Journal of Contemporary Research in the Built Environment ISSN: 2636-4468
Volume 3, Number 2, September 2019

AN EVALUATION OF CLASSIFICATION PERFORMANCE OF ARTIFICIAL NEURAL NETWORK AND LOGISTIC REGRESSION AS LOAN SCORING MODELS

Amos Olaolu Adewusi[1]
Department of Estate Management, Federal University of Technology, Akure, Nigeria

ABSTRACT

Purpose: Application of credit risk evaluation techniques has continued to receive more research attention in the advanced economies than emerging economies. However, the findings on the classification performance of these techniques have been mixed. The current paper aims at assessing the classification performance of Artificial Neural Network (ANN) and Logistic Regression Model (LRM) in credit transaction using data from Nigeria emerging economy.
Design/methodology/approach: 2,300 data samples comprising of fully and partially recovered loan accounts were obtained from the databases of eleven commercial banks and sixteen primary mortgage institutions practicing in Lagos metropolis, Nigeria. Also, data on 14 variables comprising one dependent variable (loan recovery status) and thirteen independent variables were collected on each of the data samples. To construct LRM & ANN models, the total samples were subdivided into training/validation and testing samples. 73% of the total samples (1,679) were used for training and validation of the models while 27% of the total samples (613) were used in testing the classification performance of LRM and ANN using overall accuracy, specificity, sensitivity, Type I and Type II errors as criteria for performance measurement. SPSS version 21 was adopted for data analysis.
Findings: The result of the analysis reveals among others that LRM and ANN models generated good overall accuracy values of 76.6% and 91% respectively. However, the performance of ANN is comparatively more efficiently better than that of LRM in detecting 'good' loan applicants, 'bad' loan applicants and in generating lower Type I and Type II errors than LRM. The use of ANN is therefore recommended as a credit risk evaluation technique due to its consistent performance across the performance metrics.
Research limitation/implications: The findings of the paper provide input for lending decision in lending institutions in Nigeria which is capable of minimizing bad debts & non-performing loans thereby enhances stability in financial institutions. However, the finding of the current research is of country-specific, further study may compare the classification capacity of LRM and ANN across advanced and emerging economies. Also, future studies may adopt larger samples than the ones adopted in the current study for more inclusive researches.
Originality/value: As noted above, studies on classification performance of ANN & LRM have been well documented in the advanced economies like UK, USA, China etc, the current paper extends the frontiers of knowledge to the existing body of literature in the advanced economies

[1]aoadewusi@futa.edu.ng

while making one of the foremost empirical articles to examine the subject matter in emerging economies.

Keywords: Artificial neural network; classification performance; default; loan evaluation; logistic regression; loan recovery.

1. INTRODUCTION

Accessibility to credit facility is a strategic input in investment initiation and expansion which also determines the performance of any nation's economy (Omondi and Jagongo, 2018). Generally, loans are sourced from lending institutions usually with covenants of repayment by borrowers (Taiwo, Falodun and Agwu, 2016; Berger, et al, 2019). It is pertinent to note that the approval decision of credit by lending institutions is often associated with risks in terms of default and non-recovery of defaulted accounts. Loan default and poor recovery of same are global phenomena though at varying degrees (Adewusi, 2015 and Guha et al, 2020), for example, International Swaps and Derivative Association Inc (2008) reveals that the global volume of credit default grew from US$631.5 billion in 2001 to US$54.6 trillion by mid-2008 yielding an alarming growth rate of 854.6% in less than a decade. The non-performing loans range from 19% to 48% in the banking portfolio (Nigerian Deposit Insurance Co-operation 2010). It is noteworthy that the challenge on non-performing loans is yet to significantly respond to policy remedies in developing countries especially in Nigeria (Adewusi et al 2016 and Sofayo, 2017). The decision to approve or reject loan applications portends risk of losing returns/principal (Singh, 2020). The decision to wrongly reject loan application reduces the lender's returns while inappropriate approval of 'bad' loan applications produces the risk of losing both the principal and accruable interest. The huge capital loss through non-performance of loans can be ascribed to the use of credit risk evaluation techniques that cannot cope with the reality of uncertainties (Bekhet and Eletter, 2014; Nyasaka, 2017; Leah, 2018 and Grima, and Thalassinos, 2020).

The global financial crisis in 2008 resulted from an unreliable credit rating assignment to mortgage-backed securities (Bae and Kim 2015). Risk estimate is a major factor contributing to any credit decision and it is a key determinant of profitability in lending transactions. However, while the risk is desirable and unavoidable in the business of lending institutions, inability to precisely determine it may adversely affect credit management (Bekhet and Eletter, 2014 and Agustia, et al, 2020). Credit scoring models are employed to determine loan applicant's ability to repay financial obligation (Emel, et al, 2003 and Maldonado, et al,2020), the objective of the credit scoring model is to classify loan applicants into low risk and high-risk categories (Akkoe, 2012 and Lahsasna, et al 2012). The use of subjective/personal judgment of loan officers on the ability of loan applicants concerning loan repayment is prone to human error and inconsistency (Sofayo, 2017). It is necessary to rely on models and algorithms rather than human judgment in consumer lending because of the vast decisions that are involved (Khandani, et al, 2010 and Zeng, 2020). A little improvement in the accuracy of the credit decision might reduce credit risk and translates into important future savings (Chen and Huang, 2003, Tsai and Wu, 2008 and Lahsasna et al, 2010)

Due to the significance of credit risk evaluation, several studies have proposed embracing data mining tools in lending institutions to improve their risk assessment and management of same (Akkoe, 2012). Logistic Regression Model, Discriminant

Analysis, Artificial Neural Networks, Genetic Algorithms, Genetic programming. K-Nearest Neighbour model, Decision Tree, Support Vector Machine and some hybrid models have been used to evaluate credit risk with promising results in terms of predictive accuracy (Bekhet and Eletter, 2014). It is noteworthy that the findings of the predictive capabilities of these loan evaluation algorithms are to a large extent inconclusive in terms of their appropriateness, suitability, and superiority of one over the others. Although credit risk appears to be a global challenge, the advanced countries of the world have done considerable works by adopting more futuristic and sophisticated credit risk assessment techniques (Adewusi, et al, 2016). However, it is noteworthy that the subjective method of assessing loan applicant's creditworthiness appears prevalent in most developing countries (Sofayo, 2017).

Artificial Neural Networks and Logistic Regression models are one of the most efficient and widely used methods for credit risk assessment (Bekhet and Eletter, 2014). However, the findings on the relative predictive ability of the two models require more research efforts especially in terms of their suitability and appropriateness in the loan classification task. The questions are, what is the classification performance of each of the techniques? How reliable are these models in detecting problem loan applications and non-problem loan applications? Against this backdrop, the current study assesses the predictive classification accuracy of LRM and ANN to determine their relative appropriateness and suitability in credit risk measurement.

The rest of the paper is structured as follows: the first section focuses on background introduction followed by a review of relevant literature, research methodology, discussion of result and concluding remarks.

2. Literature Review

Research efforts have continued to focus on credit risk assessment in lending institutions, the studies include the following but are not limited to them. Desai, et al (1996) compare the performance of multilayer perceptron, LRM and Linear discriminant analysis. The authors conclude that the difference in the performance of these models in risk evaluation is not statistically significant especially between LRM and ANN. Chem and Huang, (2003) examine credit rating prediction models using support vector machine (SVM) and Back- Propagated Neural Network (BPNN) to evaluate credit risk using 21 variables from the USA and Taiwanese financial markets. The results show that SVM achieved a better overall accuracy compared to BPNN. Limsombunchai et al (2005) investigate Agricultural loans in Thailand using LRM and ANN, 16,560 data of borrowers between 2001 and 2003 were gathered. The study reveals that in terms of precision, the ANN model might not necessarily predict the borrower's creditworthiness and default risk better than LRM, However, most ANN models can detect Type 1 error than LRM.

Moreover, Koh et al (2006) assert that the best performing credit risk evaluation technique is logistic regression, Neural Network, and decision tree. Similarly, Odeh et al (2006) evaluate loan default in eleven states of the USA (Arkansas, Illinois, Indiana, Kentucky, Michigan, Minnesota, Missouri, North Dakota, Ohio, Tennessee, and Wisconsin) with a large sample dataset of 157,853 credit transactions. ANN and LRM were used in estimating the probability of borrower default, the results show that the LRM model gives the best performance in identifying problem loans but exhibits poor performance in predicting problem-free loan applicants, however, ANN displayed the best accuracy in correctly predicting problem free loans.

Gouvea and Goncalves (2007) assess the predictive performance of ANN, LRM and Genetic Algorithms (GA) in Brazilian financial institutions with a dataset of 20,000 comprising of 'good' and 'bad' loan applicants spanning between 2002 and 2003, the result of the study indicates that LRM presented a slightly better result than ANN but both models were better than the model generated by GA.

Also, Youn and Gu (2010) compare the accuracy of LRM and ANN in predicting business failure for Korean lodging firms. The results indicate that ANN records better overall accuracy and correctly classified 'good' loan applicants while both models give the same Type 1 error. Yap, et al (2011) investigate the creditworthiness of borrowers and compared the accuracy level of LRM and decision tree (DT). The result concludes that there is no significant difference in the performance of LRM and DT. In another study, Blanco, et al (2013) use a multilayer perceptron Neural Network to develop a specific micro-finance credit screening model, they compared the performance of the MLP model against three statistical techniques- linear discriminant analysis, quadratic discriminant analysis, and Logistic Regression Model. The MLP model achieved higher accuracy with lower misclassification errors than others. The finding confirms the superiority of MLP over the parametric statistical techniques.

Furthermore, Bekhet and Eletter (2014) investigate default risk in Jordanian Commercial banks between 2006 and 2011 using 496 cases of creditworthy and non-creditworthy loan applicants. The creditworthiness of borrowers was evaluated using LRM and Radial Basic Function- RBF (a form of Neural Network). It was revealed that both models show promising results and can be concluded that there is no overall best model, Although, LRM performed better than RBF about the overall classification accuracy while the RBF model outperformed the LRM model in screening rejected applicants, identifying potential defaulter and also minimize Type II error.

Bae and Kim (2015) investigate personal credit rating prediction using Ubiquitous data mining (UDM), Bayesian style frequency Matrix (BFM), Multi-Layer Perception (MLP), classification tree (CT), Neural Networks rule extraction (NR) and Logistic Regression (LR) The study employed a dataset of 10,062 from January 2004 to December 2008 and concludes that LR, MLP, CT BFM, NR, and UDM have performance accuracy values of 69.%, 72.7%, 69.1%, 68.6%, 70.6% and 78.6% respectively with UDM rated more efficient than the rest models. Adewusi, Oyedokun, and Bello (2016) assess the classification performance of ANN using data from defaulted accounts obtained from the databases of commercial banks and Primary Mortgage Institutions in Nigeria from 2007 to 2015. The results reveal among others that ANN recorded better classification performance in detecting problem loan applications than non- problem loan applications.

In a more recent research effort, Allam et al (2019) explore the systematic application of neural network-based models versus logistic regression for predicting 30 days all-cause re-admission after discharge from a heart failure (HF) admission. Using a large administrative dataset, the result shows that neural network models and logistic regression have a comparable performance on HF re-admission prediction and that patient timeline data boosts prediction performance. Another study by Christodoulou et al (2019) compares the performance of the logistic regression model (LRM) with machine learning (ML) for clinical prediction modelling. The results show no evidence of the superior performance of ML over LRM.

It is pertinent to note that due to growing uncertainty in society, more studies may be needed to establish the relative appropriateness of ANN and LRM as decision support techniques.

3. Research Methodology

3.1. Variable description, dataset, and variable measurement

Table 1 shows the variables used in the analysis, the dependent variable is loan recovery status which is measured as fully recovered and partially recovered loans. The fully recovered loans refer to loan accounts with optimal performance in terms of recovery while partially recovered loans are loan accounts with poor recovery performance. The dependent variable (loan recovery with a value of 1 and 0 for fully and partially recovered loans respectively). The independent variables were segmented into macroeconomic variables (GDP, INF and INT rates), borrower characteristics (Type of borrower, years of the borrowers' relationship with the bank and history of default of a borrower), loan characteristics (loan duration, loan-to-value ratio, loan size, and loan supervision) and collateral characteristics (value of collateral, Age of collateral asset). In the aspect of variable measurement, GDP, INF and INT rate, loan size, loan duration, loan to value, age of collateral and value of the collateral are measured on continuous scale measurement while recovery status, type of borrower, loan supervision and monitoring of borrower are measured as dummy variables.

Table 1: Variable definition and measurement

Variable	Definition	Measurement
	Dependent Variable	
RECOV_STATUS	Loan recovery status	Dummy (1, if the loan is fully recovered, 0 if otherwise)
	Independent Variables	
Macro-economic indicators		
GDP	The annual growth rate in GDP	Scale (%)
INF Rate	Inflation rates	Scale (%)
INT Rate	Interest rates	Scale (%)
Borrower characteristics		
TYPE_BORR	Type of borrower	Dummy (1, if private borrower, 0 if corporate borrower)
RELAT_BORR	Years of relationship with the bank	Scale (year)
HIST_BORR	Borrower history of default	Dummy (1, if the borrower has a history of default, 0 if otherwise)
Loan characteristics		
LOAN_SIZE	Loan size	Scale (₦ million)
LOAN_DURA	Loan duration	Scale (month)
LTV	Loan-to-Value	Scale (ratio)
SUPERV_LOAN	Loan supervision and monitoring by the lender	Dummy (1, if bank partners with the borrower, 0 if otherwise)
Collateral characteristics		
AGE_COLLA	Age of collateralized real property	Scale (year)
VALUE _COLLA	Value of the collateralized property	Scale (₦ million)

Data on the fourteen variables (1 dependent and 13 independent) were collected on each of the data samples (2,300), obtained from the databases of the eleven Commercial Banks and sixteen Primary Mortgage Institutions in Nigeria. The variable definitions and measurements are contained in Table 1. The variables used are either dichotomous or continuous. Following the procedure of Bekhet and Eletter (2014), the categorical variables were converted into numerical values to be used by the neural network model while all the scale variables were standardized using the rescaling of covariate option in SPSS to improve the network training. SPSS software version 21 was adopted to perform the analysis in the study. These problem loans were later classified as fully-recovered (FR=1192) or partially-recovered (PR=1108) depending on the success achieved after recovery actions were taken. LRM and ANN models were adopted in this study as risk assessment techniques for ANN model development, the dataset is sub-divided into three groups, which are training, validation (to prevent model overfitting) and holdout samples. 1,074 (46.70%) and 613 (26.65%) data samples representing a combined total of 73.35% of the total samples are adopted for both training and validation while the remaining 613 (26.65%) was taken as holdout samples for evaluating the classification accuracy.

3.2. Classification models

Logistic Regression, Factor Analysis, Multiple Regression, Discriminant Analysis, ANN, Vector Support Machine, Decision Tree, K-nearest Neighbor Algorithm, Genetic Algorithm, Fussy Logic and a host of other hybrids have been used as credit risk assessment techniques (Bekhet and Eletter, 2014), the study chooses Logistic Regression and ANN as loan classification models because of the nature of our dependent variable which is dichotomous. Both models have enjoyed increasing use in computing the conditional probability of class membership (Yap, et al, 2011). However, the findings on their relative predictive classification abilities are mixed (Khieu et al, 2012). Logistic Regression is a predictive model widely used in classification problems (Bekhet and Eleter, 2014 and Chen, Xie, Wang, Hong, Bui and Ma, 2017). LRM is a linear regression in which the target variable is a non-linear function of the probability of being 'good' (Thomas, 2000). LRM is a common modelling technique that classifies between two groups using a set of predictor variables (Akkoc, 2012). It has the flexibility of incorporating both qualitative and quantitative factors and it is more efficient in classification problems than other statistical methods (Bandyopahya and Saha, 2009).

The works of Dermine and Neto de Carvalho, (2006) and Khieu, et al (2012) on the appropriateness of Quasi-Maximum Likelihood Estimation (QMLE) to classification and group membership are adopted in the current study. Hence, the effects of the explanatory variables on the criterion variable can be expressed in equations 1, 2, 3 and 4:

$$\text{LR} = \propto_0 + \sum_{j=1}^{n} \beta_j X_{ij} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 1$$

Where LR = loan recovery status, taking the values of 1 if the defaulted loan was fully recovered eventually and 0 if otherwise.

$\propto_0$ is constant; β_j = coefficient associated with the exogenous predictors; X_{ij} = explanatory factors (as contained in Table 2).

If P_I be the probability of recovery of defaulted loans of borrower i.

$$\text{LR} = \frac{e^{\sum_{j=1}^{n} \beta_j X_{ij}}}{1 + e^{\sum_{j=1}^{n} \beta_j X_{ij}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

Transforming equation 1 we derive the odd ratio as;

$$\frac{p_i}{1-p_i} = \exp\left(\propto_0 + \sum_{j=1}^{n} \beta_j X_{ij}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3$$

And the Logit Model is then expressed in equation 4;

$$\text{Log}\left\{\frac{p_i}{1-p_i}\right\} = \propto_0 + \sum_{j=1}^{n} \beta_j X_{ij} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 4$$

On the other hand, Artificial Intelligence (AI) forecasting techniques such as artificial neural networks have been receiving much attention lately (Reddy and Kavitha, 2010). ANN has been described as having the ability to learn like humans by accumulating knowledge through repetitive learning activities (Paliwal and Kumar, 2009). ANN is often given different names: (1) Connectionist models (2) Parallel distributed processing (3) Neuromorphic system and (4) neural computing. ANN is a branch of Artificial Intelligence (AI) in which structures are based on the biological nervous system. It can learn from experience and generalize from previous examples to new problems. ANN is a powerful data modelling tool that can capture and represent complex input/output relationships. It can represent both linear and non-linear relationships.

It is a technique used for classification, prediction, and recognition of a pattern in data when the relationship between the independent and dependent variables is complex and non-linear (Phatilwuttipat et al, 2012).

3.3. ANN model

The model consists of the three main layers as shown in figure 1 below: the input layer, hidden layers, and the output layer.

3.3.1. Input Layer

The activity of the input layer units represents the raw information that is fed into the network. The input layer presents data to the network and the size of the input layer or of nodes is determined by the number of data. The network can be designed to accept sets of input values that are binary or continuous (Ballal, 1999). In this regard, our input variables include GDP, Inflation rate, Interest rate, loan to value ratio, loan amount, loan duration, borrower status, borrower's history of default, year of business relationship with the bank, loan supervision, collateral value and age of collateral.

3.3.2. Hidden layers

The activity of each hidden unit is determined by the activities of the input units and the weights (wij) in the connections between the input and hidden units (wijxi). Hidden layers act as layers of abstraction, pulling features from inputs, they do not directly interact with the outside world (Goh, 1996). The final number of hidden layers requires determination which is by a trial and error process. While Fj is the transfer function from the hidden to output layers as shown in Figure 1.

3.3.3. Output layers

The output layer provides the forecast values. The behavior of the output unit depends on the activity of the hidden units and weights between the hidden and output units.

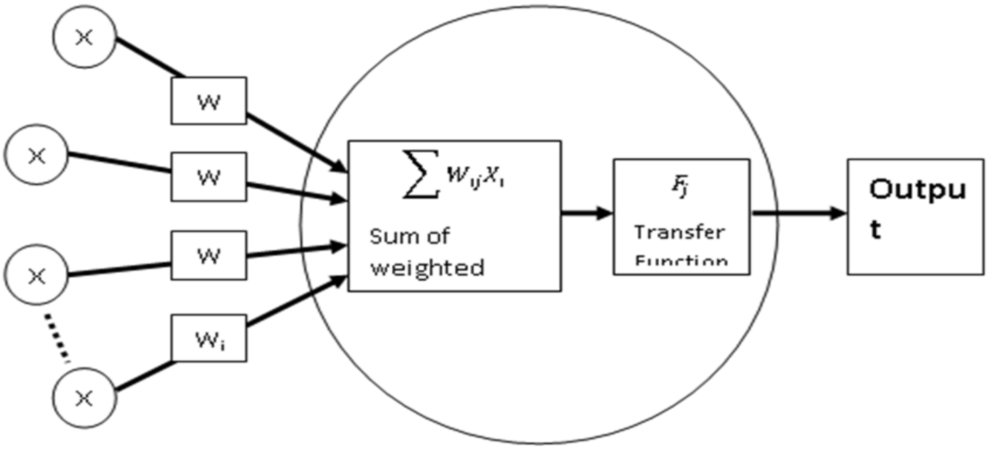

Figure 1: ANN structure of a computational unit
Source: Adapted from James and Carol (2000)

4. Results and Discussion of Findings

4.1. Descriptive statistics of the variables

The result of the analysis is discussed in this section. Table 2 presents the frequency distribution of the samples.

Table 2: Variable frequency distribution

Variables		Frequency	Percent
Rec_status	Partially recovered	1108	48.2
	Fully recovered	1192	51.8
	Total	2300	100.0
Borrower Status	Corporate	891	38.7
	Private	1409	61.3
	Total	2300	100.0
Borrower History of Default	Once in default	582	25.3
	Never defaulted	1718	74.7
	Total	2300	100.0
Bank Participation in Project	Not participated	2140	93.0
	Participated	160	7.0
	Total	2300	100.0

The frequency distribution of the variables in the study is presented in table 2. As shown in table 2, 48.2% (1109) and 51.8 (1192) of the 2,300 loan accounts sampled were partially and fully recovered respectively. A total of 61.3% were private borrowers which indicates that financial institutions extended a greater percentage of loans to the private sector, this may have a multiplier effect on the entire economy.

The majority of these loans (74.7%) were granted to borrowers who never defaulted in their past transactions with the banks while only 25.3% were granted to one-time defaulted borrowers.

Table 3: Descriptive statistics of the variables

	N	Range	Mini.	Max.	Mean	Std. Dev.	Skewness	Kurtosis
Rec_status	2300	1	0.0	1.0	0.5	0.5	-0.1	-2.0
Gross Domestic Product	2300	3.6	6.0	9.6	7.7	1.1	-0.1	-1.3
Inflation	2300	128.3	-52.5	75.8	14.7	37.4	0.3	-0.4
Interest Rates	2300	49.5	-16.1	33.4	0.8	10.1	0.8	0.5
Loan Amount	2300	6999.95	0.1	7000.0	167.5	759.4	7.7	62.8
Value of Collateral	2300	44999.2	0.8	45000.0	352.0	2752.0	15.0	239.2
Loan-to-value	2300	0.99	0.0	1.0	0.7	0.2	-1.3	2.2
Loan Duration	2300	106	6.0	112.0	20.1	12.3	1.4	2.3
Borrower Status	2300	1	0.0	1.0	0.6	0.5	-0.5	-1.8
Borrower History of Default	2300	1	0.0	1.0	0.8	0.4	-1.1	-0.7
Number of Year of Relationship with bank	2300	15	0.0	15.0	4.9	3.3	0.4	-0.4
Bank Participation in Project	2300	1	0.0	1.0	0.1	0.3	3.4	9.5
Age of Collateral	2300	27	1.0	28.0	14.0	5.5	-0.3	-0.4

Table 3 presents descriptive statistics of the sampled variables, from the table, GDP has a mean value of 7.7, Inflation has the highest mean value of 14.7 of all the macroeconomic variables considered. High inflation could pose economic challenges to the entire economy in the country thereby affecting its performance (Bernanke, Laubach, Mishkin, and Posen, 2018). An appropriate policy formulation is needed to stabilize the rate of inflation to achieve a healthier economic growth.

4.2. Test of fitness for logistic regression model

Table 4 presents the test of fitness for LRM, the result of chi-square was presented using an omnibus test which tests the significances of the LRM model. The table provides statistical evidence that there is a relationship between the selected variables and the dependent variable. (loan Recovery). The Chi-square figure of 851.89 is significant at P≤0.00 which indicates that the stated null hypothesis (there is no significant relationship between the predictor variables and the dependent variable) is rejected. Because of the significant relationship at P≤0.00, then the alternative hypothesis is accepted, that there exists a relationship between the selected variables and loan recovery.

Table 4: Logistic regression model test of fitness

	Chi-square	df	Sig.
Step	776.523	12	.000
Block	776.523	12	.000
Model	776.523	12	.000
Omnibus Test			
-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square	
1560.230a	.369	.492	
Hosmer and Lemeshow Test			
Chi-square	df	Sig.	
258.23	8	.270	

Also, Hosmer and Lemeshow test were carried out, the Hosmer and Lemeshow (H-L) are used basically to test the hypothesis that there is no linear relationship between the predictor variables and the log odds of the response variable. Peng et al (2002) assert that when the H-L test is not significant then it could be concluded that the model developed fits the data well. Table 4 also shows the H-L test result which indicates a chi-square figure of 258.23 at P≤0.270 which is not significant. This shows that the model fits the data well and that the null hypothesis should be rejected and the alternate hypothesis, that there is a linear relationship between the weighted combination of the predictor variables and log odds of the criterion variable be accepted. This is consistent with Homer and Lemeshow (2001) that low value and non- significance of the H-L test indicates a good fit to the data.

Table 5: Variables in logistic regression model

Variables	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
GDP	.225	.064	12.551	1	.000	1.253	1.106	1.419
INFLA	-.023	.002	149.102	1	.000	.977	.974	.981
INT	-.086	.007	151.961	1	.000	.917	.905	.930
Loan Amount	-.004	.002	4.045	1	.044	.996	.992	1.000
Value of Collateral	-.002	.001	3.877	1	.049	.998	.995	1.000
Loan-to-value	-1.087	.403	7.266	1	.007	.337	.153	.743
Loan Duration	.016	.006	7.589	1	.006	1.017	1.005	1.028
Borrower Status	.728	.136	28.641	1	.000	2.071	1.586	2.704
Borrower History of Default	.581	.153	14.429	1	.000	1.787	1.325	2.412
Number of Year of Relationship with bank	.143	.021	45.816	1	.000	1.154	1.107	1.202
Bank Supervision and monitoring	-.388	.283	1.880	1	.170	.679	.390	1.181
Age of Collateral	.082	.012	47.269	1	.000	1.085	1.060	1.111
Constant	-2.955	.644	21.044	1	.000	.052		

PredictRecStatus = ((0.225*GDP) - (0.023*INFLA) - (0.086*INT) - (0.004*Loan_Amount) - (0.002*Value_colla) - (1.087*LTV) + (0.016*Loan_duratn) + (0.728*Borrw_Stat) + (0.581*Borr_HisDEFT) + (0.143*Num_yrrela) - (0.388*Bank_partic) + (0.082*Age_Collate)) - 2.955 --5

Table 5 shows the effect of each predictor through the B coefficient. The sign of each B determines the direction of the relationship between each variable and the criterion variable (Bekhet and Eletter 2014). The table shows that all the variables in the model except bank participation in the project, the implication of this is that good attention must be given to those variables that are significant, this will enhance loan recovery in lending institutions. Variables with negative B indicate a decrease in the probability of loan recovery. In this regard, Gross Domestic product, Borrower status, Borrower history of default, Year of business relationship with the lender, loan duration, Age of collateral and value of the collateral will increase the probability of loan recovery while such factors as interest rate, inflation rate, and loan supervision will decrease the probability of loan recovery. To assess the riskiness of each of the 613 prospective borrowers, equation 5 is used in predicting the risk associated with prospective loan applicants, the cut-off value of 0.5 as adopted by Bekhet and Eletter (2014) was used as the threshold value which implies that when $P<0.5$ the applicant is classified into the set of low-risk category while when $P> 0.5$, the applicant is classified into a high-risk category. To determine the riskiness of each of the 613 prospective applicants, their data were imputed into the derived probability equation 5 which results in LRM classification Table 6:

Table 6: LRM classification ability

			PredictedRecStatus		Total
			Partially recovered	Fully recovered	
LRM	Partially recovered	Count	213	80	293
		% within Rec_status	72.7%	27.3%	100.%
	Fully recovered	Count	74	246	320
		% within Rec_status	23.1%	76.9%	100.%

4.3. ANN credit scoring

ANN model was built using the same dataset used in developing the LRM model to compare the accuracy of the two models. The same numbers of independent variables were also examined.

As earlier indicated, LRM and ANN models were adopted in this study as risk assessment techniques, 2,300 loans were used for LRM and ANN. For ANN model development, the dataset is sub-divided into three groups, which include training sample, testing sample (to prevent model overfitting) and holdout sample, 46.7% (1074) and 26.7% (613) of the cases were set aside as training and testing samples respectively while the rest 613 (26.7%) taken as holdout sample for evaluation of classification accuracy as shown in Table 7.

Table 7: ANN model development

			Sample allocation			Total
			Training	Test	Holdout	
Rec_status	Partially recovered	Count	523	292	293	1108
		% within Rec_status	47.2%	26.4%	26.4%	100.%
	Fully recovered	Count	551	321	320	1192
		% within Rec_status	46.2%	26.9%	26.8%	100.%
Total		Count	1074	613	613	2300
		% within Rec_status	46.7%	26.7%	26.7%	100.%

ANN network information is presented in Table 8 showing all the variables. Table 8 indicates ANN layer network information as generated by SPSS version 21. The model selected the best architecture for the network automatically, the input layer houses 12 units which represent the independent variables, the hidden layer has 6 units, the activation function being hyperbolic while the output layer generated 2 units with activation function being softmax. Relative classification performance of both models are presented in Table 9.

Table 8: ANN network information

Input Layer	Covariates	1	GDP
		2	NFL
		3	INT
		4	TYPE_BORR
		5	RELAT_BORR
		6	HIST_BORR
		7	LOAN_SIZE
		8	LOAN_DURA
		9	LTV
		10	SUPERV_LOAN
		11	AGE_COLLA
		12	VALUE_COLLA
	Number of Units		12
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1a		5
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	RECOV_STATUS
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy

a. Excluding the bias unit

Overall classification accuracy, true classification percent correct (sensitivity and specificity) and misclassification error (Type I and Type II error) were evaluated and the result is shown in table 9, this is done to discover the predictive capacity of the two models. The classification accuracy rate is among the most common quantitative measures used in evaluating the predictive accuracy of classification models (Abdou, et al 2007). The overall percentage correct attained by a model may sometimes be misleading, thus, to interpret the model performance in a more meaningful way, the misclassification cost of Type I and Type II errors must be differentiated (Bekhet and Eletter, 2014). For better research precision, the current study adopts the positions of the two authors for a better understanding of the predictive performance of the two models.

The overall classification accuracy of the two models using test samples is 79.1% and 91.2% for LRM and ANN respectively. Percentages of 65% and above are acceptable to practitioners as a standard classification percentage for decision making (Gouvea and Goncalves, 2007). While the two models are adequate as loan classification tools judging from their respective overall accuracy, however, the overall classification capacity of ANN is higher than that of LRM. This finding is consistent with the finding of Youn and Gu (2011) where they claimed that ANN has a better overall predictive classification percentage than LRM. However, the finding of Bekhet and Eletter (2014) is contrary to the finding of the current study as they asserted that LRM had a better overall classification percentage than ANN.

Table 9: Relative classification performance of LRM and ANN

Item	Predicted Value of ANN and LRM						Remarks
	LRM			ANN			
	Partially Recovered Loan	Fully Recovered Loan	Percent Correct (%)	Partially Recovered Loan	Fully Recovered Loan	Percent Correct (%)	
Partially Recovered Loan	TN (213)	FP (80)	72.7	TN (263)	FP (30}	89.8	ANN> LRM (ANN Detects Problem Loan Applicants Better Than LRM)
Fully Recovered Loan	FN (74)	TP (246)	76.9	FN (24)	TP (296)	92.5	ANN> LRM (ANN Detects Good Loan Applicants Better Than LRM)
Overall percentage			74.9			91.2	ANN> LRM (ANN has a better overall accuracy Better than LRM)
Type 1 error			27.3			10.2	ANN> LRM (ANN has a lower value of bad loan applicants wrongly classified as good loan applicants than LRM)
Type 2 error			23.1			7.5	ANN< LRM (ANN has a lower Percentage of good loan applicants wrongly classified as good loan applicants than LRM)

TN=True Negative, FN = False Negative, TP=True Positive, FP= False Positive

In terms of sensitivity analysis (percent correct of good samples), that is, the percentage of good samples classified as good samples. LRM has 76.9% while ANN has 92.5% of good applicants classified as good applicants. This indicates that the ANN model is efficiently better than LRM in classifying ‘good’ applicants as ‘good’ applicants –a difference of 15.6%. This finding is consistent with Odeh et al (2006), ANN displayed the best accuracy in correctly predicting ‘good’ applicants as ‘good’ applicants. In terms of specificity analysis (percent correct of bad samples), this is the ability of a model to classify ‘bad’ applicants as ‘bad’ applicants. LRM obtained 72.7% and ANN generates 89.8% accuracy when classifying ‘bad’ loan applicants as ‘bad’ loan applicants. Again, ANN classification accuracy of ‘bad’ loan applicants classified as ‘bad’ loan applicants is 17.1% higher than that of LRM. This finding agrees with the work of Limsombunchai et al (2005) that ANN has often predicted the borrowers’ worthiness better than LRM. Lastly, misclassification errors of the models were examined from two perspectives, firstly, Type I error which is the wrong classification of ‘bad’ applicants as ‘good’ loan applicants, secondly, Type II error which is the wrong classification of ‘good’ loan applicants as ‘bad’ applicants.

From table 9, the model generated a Type 1 error of 27.3% and 10.2% for LRM and ANN respectively. It thus means that ANN wrongly classified only 10.2% of the total samples as ‘good’ loan applicants while LRM has 27.3% of ‘bad’ applicants wrongly classified as ‘good’ applicants having a difference 17.1%, this implies lending decisions.

The model with a low misclassification Type I error possesses sufficient merit in supporting lending decisions. This corroborates the findings of Youn and Gu (2010) and Adewusi (2015) that most of ANN models can detect Type 1 error much better than LRM model. Also, Type I error holds significant implications on lenders as this may lead to loss of both principal and interest.

Type II error of the two models was also examined; Type II error is the incorrect classification of 'good' loan applicants as 'bad' loan applicants. Table 9 also shows that LRM has 23.1% of 'good' applicants incorrectly classified as 'bad' applicants while ANN has 7.5% of 'good' applicants incorrectly classified as 'bad' applicants. In this regard, LRM still has a higher value of good applicant wrongly classified as bad applicants than ANN, this could impact negatively on lenders' profit as Type II error results in loss of interest to lenders. Most lenders are out to make a profit for the credit granted to borrowers, in such a situation, ANN models might produce a better result than LRM.

5. CONCLUSION

The current study attempts to assess the relative predictive performance of LRM and ANN as credit scoring models. The samples were divided into training/validation and hold-out sub-divisions. The predictive efficiency of the models was assessed in table 9 using 613 data samples of unknown borrowers, it was revealed that the ANN model consistently outperformed LRM using the overall percentage accuracy, percent correct of loan applicants, Type I and Type II errors.

From the foregoing analysis, the two models have performed beyond 65% considered as acceptable standards by lenders (Gouvea and Goncalves, 2007) which indicates that the two models are good as credit scoring models. The study concludes that while LRM and ANN models accuracies are found within the range of 65% and above acceptable to most lenders, the paper observed that ANN is consistently better than LRM as claimed by some studies (Pandey, Choudhury, Jagdev, and Dehuri, 2013; Iturriaga, and Sanz, 2015 and Koh, Tan and Goh, 2015). ANN model may provide a more comfortable platform for lending decision making.

Furthermore, the finding of the current study is limited to the size of the data utilized and obtained from the study area, further study may consider larger samples across some selected countries for comparison purpose.

6. ACKNOWLEDGEMENTS

The author acknowledges the efforts of Dr. T.B. Oyedokun for guiding the use of ANN. The assistance of loan recovery officers of the participating lending institutions is equally and highly appreciated.

REFERENCES

Abdou, H., El-Masry, A.and Pointon, J. (2007). On the applicability of credit scoring models in Egyptian banks. *Banks and Bank systems*, *2*(1), 4-20.

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*, *9*(2), 222-238.

Adewusi, A.O. (2015); Determinants of Loan Recovery in Lagos Metropolis, Nigeria. Doctor of Philosophy Thesis Submitted to the Department of Estate Management, Federal University of Technology, Akure, Nigeria.

Agustia, D., Muhammad, N. P. A., and Permatasari, Y. (2020). Earnings management, business strategy, and bankruptcy risk: evidence from Indonesia. *Heliyon*, *6*(2), 3330-3317.

Akkoç, S. (2012). An empirical comparison of conventional techniques, neural networks and the three-stage hybrid Adaptive Neuro-Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research*, *222*(1), 168-178.

Bae, J. K., & Kim, J. (2015). A personal credit rating prediction model using data mining in smart ubiquitous environments. *International Journal of Distributed Sensor Networks*, *11*(9), 179060.

Bandyopadhyay, A., & Saha, A. (2009). Factors driving Demand and default risk in residential Housing Loans: Indian Evidence. MPRA Paper No. 14352, posted 1. April 2009 04:21 UTC. Online at http://mpra.ub.uni-muenchen.de/14352/

Bekhet, H. A., &Eletter, S. F. K. (2014). Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. *Review of Development Finance*, *4*(1), 20-28.

Berger, A. N., Makaew, T., & Roman, R. A. (2019). Do Business Borrowers Benefit from Bank Bailouts? The Effects of TARP on Loan Contract Terms. *Financial Management*, *48*(2), 575-639.

Bernanke, B. S., Laubach, T., Mishkin, F. S., & Posen, A. S. (2018). *Inflation targeting: lessons from international experience*. Princeton University Press

Blanco, A., Pino-Mejías, R., Lara, J., & Rayo, S. (2013). Credit scoring models for the microfinance industry using neural networks: Evidence from Peru. *Expert Systems with Applications*, *40*(1), 356-364.

Chen, M. C., & Huang, S. H. (2003). Credit scoring and rejected instances reassigning through evolutionary computation techniques. *Expert Systems with Applications*, *24*(4), 433-441.

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T. and Ma, J. (2017). A comparative study of the logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena*, *151*, 147-160.

Dermine, J., & De Carvalho, C. N. (2006). Bank loan losses-given-default: A case study. *Journal of Banking & Finance*, *30*(4), 1219-1243.

Desai, V. S., Crook, J. N., & Overstreet, G. A. (1996). A comparison of neural networks and linear scoring models in the credit union environment. European Journal of Operational Research, 95(1), 24 – 37.

Emel, A. B., Oral, M., Reisman, A., & Yolalan, R. (2003). A credit scoring approach for the commercial banking sector. *Socio-Economic Planning Sciences*, *37*(2), 103-123.

Goh, A. T. (1996). Neural-network modelling of CPT seismic liquefaction data. *Journal of Geotechnical engineering*, *122*(1), 70-73.

Gouvêa, M. A., & Gonçalves, E. B. (2007). Credit Risk Analysis Applying Logistic Regression, Neural Networks, and Genetic Algorithms Models. Paper presented at the Production and Operations Management Society (POMS), Dallas, Texas, U.S.A.

Grima, S., & Thalassinos, E. I. (2020). Financial Derivatives Use: A Literature Review. In *Financial Derivatives: A Blessing or a Curse?* Emerald Publishing Limited.

Guha, R., Sbuelz, A., & Tarelli, A. (2020). Structural recovery of face value at default. *European Journal of Operational Research*, *283*(3), 1148-1171.

Hosmer, D.W. and Lemeshow, S. (2000), *Applied Logistic Regression*, 2nd edition. John Wiley & Sons, Inc., New York.

ISDA (2008). International Swaps and Derivative Association of June.

Iturriaga, F. J. L., & Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of US commercial banks. *Expert Systems with Applications*, *42*(6), 2857-2869.

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, *34*(11), 2767-2787.

Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Systems with Applications*, *37*(9), 6233-6239.

Koh, C. L., Hsueh, I., Wang, W. C., Sheu, C. F., Yu, T. Y., Wang, C. H., & Hsieh, C. L. (2006). Validation of the action research arm test using item response theory in patients after stroke. *Journal of rehabilitation medicine*, *38*(6), 375-380.

Lahsasna, A., Ainon, R. N., & Wah, T. Y. (2012). Enhancement of transparency and accuracy of credit scoring models through the genetic fuzzy classifier. *Maejo International Journal of Science and Technology*, *4*(1), 136-158.

Leah, A. A. (2018). *Factors Affecting the Effectiveness of Bank Credit in Enhancing the Performance of Small and Medium Enterprises in Kenya: A Case of Kisumu City* (Doctoral dissertation, JKUAT-COHRED).

Limsombunchai, V., Gan, C., & Lee, M. (2005). An analysis of credit scoring for agricultural loans in Thailand.

Maldonado, S., Peters, G., & Weber, R. (2020). Credit scoring using three-way decisions with probabilistic rough sets. *Information Sciences*, *5(7)*, 700-714.

Nyasaka, F. O. (2017). *The Relationship between Credit Risk Management Practices and Non-Performing Loans in Kenyan Commercial Banks: A Case Study of KCB Group Limited* (Doctoral dissertation, United States International University-Africa).

Odeh, O. O., Featherstone, A. M., & Sanjoy, D. (2006). Predicting Credit Default in an Agricultural Bank: Methods and Issues. In *Southern Agricultural Economics Association Annual Meeting, Orlando, FL*.

Omondi, R. I., & Jagongo, A. (2018). Microfinance services and financial performance of small and medium enterprises of youth SMEs in Kisumu County, Kenya. *International Academic Journal of Economics and Finance*, *3*(1), 24-43.

Paliwal, M., & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert systems with applications*, *36*(1), 2-17.

Pandey, T. N., Choudhury, D., Jagdev, A. K., & Dehuri, S. (2013). Comparison of classification techniques used for credit risk assessment in financial modelling. *International Journal of Management, IT and Engineering*, *3*(5), 180.

Papke, L., Wooldridge, J. M., 1996. Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. Journal of Applied Econometrics 11, 619-632.

Phatiwuttipat, P., Kongprawechnon, W. and Komolavanij, S. (2012), “Neural network and support vector machine approach for analyzing small and medium-sized enterprises lending decisions in Thailand”, *Proceedings of the 2012 International Conference on Business and Information (BAI 2012)*, *Sapporo, Japan, July 3-5, 2012*, pp. F459-F468

Reddy MVJ, Kavitha B. Neural Networks for Prediction of Loan Default Using Attribute Relevance Analysis. Proceedings of the 2010 International Conference on Signal Acquisition and Processing. 1749237: IEEE Computer Society; 2010: 274-277

Taiwo, J. N., Falohun, T. O., & Agwu, E. (2016). SMEs financing and its effects on Nigerian economic growth. *European Journal of Business, Economics and Accountancy*, *4*(4).

Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting the financial risk of lending to consumers. *International journal of forecasting*, *16*(2), 149-172.

Tsai, C. F., & Wu, J. W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert systems with applications*, *34*(4), 2639-2649.

Wu, C., Guo, Y., Zhang, X., & Xia, H. (2010). Study of personal credit risk assessment based on support vector machine ensemble. *International Journal of Innovative Computing, Information and Control*, *6*(5), 2353-2360.

Xie, T., Yu, H., & Wilamowski, B. (2011). Comparison between traditional neural networks and radial basis function networks. In *Industrial Electronics (ISIE), 2011 IEEE International Symposium on* (pp. 1194-1199). IEEE.

Yap, B. W., Ong, S. H., & Husain, N. H. M. (2011). Using data mining to improve the assessment of creditworthiness via credit scoring models. *Expert Systems with Applications*, *38*(10), 13274-13283.

Youn, H., & Gu, Z. (2010). Predicting Korean lodging firm failures: An artificial neural network model along with a logistic regression model. *International Journal of Hospitality Management*, *29*(1), 120-127.

Zeng, H. (2020). A Comparison Study on the Era of Internet Finance China Construction of Credit Scoring System Model. *Bangladesh Journal of Multidisciplinary Scientific Research*, *2*(1), 1-22.