THE ASSESSMENT OF FINANCIAL INSTITUTIONS' AWARENESS AND APPLICATION OF MACHINE LEARNING TECHNIQUES FOR CREDIT RISK PREDICTION- THE CASE OF NIGERIA

Asongo Abraham Iorkaa[1], Modu Barma[1], Hammandikko Muazu[1]

[1]Department of Statistics and Operations Research, Modibbo Adama University, Yola, Nigeria

Corresponding Author: Asongo Abraham Iorkaa, absongo@yahoo.com

ABSTRACT: The role of credit risk's models for financial institutions and the economy at large in making lending decisions cannot be overemphasised. However, there is much doubts if these scientific models used to predict credit risks are applied by financial institutions in developing countries going by huge nonperforming loans recorded banks in these nations annually. This research examined the awareness and application of machine learning models for credit risk predictions among financial institutions in developing countries with focus on Nigeria. Structured questionnaires were developed for data collection. The statistical package for Social Sciences version 17 (SPSS 17) was used to calculate the mean responses. Mean $(\overline{X})$ was used to answer the research questions. The result shows a very low awareness on machine learning techniques while its application for credit risk prediction is completely not in place among the financial institutions in developing countries. The study therefore recommends the deployment and application of machine learning techniques for credit risk prediction in developing countries which will serve as an objective basis for assessing loan applicants to replace the judgmental system which is mainly a subjective decision that is bond to errors and human interference.
KEYWORDS: Machine Learning, Credit Risk, Artificial Intelligence, models and technique

1. INTRODUCTION

Machine learning is a discipline within artificial intelligence (AI) concerned with the programming of computers to automatically adapt and learn from data or past experience [18]. This can be achieved using an algorithm that specifies a sequence of instructions which transforms the input to output [2]. In machine learning, algorithms are used to distinguish between meaningful and irrelevant patterns in data. Examples of machine learning applications include the provision of accurate medical diagnostics (e.g. breast cancer), real-time map-based monitoring of environmental disasters (e.g. forest fires), and sensory monitoring in the industrial process (e.g. mechanical failure). Banks are relevant to economic development through the financial services they provide. Their intermediation role can be said to be a catalyst for economic growth. The efficient and effective performance of the banking industry over time is an index of financial stability in any nation [21]. The extent to which a bank extends credit to the public for productive activities accelerates the pace of a nation's economic growth and its long-term sustainability. This involves lending which has to do with taking risks, accessing the risk of defaults and movement of interest rate. It is concerned with granting credit facilities to customers which could be either individual or business organization. According to [23], bank credit constitutes an important, if not the main, part of the work performed by commercial banks. Not only do credit facilities represent a sizeable portion of commercial banks assets, but they are also considered to be an important source of bank returns. However, a bank's capacity to grant loans is limited as it is determined by the difference between the total numbers of deposits and the part kept by the bank in the form of liquid funds to meet the demands of clients. Bank credit is also important because of its role in bank activities, i.e. accepting deposits and granting loans, while at the same time observing the possibility of some clients withdrawing all or part of their deposits. In other words, the funds to be granted as credit are those which will not probably be withdrawn by depositors during the period appointed for payment of the loan. Thus, banks grant loans from the funds deposited with them with a higher interest rate than that paid to clients for their deposits to avoid risk resulting any from possibility of insolvency. The decision by a bank to grant credit is clearly important. Because a wrong credit decision will result to non-performing loans eventually [3]. In view of this banks must be aware of the potential risks which may be faced if the client fails to repay. However, despite the advancement in technology and scientific models in predicting credit risks, there are many doubts on whether these scientific models are being applied by financial institutions (FIs) in developing countries.

Therefore, the objective of this research is to find out the awareness and application of machine learning techniques for credit risk prediction among the FIs in developing countries with focus on Nigeria.

2. RELATED LITERATURE

2.1 Expert systems

Machine learning models and artificial intelligence models are also referred to as expert systems [21]. Expert systems are processes of decision making that are performed by a computer program that contains stored knowledge and solves problems in a specific field of human expertise [15]. Expert systems consist of three main parts: (i) a knowledge base which consists of stored rules, (ii) an inference engine that applies the knowledge to the problem, and (iii) a user interface that presents information and questions to the operator and supplies the operator's response to the inference engine. In a credit modeling context, however, expert systems have few examples. [29] showed that the reason for this is because the details of such systems are not usually generic. None has been published to give exact details. However, [25] applied it to small business loans and [19] and [28] applied it to commercial loans.
Artificial intelligence techniques, especially machine learning techniques such as neural networks have been used in default prediction and bankruptcy prediction as well as credit rating analysis [14]. [26] examined the change in credit score migration rates probabilities across business cycles (which may be used as proxy for the systematic risk). [12] looked at credit risk migration and downgrading. One of their findings shows that the least likely borrowers to face credit risk downgrade are "the borrowers at the level of retiring" and the ones that are "actively involved in their business but with children past college age".
[11] used a theoretical model to distinguish between unsuccessful and successful agribusiness loans that included the lender experience and other parameters. Also [11] used primary loan data that contains financial and non-financial variables and applied the logistic regression method to differentiate between successful and non-successful agribusiness loans. He found that unsuccessful loans were associated with "less experienced primary and supervisory loan officers, and repayment projections prepared more often by the borrower or accountant" [11]

2.2. Machine Learning Techiques

The following are summary of some Machines Learning techniques used for credit Risk predictions:

2.2.1 Decision Tree Methodology

Decision trees is one of the Machine Learning Algorithm. According to [21], decision tree approach or the recursive partitioning algorithm (RPA) is a non- parametric, complex and computerized intensive sorting algorithm. The basic idea is to split the sample responses into the new sub-samples that are as homogeneous as possible and as different from each other, and then to repeatedly split the sub-sample into subgroups until it generates the possibility for decision- making. The entire sample is the root node, while the sub-samples are called nodes. Each terminal node is classified as either of 'safe' or 'risky'. Figure1 illustrates a typical RPA.
Assuming we are to decide to give a loan using past credit history, income and loan term. In this Figure terminal subsets are represented by decision boxes while non-terminal subsets are represented by oval shape text boxes. The tree classifier predicts a class for safe or risky loans. If the past credit history was excellent, the classifier predicts safe loan, if the payment of the previous loan was poor, then the classifier look at the income of the customer, if the income is low the classifier predict risky loan, if the income is high the classifier predict safe loan. The process stops when the subsets meet the requirements to be a terminal node of the tree.

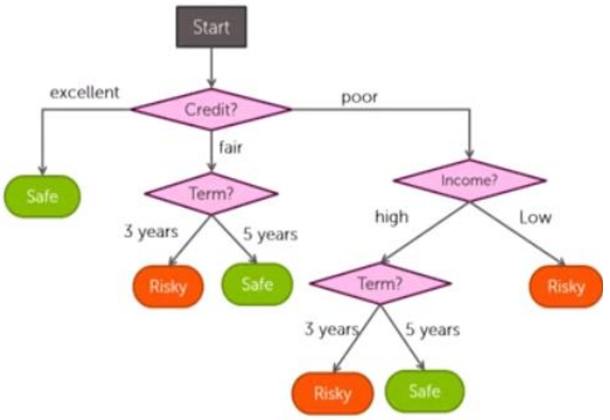

Fig.1. Decision Tree for Credit Risk Prediction Illustration

The objective of decision tree in respect to credit modeling is to develop a sequence of questions or rules which either approve or disapprove to classify loans data into good loans and bad loans. The decision tree itself looks for defining cut-off values for default risk. The best splitting rule is defined as one that allows the decrease in the total impurities of the resulting sub samples compared with the impurity of the root node [21]. The tree is completed when terminal nodes are assigned to all possible default risks and all loans data is finally classified. The outcome of the decision tree is a set of terminal nodes representing the minimisation of the observed expected cost of misclassification of each assignment. There are many methods of splitting used in decision tree approach, including the Kolmogorov- Sminov statistic [27], the basic impurity index, the Gini index, the entropy index, and the maximize half-sum of

squares. Hu and Ansell applied it to a sample of corporate credit data [27]. All these studies have generally shown that the RPA provides better classification precision than many of the other techniques examined.
[17] state that the classification and regression tree (CART) and the analytical neural networks provide an alternative to logistic regression especially when the relationships between dependent and independent attributes are highly nonlinear. Their decision to use CART is based on a previous study that proved that CART is essentially non-parametric. [5] used data from two banks in India that provide agricultural production loans to farmers. They examined two classifiers: logistic regression and decision trees. They state that decision trees have been considered as white boxes, compared and evaluated the accuracy and efficacy of the two classifiers. They acknowledge the universal approximation property of neural networks in credit scoring and state its lack of explanation capability when used for decision-making. Artificial intelligence researchers studied two approaches for classification problems: the "symbolic approach" based on decision trees and the "connectionist approach" which is mainly based on neural networks [5]. Neural networks are not new to economics and finance; in fact, they have been used to solve problems in these areas previously [4]. Problems solved with neural networks can have different forms: classification and discrimination, function approximation and optimization, and series prediction [4].

2.2.2 Nearest-Neighbours Methodology

The nearest-neighbours method (also called pattern recognition) is a nonparametric approach for classifying and was first proposed by Fix and Hodges (1952). It is known to be one of the machine learning algorithms. The idea of this method is to select a metric on the space of application data to measure how far apart any two applicants are (Thomas *et al.*, 2002). In the nearest methods, credit risk is assessed for an applicant from a sample of past applicants as a representative standard and then classified by the proportions of 'good' and 'bad' among the κ nearest applicant 'most similar' points in a training sample. The similarity of points is assessed using an appropriate distance metric. The selection of a suitable distance measure is an important part of the κ nearest neighbour method [16]. The purpose of selecting a metric is to improve classification accuracy by comparing with some specific criterion. Therefore, an applicant is classified as good, if a majority of its neighbours are good; if not, the applicant is classified as bad.
[13] suggested a metric of the form:

$$d(x,y) = \{(x-y)(1+dww)^T(x-y)\}^{\frac{1}{2}} \quad (1)$$

Where X and Y are points in the feature space, I is the identity matrix, D is a distance parameter, and w is a particular direction in the measurement space. The nearest neighbour method has some advantages in credit modeling [13]: (i) this method can build a model of irregularities in the risk function over the feature space because of the nonparametric nature; (ii) the performance of this method is superior to other nonparametric techniques, such as Kernel methods, when the data are multidimensional; (iii) it is straightforward to update dynamically by adding applicants to the design set when their true class becomes known and by dropping older cases. A simple nearest neighbour was applied to credit modeling data by [7] on personal loan applications. [7] found that the method was quicker and far less subjective than many numerical credit modeling methods. [13] also compared a variety of credit modeling models, including nearest neighbour for home loans. [10] compared the performance of some credit modeling models, including nearest-neighbour method and the result showed a good performance for the nearest-neighbour method. [13] used nearest-neighbour method for loan applications compared with modeling methods and found that the performance of the nearest-neighbours method achieved the lowest expected bad risk rate, but it had the advantage over the other methods in relation to classification. However, [6] found the nearest-neighbour method was worse than other methods such as neural networks, discriminant analysis and logistic regression when they applied it to loan data from different countries.

2.2.3 Goal programming Methodology

Goal programming is a branch of mathematical programming that is concerned with the optimal allocation of limited resources to achieve a desired goal by maximizing and minimizing values as well as an efficient part of operations research to solve many practical problems [8].
[20] was the first to study suggested goal programming in classifying problems where there are two groups and there is a separating hyperplane, which can separate the two groups accurately. [8] proposed that goal programming can be applied for discriminant problem when the two groups are not necessarily separable by using objectives such as minimization of the sum of absolute errors or minimizing the maximum error [29].
In a credit modeling context, goal programming aims to determine a relationship between weights, or scores associated with explanatory variables and a scalar that defines a cut-off point between good and bad loans. To solve the problem by linear programming, a popular formulation is as follows:

$Minimize: \ a_1 + a_2 + \cdots + a_{nG+nB}$

$Subject\ to$:

$w_1x_{i1} + w_2x_{i2} + \cdots + w_p\, x_{ip} \geq c - a_i,\ 1 \leq\ \mathrm{I}\ \leq n_{nG}$

$w_1x_{i1} + w_2x_{12} + \cdots + w_px_{ip} \leq c + a_{i,}n_G + 1 \leq$

$i \leq n_G n_B \quad a_i \geq 0, 1 \leq i \leq n_G + n_B$ (2)

Where $a_1 + a_2 + \cdots + a_{nG+nB}$ is the objective function that includes the possible errors (all are positive or zero),
$w_1, w_2, \ldots, w_p$ are the weights that minimize the sum of the absolute values of errors,
$x_{i1}, x_{i2}, \ldots, x_{ip}$ the predictive value of a borrower, i and c the cut-off.

If an applicant in the sample is good, we could require

$$w_1x_{i1} + w_2x_{i2} + \cdots + w_p\, x_{ip} \geq c - a_i$$

However, if an applicant in the sample is bad, we require

$$w_1x_{i1} + w_2x_{12} + \cdots + w_px_{ip} \leq c + a_{i,}$$

[12] suggested a mixed integer programming for classification and showed that it can be expanded to allow variables to be chosen in a stepwise way. However, it has two major disadvantages [29]. First, it takes much longer than goal programming to solve and hence can deal with only very small sample sets with the number of cases in the hundreds. Second, there are often a number of optimal solutions with the same number of misclassifications on the training set but with quite different performances on the holdout sample.

2.2.4 Integer programming

Any decision problem (with an objective to be maximized or minimized) in which the (quantifiable) decision variables must assume non-fractional or discrete values may be classified as an integer optimization problem [4]. If only some of unknown variables are required to be integers, then the problem is called mixed integer programming which prevents a trivial solution. This method overcomes the limitations of linear programming resulting from minimization and maximization of deviation.
In this technique, at least some of the variables will have to be integer (0, 1, 2, etc.).
[13] provided the following model:

$minimize\ L(d_1 + \cdots + d_{nG} + D(d_{nG+1} + \cdots +$

$d_{nG+B})$ (3)

$Subject\ to$

$w_1x_{i1} + \cdots + w_px_{ip} \geq c - Md_i,$

$1 \leq i \leq n_{G,}$

$w_1x_{i1} + \cdots + w_px_{ip} \leq c + md_i,$
$0 \leq d_i \leq 1, 0 \leq d_i \leq 1,\ \ d_i integer$

Where L is the cost of misclassifying a good as a bad, D the cost of misclassifying a bad as a good, M is a positive number, and d_i is a variable that is 1 if a customer in the sample is misclassifying a good and 0 otherwise.
[13] found that the integer model is a better classification model than the goal programming model.
[12] suggested a mixed integer programming for classification and showed that it can be expanded to allow variables to be chosen in a stepwise way. However, it has two major disadvantages. First, it takes much longer than goal programming to solve and hence can deal with only very small sample sets with the number of cases in the hundreds. Second, there are often a number of optimal solutions with the same number of misclassifications on the training set but with quite different performances on the holdout sample.

2.2.5 Genetic Algorithms

Genetic (or evolutionary) algorithms are one of the artificial intelligence methods used in credit modeling. Genetic algorithms were pioneered by [9] who took the same ideas of the general principles of evolutionary natural selection suggested by Charles Darwin and used them on unconstrained optimization problems. The idea of this method is to attempt to simulate the survival of the fitness rule of genetic mutation to develop optimisation algorithms [29].
A basic genetic algorithm represents selecting a population of candidate solutions (called individuals) to a problem. Solutions are represented as strings of genes (called chromosomes).
Genetic algorithms basically assess the performance (called fitness) of each possible solution in each generation and then calculate the fitness of each string to achieve a given objective.

3. METHODOLOGY

The objective of this research was to assess the current procedures applied by financial institutions in Nigeria as a developing country for making credit decisions. To achieve this, questionnaires were designed, validated, pretested and administered to Banks in Nigeria.

3.1 Sample size and sampling procedure

Random sampling method was used to select samples from the 39 registered banks in Nigeria by categories (i.e. deposits money banks, development banks, merchant banks, mortgage banks and microfinance banks). This allowed each Bank to have an equal probability of being picked and each Bank to have an equal chance of being included in the sample. The following formula for sample size of a smaller population was used:

$$n_f = \frac{n}{\left(1+\frac{n}{N}\right)} \quad (4)$$

Where n_f = the desire sample size when the population is less than 10,000,
n = the desired sample size when the population is more than 10,000 (n is obtained by the formula

$$\frac{Z^2\, p_q}{(d^2)}$$

Where Z = the standard normal deviation usually set at 1.96,
p= the proportion in the target population usually set at 0.5,
q= 1-p and d = degree of accuracy desired, usually set at 0.05)
N= the estimate of the population size.

$$n_f = \frac{384}{\left(1+\frac{384}{38}\right)} = 36$$

Therefore, n= 384 (i.e. the desire sample size when the population is above 10,000)
N = 38 (i.e. total number of various categories of Banks registered with CBN as at the time of data collection).
Therefore, a sample size of 36 banks were chosen.

3.2 Instrument of primary data collection

The instrument for primary data collection was structured questionnaire developed by the researcher from literature. It contains two sections, namely section A for information on the personal profile of the respondents while section B is concerned with criteria for granting credit. Items in section B which were the main data sought for in the questionnaire were treated on a five-point scale:

Table 1: Scale for Questionnaire Assessment

Strongly Agree	5 points
Agreed	4 points
Undecided	3 Points
Disagree	2 Points
Strongly Disagree	1 Point

3.3 Validation of the instrument

As outlined in [1], after the questionnaire was vetted and approved by the supervisor, the instrument was sent to Head of lending of Zenith Bank Plc, Yola, Nigeria to validate the questionnaire in terms of clarity of the items, the appropriateness of the Banking terminology and phrases used. The effecting of the observations and suggestions of the validators produced the final copy of the questionnaire which was then administered.

3.4 Reliability of the instrument

In order to obtain a reliable instrument, the final copy of the questionnaire was trial tested on 10 respondents drawn from Standard Microfinance Bank Limited. The test was to determine the reliability coefficient of the instrument which was found to be appropriate [1].

3.5 Data Analysis

Mean $\overline{X}$ was used to answer the research questions. The statistical package for Social Sciences version 17 (SPSS 17) to calculate the mean responses. The study considers the mean appropriate because, according to [22], it is the most accurate and representative measure of central of tendency.
The table of true limits of real numbers adopted from [24] was used as a benchmark for the mean responses as shown in table 2.

Table 2: Table of True Limits of Real Numbers

LIMITS	DESIGNATION
4.50-5.00	Strongly Agree
3.50-4.49	Agreed
2.50-3.49	Undecided
1.50-2.49	Disagree
0.50-1.49	Strongly Disagree

To effect decision, a mean of 3.50 and above was regarded as “Agree”. A mean between 2.50 and 3.49 was regarded as “Undecided” while a mean of 2.50 and below was regarded as “Disagree” (Spiegel and [24].

4. RESULTS AND DISCUSSION

Table 3 below is a description of the results obtained from the data analysis.
Table 3 present data on the mean responses of Bank official regarding the currently Loan process in Nigerian Banks. Looking at the mean of the items, it indicates that the respondents rated all the items agree, except item 10 which the mean responses from Bank officials suggest that they have heard of machine learning techniques in the past though never use it for loan process.

Table 3: Respondents Mean Scores

SN	Items	Mean $(\bar{X})$	Remarks
1	Non-Performing loans leads to Banks' Insolvency	3.60	Agree
2	Decision to Give Loans ins Nigerian Banks is currently Based on Manual Risk Assessment of the Customer	4.51	Agree
3	Disbursement is done only when customer meets those criteria	4.52	Agree
4	Decision to Give Loans ins Nigerian Banks is currently subjective and Based on customer 5Cs	4.39	Agree
5	Subjective Decisions leads to Mistake and high Default	4.43	Agree
6	Use of Scientific models and analytic software during loan process will help Banking industry reduce default	3.57	Agree
7	Use of Machine Learning techniques for Default Prediction will Reduce default	4.48	Agree
8	I have not heard of Machine Learning Before Now	2.48	Disagree
9	I heard of Machine Learning but Never use it for Loan Appraisal	4.35	Agree
10	Need for training Bank staff to apply Machine Learning Techniques in Loan Appraisal.	4.05	Agree
	Total Grand Mean	3.99	Agree

The table clearly proved that the current practice of lending in Nigerian Banks is based on subjective decision (see item 6) and that such decisions subject to mistakes (see item 7). Hence, the necessity of having Machine learning models that will help them in loan prediction (see item 9). It further justifies the imperative and timeliness of this research.

5. CONCLUSIONS

This research has shown that the financial institution in Nigeria as one of the developing economies relies on subjective decision in deciding who is eligible to be granted a credit facility. Subjective decisions are prone to mistakes hence the reason for the increase in nonperforming loans across the sector. The research has further shown that the awareness of machine learning techniques as a scientific tool for credit risk prediction is very low while its application for credit risk prediction among financial institutions in developing economy (Nigeria as a case study) is completely not in place. Furthermore, results from the respondents have shown the willingness of financial institutions' employees to be trained to make use of models that will be able to reduce wrong credit decisions. We therefore recommend the deployment and application of machine learning techniques for credit risk prediction in developing countries which will serve as an objective basis for assessing loan applicants. This will in no small measure help in deciding on whether to grant credit or not and it will replace the judgmental system which is mainly a subjective decision that is bond to errors and human interference.

REFERENCES

[1]. **E. O. Akuezuilo, N.N. Agu** - *Research and statistics in Education and Social Sciences.* Awka: Nuel Centi Publishing and Academic Press Ltd, 2013

[2]. **E. K. Alpaydin** - *Introduction to machine learning*. The MIT Press, Cambridge, MA., USA, 2004

[3]. **F. O. Amah -** *Determinants of Non-performing Loans (NPL) in Emerging Economies: Evidence from Nigerian Banking Industry*. PhD Thesis, University of Nsukka, 2016

[4]. **E. Angelini, D. Giacomo, R. Andrea** - *A Neural Network Approach for Credit Risk Evaluation*. The Quarter Review of Economics and Finance: 733-755, 2008

[5]. **S. Arns, T. Maria, J. Pedro, N. Steiner, Y. Nei, S. Tamio, C. Julio -** *Using Neural Network Rule Extraction for Credit-Risk Evaluation.* International Journal of Computer Science and Network Security, 12(3), 6-17, 2006

[6]. **B. Baesens, T. Van Gestel, M. Stepanova, J. Suyken, J. Vanthienen** - Benchmarking *state-of-the-art classification algorithms for credit scoring.* Journal of the Operational Research Society, 54(6), 627–635, 2003

[7]. **S. Chatterjee, S. Barcun -** *A nonparametric approach to credit screening*. Journal of the American Statistical Association, 65(320)**,** 150–154, 1970

[8]. **B. R. Feiring -** *Linear programming: An Introduction.* Beverly Hills: Sage Publications, 1986

[9]. **E. Fix, J. Hodges -** *Discriminatory analysis, nonparametric discrimination, consistency properties*, Report 4, Project 21-49-004, School of Aviation Medicine, Randolph Field, Taxes, 1952.

[10]. **T. Fogarty, N. Ireson** - *Evolving Bayesian Classifiers for Credit Control - A comparison with other Machine Learning Methods*. IMA Journal of Mathematics Applied in Business and Industry, 5,63–75, 1993

[11]. **R. L. Gallagher** - *Distinguishing Characteristics of Unsuccessful versus Successful Agribusiness Loans*. Spring: Agricultural Finance Review: 19-35, 2001

[12]. **B. Gloy, L. Eddy, A. Michael** - *Credit Risk Migration and Downgrades Experienced by Agricultural Lenders*. Agricultural Finance Review: 1-16, 2015

[13]. **W. E. Henley, D. J. Hand –** A *k-nearest-neighbour classifier for assessing consumer credit risk*. The Statistician, 45(1)**,** 77–95, 1996

[14]. **Z. Huang, H. Chen, C. Hsu, S. Wu** - *Credit rating analysis with support vector machines and neural networks: a market comparative study*. Decision support systems, vol. 37.543–558, 2004

[15]. **O. Hubal, S. Meisser** - *The Swiss Credit Market.* HEC University of Lausanne, 2000

[16]. **Y. J. Jiangsheng** - *Method of k-Nearest neighbors.* Working Paper, Institute of Computational Linguistics Peking University, China, 2002

[17]. **J. Kao, C. Chih** - Mining *the Customer Credit by Using the Neural Network Model with Classification and Regression Tree Approach.* Vancouver.

[18]. **K.K. Kennedy** - *Credit scoring using machine learning*. PhD thesis. Dublin Institute of Technology: doi:10.21427/D7NC7J, 2013.

[19]. **K.K. Leonard** - *Detecting credit card fraud using expert systems*. Computers and Industrial Engineering, 25(1), 103–106, 1993

[20]. **O. S. Mangasarian** - *Linear and nonlinear separation of patterns by linear programming*, Operations Research 13, 444–452, 1993

[21]. **A.H. Mohamed** - *Credit Risk Modeling in Developing Economy: The Case of Libya: PhD thesis.* Griffith University, 2009

[22]. **B.G. Nworgu** - Education Research: Basic issues and Methodology. Ibadan: Wisdom Publishers, 2006

[23]. **J. Ojo, R. Samoye** - The Impact of commercial Bank's non-performing Loans on financial Development in Nigeria. Journal of Economics and Business Vol.9 (1), 87-99, 2013

[24]. **P.A. Omozopkia** - *Analysis of Constraint Militating against Production Work Engineering Trades Cluster of Technical College in Engineering Technology:* Industrial Application, 1(3), 12-16, 2001

[25]. **V.V. Pearson** - *Designing a Pre-screening Scorecard for Small Business Customers*. Journal of Commercial Lending, 77(4), 39–44, 1994

[26]. **J. Phillips, L. Ani** - *Credit Score Migration Analysis of Farm Businesses: Conditioning on Business Cycles and Migration Trends*. Agricultural Finance Review. 1-15, 2004

[27]. **B.K. Roy** - *Application of decision trees for predicting the performance of microfinance institutions.* PhD thesis, University of Reims, 2013.

[28]. **S.S. Shaffer** - *The winner 's curse in banking*. Journal of Financial Intermediation, 7(4), 359–392,1998

[29]. **L. Thomas, B. Edelman, N. Crook** - *Credit Scoring and its Applications.* Philadelphia, PA, Society for Industrial and Applied Mathematics, 2002.