion	Nominal
ETHN (+)	Ethnicity	Nominal
FAMS (+)	Family size	Actual in number
AGE (+)	Age	Actual in year
EDUC (+)	Education	Ordinal
OCCP (+)	Occupation	Nominal
INCY (+)	Income	Scale (₦)
PPTY (+)	Property type	Nominal
REFS (+)	Reference source	Nominal
REWM (+)	Relationship with the property manager	Nominal
DUOD (+)	Duration of default	Scale (Month)
TENH (+)	Tenant history	Dummy (1, if an applicant has a bad rental history, 0 if otherwise)

Table 1: Operationalization of Variables

3.2. Performance Measures

As shown in table 3, the dataset is imbalanced since the majority class (selected applications) are more than the minority class (non - selected applications) being 70% and 30% respectively. The issue of imbalance dataset is not uncommon in classification tasks (Meira et al., 2017), however, a number of methods have been proposed to balance the imbalanced dataset including oversampling the minority and under-sampling the majority class (Li, 2019) or at another time, the dataset might just be analyzed as they are (He and Nawata, 2019). Every method of dealing with an imbalanced dataset comes with its own advantages and disadvantages (Ghori et al, 2019). The issue of merits and demerits of the methods often used in handling the imbalanced datasets has been well reported in the literature (Kim et al, 2016). It is not within the scope of the current study to examine the methods of resampling of imbalanced datasets all over but the study essentially focuses on using the performance metrics that are insensitive to class imbalance. Modeling the imbalanced dataset is a challenging task due to the fact that the distribution of classes (target variable Y) is non-uniform (Bhattacharyya et al., 2011). The situation becomes more difficult if the built model focuses on the minimization of negative class (Ghori et al., 2019). In such a situation, the performance metric that over-represents the majority (selected) class may not really reveal the model's predictive power. The important thing in a situation of the imbalanced dataset is that the performance measure that over-represents the majority class at the expense of the minority class should be disparaged when determining the efficiency of any built model. Accuracy, which is one of the commonest measures used in determining the actual performance of a model is biased toward the positive class (selected class) and as such it is inappropriate as an indicator of model performance. In this regard, measures that evaluate the correctness of the positive and negative classes separately are often preferred (Ghori et al., 2019). To achieve this, a basic confusion matrix is used to calculate true positive (TP), true negative (TN), false positive (FP), and false-negative (FN). In this study, the basic confusion matrix was used to calculate other more complex metrics that represent positive and negative classes separately or in a combined way (Razzak et al., 2019). Thus, specificity, sensitivity, precision, Type I and Type II errors, F1 score, adjusted F- measure, MCC and AUC metrics are used in the evaluation of the model due to the fact that accuracy lacks the power to realistically predict FP and FN in an imbalanced dataset, the confusion matrix -based measures are briefly discussed as follows;The confusion matrix showing the predicted and real values is indicated in table 3;

		Predicted Value	
		Positive (P)	**Negative (N)**
Real Value	Positive (P)	True Positive (TP)	False Negative (FN)
	Negative (N)	False Positive (FP)	True Negative (TN)

Table 2: A Typical Confusion Matrix Structure

3.2.1. True Positive (TP)

This is the correct classification of a good/performing rental applicant as a good rental applicant by the classifier i.e., good applicants classified as good.

3.2.2. True Negative (TN)

It is a correct classification of a bad rental applicant as a bad rental applicant i.e., bad applicants classified as bad. False Positive (FP) refers to an incorrect classification of a bad rental applicant as a good rental applicant i.e., bad applicants classified as good, while False Negative (FN) refers to the incorrect classification of a good rental applicant as a bad rental applicant i.e., good applicants classified as bad.

3.2.3. True Positive Rate (TPR)

Also known as sensitivity or recall, measures the total number of the selected performing tenants that are correctly classified as such by the classifier. Increase percentage of sensitivity will lower the percentage of false negative. This implies that the model that has the ability to increase the number of the selected performing tenant will in turn affects the property return level. It would be better to have a high Recall as the property investors do not want to lose their returns, in such a situation it will be a good idea to alarm the property managers even if there is a slight doubt about the rent payment ability of the rental applicants. Therefore, any classifier that obtains such predictive power is considered effective in tenant selection is expressed as;

$$Sensitivity = \frac{TP}{TP + FN} \quad 1$$

Specificity refers to the proportion of correctness of the false samples classified as false samples. It is used to measure the model's performance on negative class and the calculation is given by;

$$Specificity = \frac{TN}{TN + FP} 2$$

Where TP, TN, FP, and FN areas earlier stated.

3.2.4. Harmonic Mean

Harmonic mean is suitable for measuring the imbalanced set for its ability to equally distribute the majority and minority classes. When harmonic mean is closer to zero, it is an indication that both precision and recall are high (Sun et al., 2007).

3.2.5. Precision

Precision is the ratio of true positivesto the total of the true positives and false positives. Precision looks to see how much junk positives got thrown in the mix. If there are no bad positives (those FPs), then the model had 100%

precision (Ghori et al, 2019). The more FPs that get into the mix, the uglier that precision is going to look. It is also expressed as;

$$\text{Precision} = \frac{TP}{TP + FP} \quad 3$$

3.2.6. Geometric Mean

According to (Comert and Kocamaz, 2017), the distribution of data between classes may be unbalanced in many cases, in which case the geometric mean (GM) metric becomes extremely useful to achieve more objective results. GM maximizes the classification accuracy of the entire population based on the balance accuracy between the positive and negative classes. The Geometric Mean (G-Mean) measures the balance between classification performances on both the majority and minority classes (Mustika et al., 2020). A low G-Mean is an indication of poor performance in the classification of the positive cases even if the negative cases are correctly classified as such. This measure is important in the avoidance of over fitting the negative class and underfitting the positive class (Akosa, 2017). GM is expressed as;

$$G - mean = \sqrt{Sensitivity \times Specificity} \quad 4$$

F-Measure;F-Measure is also called F1-score, and it represents the harmonic mean between recall and precision values (Sokolova et al, 2006). FM is used when the performance on both positive and negative classes are required to be high, the value ranges from 0 to1, and high values of F-measure indicate high classification performance (Tahir et al., 2019; Khor et al., 2019). It is calculated as;

$$F - Measure = \frac{2 \times Sensitivity \times Precision}{Sensitivity + Precision} \quad 5$$

The Adjusted F-Measure (AGF)is an improvement over the F-Measure especially when the data is imbalanced. The AGF metric is introduced to use all elements of the confusion matrix and provides more weights to samples that are correctly classified in the minority class (Tharmat, 2018). The AGF is calculated by first computing

$$F_2 = 5 \times \frac{Sensivity \times Precision}{(4 \times Sensitivity) + Precision} \quad 6$$

After the class labels of each case are switched such that positive cases become negative and vice versa. A new confusion matrix concerning the original labels is created and the quantity:

$$Inv\ F_{0.5} = \frac{5}{4} \times \frac{Sensitivity \times Precision}{(0.5^2 \times Sensitivity) + Precision} \quad 7$$

The AGF is finally computed by taking the geometric mean of $F2$ and $InvF_{0.5}$ as

$$AGF = \sqrt{F_2 \times Inv\ F_{0.5}} \quad 8$$

Where F2 and InvF2 are F measure and inversion of F measure respectively.

Matthew's Correlation Coefficient; The Matthews correlation coefficient (MCC) is least influenced by imbalanced data, it represents the correlation between the observed and predicted classifications, and it is calculated directly from the confusion matrix (Akosa, 2017). The value ranges from -1 to +1. A coefficient of −1 shows that there is a perfect disagreement between the actual and the predicted, +1 when there is a perfect agreement, while 0 means no better than random and -1 the worst possible prediction. (Hague et al, 2016 and Boughorbel et al., 2017).

$$MCC = \frac{TP \times TN \times -FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad 9$$

The area under the curve (AUC);AUC provides an aggregate measure of performance across all possible thresholds, when AUC value is more than 0.5 and it is close to unity, the better the discriminating power of the model, it also means that the model is good enough to distinguish between the positive and negative classes (Saji and Balachandran, 2015). AUC is expressed as;

$$\frac{1}{2}(Sensitivity + Specificity) \quad 10$$

3.3. Machine Learning Algorithms

The selected machine learning algorithms are briefly discussed below;

3.3.1. Logistic Regression

The logistic regression model tries to identify the correlation between the dependent and the independent variables. It is mainly used for binary classification where the target variable is binary and one or more independent variables can be continuous or binary. This model uses the logistic function to determine the probability of the output with respect to the inputs. The classification is performed such that a threshold is provided, and all the probability values greater than a certain threshold are assigned one class and the values less than the threshold are assigned the other class (Hosmer and Lemeshow, 2000).

An alternative to the linear DA model is logistic regression, a method that has fewer assumptions than linear discriminant models (i.e., no multivariate normality and equal dispersion assumptions) (Cox, 1970)

A logistic function having the following form is used, $y = \frac{1}{1+e^{y}}$ 11

$$y = a + \sum_{i=1}^{n} bixi \quad 12$$

Where xi represents the set of individual variables,
bi is the coefficient of the ith variable, and

Y is the probability of a favorable outcome.

Logistic regression as a qualitative response model is appropriate when the dependent variable is categorical (Magden et al, 1978). In this study, the dependent variable selection status is binary, and logistic regression is a widely used technique in such problems (Hosmer and Lemeshow, 2000). Binary choice models have been used in studying decision-related tasks in finance, marketing, and insurance policy related decisions; In the case of insurance fraud, investigators use the estimated probabilities to flag individuals that are more likely to submit a fraudulent claim. it has been argued that identifying fraudulent claims is similar in nature to several other problems in real life including medical, real estate, and epidemiological problems (Caudill et al, 2005).

3.3.2. Support Vector Machine:

Support vector machines (SVMs) are statistical learning techniques that have been found to be very successful in a variety of classification tasks (Li, 29019). Several unique features of these algorithms make them especially suitable for binary classification problems like decision making and detection of suspicious moves (Xie et al, 2009). SVMs are rooted in statistical learning / supervised theory and were developed by Viadimir in the 1990s. The SVM looks at the extreme boundaries and draws the edges often as hyperplanes which separate two classes. SVMs are efficiently used for learning and classify users as reported in Wang and Stolfo (2003). SVM creates a margin between the two classes and tries to maximize the margin, this way it constructs an optimal decision function f(x) that can predict unseen instances with high accuracy as given in Eqn 2where sgn(g(x)) in the boundary between the positive and negative classes (Vapnik, 1999).

SVM regression aims to find an approximation to a non-linear function that maps the input data into high dimensional space. Here, a hyperplane is constructed in a way that it separates the data points with a maximal margin with linear regression (Ghori et al, 2019). Its operation is expressed as in equation 13;

$$f(x) = \operatorname{sgn}\left(g(x)\right) \qquad 13$$

Where, $f(x)$ is the decision function, $\operatorname{sgn}\left(g(x)\right)$ is the boundary between the positive and negative classes (Vapnic, 1999). The expected error in classification is expressed in equation 4;

$$R < \tfrac{t}{n} + \sqrt{h\left(\ln(\tfrac{2n}{h})+1\right) - \ln(\tfrac{n}{4}) \Big/ n} \qquad 14$$

Where R is the expected error,

t is the training error,

n is the number of training samples,

h is the dimension of the set of hyperplanes.

Applications of SVMs also include bioinformatics, machine vision, text categorization, and time series analysis (Statnikov and Wang,2007). The strength of SVMs comes from two important properties they possess — kernel representation and margin optimization.

3.3.3. Random Forest

Random forest is also a type of supervised learning algorithm which makes use of the ensemble method. It is a combination of an ensemble of decision trees (Statnikov and Wang, 2007). The attributes for the split are determined randomly and used for generating the individual trees. At the time of classification, the votes of all the decision trees are taken into consideration and the class with the majority vote is assigned as the class for the given value.

The random forest consists of multiple individual decision trees, for each tree, a separate set of training examples is selected using this offer solution to the problem of over fitting in an imbalance dataset. At the testing phase, the outcome of a sample is evaluated by using the majority voting scheme from among all the individual decision trees (Ghori et al., 2019). This approach presents a unique advantage in that the different training examples are used in every decision tree, also a variable number of nodes or cores can be used for training (Ho, 1995).Random forests are computationally efficient since each tree is built independently of the others. With a large number of trees in the ensemble, they are also noted to be robust to over fitting and noise in the data. Ensembles perform well when individual members are dissimilar, and random forests obtain variation among individual trees using two sources for randomness: first, each tree is built on separate bootstrapped samples of the training data; secondly, only a randomly selected subset of data attributes is considered at each node in building the individual trees. Random forests thus combine the concepts of bagging, where individual models in an ensemble are developed through sampling with replacement from the training data, and the random subspace method, where each tree in an ensemble is built from a random subset of attributes. Ensemble methods seek to address this problem by developing a set of models and aggregating their predictions in determining the class label for a data point.

Other studies comparing the performance of different learning algorithms over multiple datasets have found that random forest shows good overall performance (Caneana et al, 2008, Khashgoftan, 2007). Random forests have been applied in recent years across varied domains from predicting customer churn (Xie et al, 2009), image classification, to various biomedical problems. While many papers note their excellent classification performance in comparison with other techniques including SVM, a study by Statnikov and Wang, (2007) finds SVM to outperform random forests for gene expression microarray data classification. The application of random forests to fraud detection and other classification tasks especially in fraud detection, loan default, and management is relatively new, with few reported studies.

3.3.4..Decision Trees

Decision tree is one of the predictive models which maps observation about an item represented in branches to a conclusion about a target value represent in leaves (Pandey et al., 2017). It is recognized as an important supervised learning technique with each internal node or non - leaf node labelled as an input feature. The individual leaf node in the tree is labelled with a class or probabilistic distribution over the class (Bask et al., 2011; Currat and Orri, 2011). The branches between the nodes indicate the possible values that these attributes can have in the observed samples while the terminal node talks about the final value of the dependent variable (Wang et al., 2005). The popularity of decision tree models in data mining arises from their ease of use, flexibility in terms of handling various data attribute types, and interpretability. They have been found to perform favorably in comparison with support vector machine and other current techniques (Breiman, 2001).

Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too. The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data (training data). Decision trees classify the examples by sorting them down the tree from the root to some leaf/terminal node, with the leaf/terminal node providing the classification of the example. Each node in the tree acts as a test case for some attribute, and each edge descending from the node corresponds to the possible answers to the test case. This process is recursive in nature and is repeated for every sub tree rooted at the new node.

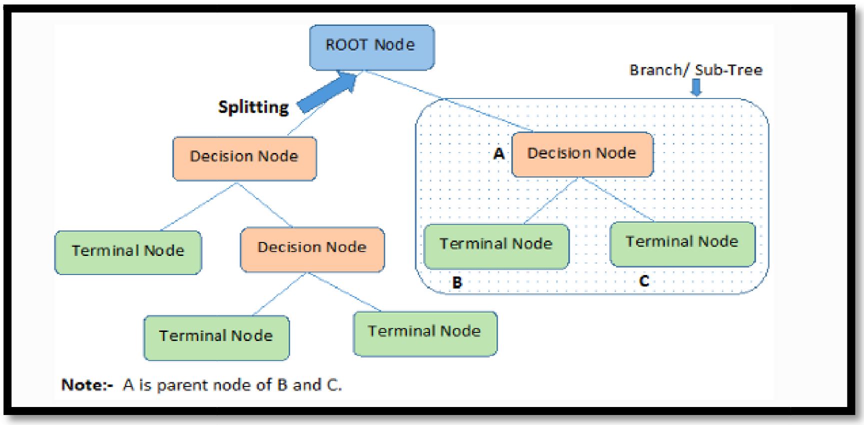

Figure 1: A Typical Decision Tree

4. Results and Discussion

Figure 2, tables 3 and 4 are used in the presentation of the results and discussion;

S/no.	Variables		Frequency	Percentage
1	Dependent variable	Not selected	193	26.7
	Prospective rental applicant status	Selected	531	73.3
		Total	724	100
2	Independent Variables	Female	241	33.3
	Gender	Male	483	66.7
		Total	724	100
3	Marital	Marital	464	64.1
		Single	260	35.9
		Total	724	100
4	Religion	Others	6	0.82
		Muslim	133	18.377
		Christianity	585	80.8
		Total	724	100
5	Ethnicity	Hausa	6	0.82
		Yoruba	503	69.47
		Igbo	215	29.69
		Total	724	100
6	Family	7-10	57	7.9
		4-6	304	42.0
		1-3	363	50.1
		Total	724	100
7	Age	41 and above	191	26.4
		31-40	381	52.6
		18-30	152	21
		Total	724	100

S/no.	Variables	Frequency	Percentage	
8	Educational status	Ph.D.	9	1.24
		MSc	79	10.91
		BSc	456	62.98
		O level	180	24.86
		Total	724	100
9	Occupation	Self-employed	282	38.9
		Private company	299	41.3
		Civil Servant	143	19.8
		Total	724	100
10	Income	400,000 and above	34	4.7
		200,00-350,000	216	29.8
		50,000-150,000	474	65.5
		Total	724	100
11	References	Community leader	16	2.2
		Religious leader	103	14.2
		Friend/ colleague	314	43.4
		Employer	291	40.2
		Total	724	100
12	Property type	Residential	724	100
13	Relationship with the property manager	None	458	63.3
		Same ethnic	63	8.7
		Church/mosque member	72	9.9
		Family	10	1.4
		Friend	121	16.7
		Total	724	100
14	Length of default	None	570	78.7
		>7 months	39	5.4
		4-6 months	27	3.7
		1-3 months	88	12.2
		Total	724	100
15	History of default	Bad History	163	22.5
		Good history	561	77.5
		Total	724	100

Table 3: Frequency Distribution of Research Variables

Table 3 shows the frequency distribution of the variables in the study, the number of non-selected prospective tenants is 193 representing 26.7% while the number of the selected prospective tenants is 531 representing 73.3%. This is a case of an imbalanced dataset which is usually encountered in the classification tasks. The frequency distributions of other variables are contained in table 3.

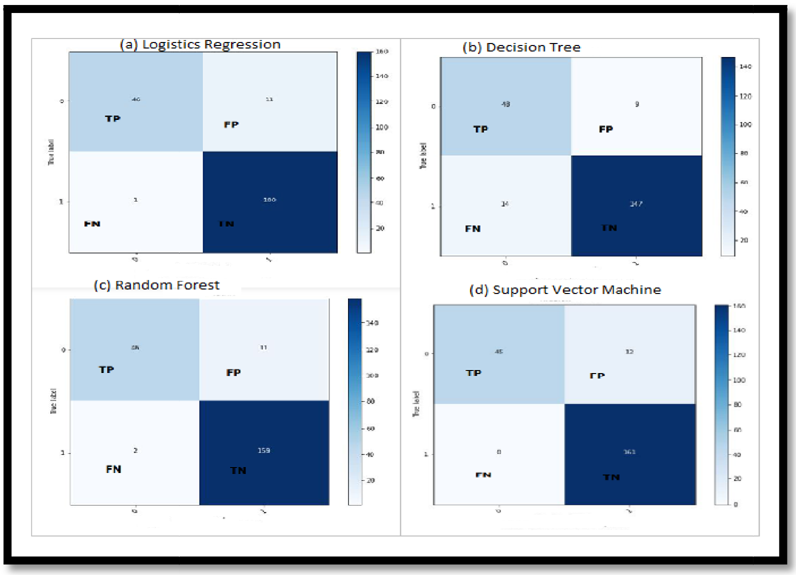

Figure 2: The Confusion Matrix Structure of the Selected Classifiers

Figure 2 and table 4 show the respective classification values of TP, FP, TN and FN

N0	Classifier	TP	FP	TN	FN
1	Logistic Regression	46	11	160	1
2	Decision Tree	48	9	147	14
3	Random Forest	46	11	159	2
4	Support Vector Machine	45	12	161	0

Table4: Values of TP, FP, TN and FN across the Four Selected Classifiers

It is necessary to note that property managers will normally seek to embrace classifiers that have, first, a high TP, second, low FP and FN in order to increase the number of the performing tenants in the vacant properties, anything short of this has the capability of affecting the expected property returns either in terms of rent default, covenant default or litigation.

From table 4 and figure 2, the predicted values of all the classifiers range from 45 to 48 with the decision tree having the highest value of TP (48), followed by random forest (46) and logistic regression (46) and the least being 45 for support vector machine. The choice of a classifier is a function classification problem at hand (Ghori et al., 2019). The implication is that when a property manager has the goal of selecting well-performing tenants, the manager should appropriately select a classifier with high predictive power in detecting a true tenant (performing tenant). The ability of a classifier in detecting good tenants is germane to the realization of investment returns. As shown in table 4, DT has the highest predictive ability in detecting true positive (good/performing rental applicants), in that regard, DT may be selected for residential rental screening. This finding contradicts the finding of Saberioon et al. (2018) which claimed that SVM outperforms other supervised training algorithms, however, the finding of the current study agrees with the finding of Ghori et al. (2019) which concluded that DT obtained the highest performance among other classifiers in detecting true and false positives.

Also, in terms of false-positive (bad rental applicants predicted by the classifiers as good applicants) and false negative (good rental applicants predicted by classifiers as a bad applicant). The former represents the Type I error while the latter indicates the Type II error. The two types of error generate cost but the cost of Type I error which is the incorrect classification of a bad rental applicant as a good rental applicant is more than the Type II error. Interestingly, the decision tree has the lowest value of false-positive of 9 samples followed by random forest (11 samples), logistic regression (11 samples), and support vector machine (12 samples).

It thus means that the decision tree as a classifier misclassified only 9 bad rental applicants as good rental applicants. A classifier with a low FP value is desirable and provides inputs for the tenant selection decision. However, in terms of a false negative, that is, the ability of the classifier to classify good tenants as bad tenants, SVM generated the best performance as it has zero value of misclassifying good tenants as bad tenants while DT, LR, and RF misclassified 14, 1 and 2 good applicants as bad applicants respectively. This finding corroborates the work of Basheri et al. (2019) which claims that SVM outperformed DT, RF, and LR in detecting Type II errors but decreased in performance with a large sample. Thus, when a manager intends to minimize good applicants misclassified as bad applicants, DT will not make an appropriate classifier but rather SVM and LR may produce better results. The cost of type II error is that the managers lose income/ performing rental applicants that would have occupied the vacant premises and contributed to the profitability of the Property.

Classifiers/Metrics	Logistic Regression	Decision Tree	Random Forest	Support Vector Machine
Precision	0.81	0.84	0.81	0.79
F Measure	0.88	0.91	0.88	0.88
AGF	0.88	0.91	0.89	0.88
MCC	0.86	0.74	0.84	0.86
AUC	0.85	0.85	0.87	0.87
SE	0.96	0.99	0.98	0.98
SP	0.74	0.71	0.75	0.77
GM	0.85	0.84	0.86	0.87

Table 5: Comparison of the Performance of the Selected Machine Learnings using Certain Criteria

Table 5 shows other performance metrics as they affect the selected classifiers. The classifier with a high sensitivity/recall value is most suitable in the classification task (Ghori et al., 2019). Sensitivity/recall refers to selected tenants that are predicted as selected tenants. The higher the value of sensitivity the better the classifier. Table 5 shows that DT outperformed other classifiers having 98.9% ability to predict good and performing tenants as good performing tenants while SVM, RF, and LR and 98.2%, 97.8%, and 96.3% respectively. From the analysis, all the selected classifiers obtained very good and high sensitivity values which range from 96.3% to 98.9%. This implies that any of the classifiers may be applied to tenant selection in the Nigerian rental property market. The implication of DT superiority is that as more of good performing tenants are selected, the property returns become more guaranteed.

As also shown in table 5, different performance metrics rank classifiers differently as revealed in the subsequent section;

- Precision, F- measure, SE, and AGF revealed that DT outperformed other classifiers

- MCC shows that LR outperformed all other selected classifiers, and
- AUC, SP, and GM also revealed that SVM is superior in performance in comparison to other classifiers.

The paper observes that the choice of the best classifier is a function of the kind of problem the property managers wish to solve. All the selected classifiers have performed very well as revealed by the eight different performance metrics. However, each of them obtained superior value than others in different areas. This finding corroborates the work of Shavlik et al. (1991) which emphasized that comparison among different machine learning algorithms suggest that no single classifier is best for all learning tasks as each of them is only best for some tasks but not all tasks

5. Conclusion

The paper attempts to examine the performance of four different machine learning algorithms including DT, LT, RF, and SVM with the goal of choosing the best among them using 8 performance metrics. A total of 724 residential rental applications were obtained from 58 property managers practicing in Lagos, the Nigerian property market. The data were subdivided into training and testing data subsets. The analysis was performed on the python 3 data environment. The test data representing 15 % of the total samples were used in evaluating the varying predictive powers of the selected classifiers. The study found that all the classifiers produced good results as revealed by the 8-performance metrics but they show superiority in the different problem areas (exhibiting varying detecting abilities). The ability of the classifier to obtain high sensitivity/ recall, low Type I, and low Type II errors is very vital to rental screening for a guaranteed regular property return. In this regard, DT outperforms LR, RF, and SVM with its highest value and lowest false positive while SVM demonstrated superiority over others by obtaining zero Type II error. The models developed in this study serves as cost and time-saving in the tenant selection process and also provide warning signals to the property managers about any doubtful residential rental applicants. Area of further study could focus on investigating the rest ML classifiers such as ANN, Baye Naive, KNN, Fuzzy logic, and other ensembles using larger datasets.

6. References

i. Adewusi, A. O., Oyedokun, T. B., & Bello, M. O. (2016). Application of artificial neural network to loan recovery prediction. International Journal of Housing Markets and Analysis, 3(3), pp 20-43.
ii. Akosa, J., (2017), Predictive accuracy: a misleading performance measure for highly imbalanced data. In Proceedings of the SAS Global Forum (pp. 2-5).
iii. Bahnsen, A. C., Stojanovic, A., Aouada, D., & Ottersten, B. (2013). Cost sensitive credit card fraud detection using Bayes minimum risk. In 2013 12th international conference on machine learning and applications (Vol. 1, pp. 333-338). IEEE.
iv. Bahnsen, A. C., Villegas, S., Aouada, D., Ottersten, B., Correa, A. M., & Villegas, S. (2017). Fraud detection by stacking cost-sensitive decision trees. Data Science for Cyber-Security.
v. Bhattacharyya, S., Maulik, U., & Dutta, P. (2011). Multilevel image segmentation with adaptive image context-based thresholding. Applied Soft Computing, 11(1), 946-962.
vi. Boughorbel, S., Jarray, F. and El-Anbari, M., (2017). Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PloS one, 12(6).
vii. Breiman, L. (2001). Random forests. Machine learning, 45(1), pp 5-32.
viii. Cananéa, I., Mariz, D., Kelner, J., Sadok, D., & Fodor, G. (2008, March). An on-line access selection algorithm for ABC networks supporting elastic services. In 2008 IEEE Wireless Communications and Networking Conference (pp. 2033-2038). IEEE.
ix. Cömert, Z., & Kocamaz, A. F. (2017). A study of artificial neural network training algorithms for classification of cardiotocography signals. J Sci Technol, 7(2), pp 93-103.
x. Dabara, D. I., Anthony, A. I., Olusegun, O. J., Eleojo, A. G., & Michael, A. O. (2017). Rent Default Factors in Residential Properties in Osogbo Metropolis Osun State, Nigeria. International Journal of Business and Management Studies, 6(1), pp 61-68.
xi. Dornadula, V. N., & Geetha, S. (2019). Credit Card Fraud Detection using Machine Learning Algorithms. Procedia Computer Science, 165, pp 631-641.
xii. Furick, M. T. (2006). *Using Neural Networks to Develop a New Model to Screen Applicants for Apartment Rentals.* Doctoral dissertation, Nova Southeastern University, Graduate School of Computer and Information Sciences
xiii. Galindo, J., & Tamayo, P. (2000). Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications. Computational Economics, 15(1-2), 107-143.
xiv. Gan, C., Limsombunchao, V., Clemes, M. D., & Weng, Y. Y. (2005). Consumer choice prediction: Artificial neural networks versus logistic models.
xv. Gbadegesin, J. T., & Ojo, O. (2013). Ethnic bias in tenant selection in metropolitan Ibadan private rental housing market. Property Management, 5(3), 23-45.
xvi. Ghori, K. M., Abbasi, R. A., Awais, M., Imran, M., Ullah, A., & Szathmary, L. (2019). Performance analysis of different types of machine learning classifiers for non-technical loss detection. IEEE Access, 8, 16033-16048.
xvii. Gonzalez Viejo, C., Torrico, D. D., Dunshea, F. R., & Fuentes, S. (2019). Development of artificial neural network models to assess beer acceptability based on sensory properties using a robotic pourer: a comparative model approach to achieve an artificial intelligence system. Beverages, 5(2), 33.
xviii. Haque, M.N., Noman, N., Berretta, R. and Moscato, P. (2016). Heterogeneous ensemble combination search using genetic algorithm for class imbalanced data classification. PloS one, 11(1).

xix. Ho, T. K. (1995). Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278-282). IEEE.

xx. HOMER, D., & Lemeshow, S. (2000). Applied Logistic Regression, 2000. Journal of Biomechanics, 29, 723-733

xxi. Iqbal, S., Ghani, M. U., Saba, T., & Rehman, A. (2018). Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN). Microscopy research and technique, 81(4), 419-427.

xxii. Karim, A., Loqman, C., & Boumhidi, J. (2018). Determining the number of clusters using neural network and max stable set problem. Procedia Computer Science, 127, 16-25.

xxiii. Khor, R.C., Nguyen, A., O'Dwyer, J., Kothari, G., Sia, J., Chang, D., Ng, S.P., Duchesne, G.M. and Foroudi, F., (2019). Extracting tumour prognostic factors from a diverse electronic record dataset in genito-urinary oncology. International journal of medical informatics, 121, pp.53-57.

xxiv. Kim, J., Kim, J., Thu, H. L. T., & Kim, H. (2016). Long short-term memory recurrent neural network classifier for intrusion detection. In 2016 International Conference on Platform Technology and Service (PlatCon) (pp. 1-5). IEEE.

xxv. Kim, S. J., Cho, K. J., & Oh, S. (2017). Development of machine learning models for diagnosis of glaucoma. PLoS One, 12(5), e0177726.

xxvi. Li, B., Qin, Y., Yuan, B., & Lilja, D. J. (2019). Neural network classifiers using a hardware-based approximate activation function with a hybrid stochastic multiplier. ACM Journal on Emerging Technologies in Computing Systems (JETC), 15(1), 1-21.

xxvii. Li, Y., Bradshaw, J., & Sharma, Y. (2019). Are generative classifiers more robust to adversarial attacks? In International Conference on Machine Learning (pp. 3804-3814). PMLR.

xxviii. Louis, I. C., Ifeanyi, E. F., &Henry, O. C. (2019). Examination of Causes of Tenancy Disruption and Its Effects on Property Investment in Anambra State, Nigeria.

xxix. Mustika, H.F., Syafiandini, A.F., Manik, L.P. and Rianto, Y., (2020). Evaluating Naïve Bayes Automated Classification for GBAORD. Computer Engineering and Applications Journal, 9(1), pp.29-37.

xxx. Mutupe, R. M., Osuri, S. O., Lencwe, M. J., & Chowdhury, S. D. (2017). Electricity theft detection system with RF communication between distribution and customer usage. In 2017 IEEE PES PowerAfrica (pp. 566-572). IEEE.

xxxi. Olawande, O. A. (2011). Harnessing real estate investment through proper tenant selection in Nigeria. Property Management, 2(5), 56- 70.

xxxii. Onyejiaka, J. C., Chinagorom, J. C., & Ndive, P. C. (2020). An Examination of Tenant Eviction in Anambra State, Nigeria. Tropical Built Environment Journal, 7(1).

xxxiii. Pandey, N., Sanyal, D. K., Hudait, A., & Sen, A. (2017). Automated classification of software issue reports using machine learning techniques: an empirical study. Innovations in Systems and Software Engineering, 13(4), 279-297.

xxxiv. Pandey, P., Deepthi, A., Mandal, B., & Puhan, N. B. (2017). FoodNet: Recognizing foods using ensemble of deep networks. IEEE Signal Processing Letters, 24(12), 1758-1762.

xxxv. Saberioon, M., Císař, P., Labbé, L., Souček, P., Pelissier, P., & Kerneis, T. (2018). Comparative performance analysis of support vector machine, random forest, logistic regression and k-nearest neighbours in rainbow trout (oncorhynchus mykiss) classification using image-based features. Sensors, 18(4), 1027.

xxxvi. Saji, S.A. and Balachandran, K. (2015) Comparative Study of various training algorithms of Artificial Neural Networks on Diabetes dataset. International Journal on Recent and Innovation Trends in Computing and Communication, 3(2), pp.378-382.

xxxvii. Shavlik, J. W., Mooney, R. J., & Towell, G. G. (1991). Symbolic and neural learning algorithms: An experimental comparison. Machine learning, 6(2), 111-143.

xxxviii. Sokolova, M., Japkowicz, N. and Szpakowicz, S., (2006). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In Australasian joint conference on artificial intelligence (pp. 1015-1021). Springer, Berlin, Heidelberg.

xxxix. Statnikov, A., Wang, L., & Aliferis, C. F. (2008). A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC bioinformatics, 9(1), 319.

xl. Stolfo, D., Uijl, A., Benson, L., Schrage, B., Fudim, M., Asselbergs, F. W., & Savarese, G. (2020). Association between beta-blocker use and mortality/morbidity in older patients with heart failure with reduced ejection fraction. A propensity score-matched analysis from the Swedish Heart Failure Registry. European Journal of Heart Failure, 22(1), 103-112.

xli. Sun, Y. J., Zhang, S., Miao, C. X., & Li, J. M. (2007). Improved BP neural network for transformer fault diagnosis. Journal of China University of Mining and Technology, 17(1), 138-142.

xlii. Tahir, M.A.U.H., Asghar, S., Manzoor, A. and Noor, M.A., (2019). A classification model for class imbalance dataset using genetic programming. IEEE Access, 7, pp.71013-71037.

xliii. Taylor, A., Leblanc, S., & Japkowicz, N. (2018). Probing the limits of anomaly detectors for automobiles with a cyberattack framework. IEEE Intelligent Systems, 33(2), 54-62.

xliv. Taylor, R. A., Moore, C. L., Cheung, K. H., & Brandt, C. (2018). Predicting urinary tract infections in the emergency department with machine learning. PloS one, 13(3), e0194085.

xlv. Thambu, E. K. (2019). Effects of Rent Default on Real Estate Investment Return in Nairobi County, Kenya (Doctoral dissertation, United States International University-Africa).

xlvi. Thanh Noi, P., & Kappas, M. (2018). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. Sensors, 18(1), 18.

xlvii. Vapnik, V. N. (1999). An overview of statistical learning theory. IEEE transactions on neural networks, 10(5), 988-999.

xlviii. Wang, K., & Stolfo, S. (2003). One-class training for masquerade detection. In *International Workshop on Mathematical Methods, Models, and Architectures for Computer Network Security* (pp. 57-81). Springer, Berlin, Heidelberg.

xlix. Wang, Y., Tetko, I. V., Hall, M. A., Frank, E., Facius, A., Mayer, K. F., & Mewes, H. W. (2005). Gene selection from microarray data for cancer classification—a machine learning approach. Computational biology and chemistry, 29(1), 37-46.

l. Xie, H., Yang, D., Sun, N., Chen, Z., & Zhang, Y. (2019). Automated pulmonary nodule detection in CT images using deep convolutional neural networks. Pattern Recognition, 85, 109-119.

li. Yacim, J. A., & Boshoff, D. G. B. (2018). Combining BP with PSO algorithms in weights optimization and ANNs training for mass appraisal of properties. International Journal of Housing Markets and Analysis.