FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 8 No. 6, December, (Special Issue) 2024, pp 163 - 173
DOI: https://doi.org/10.33003/fjs-2024-0806-2893

CREDIT RISK ANALYSIS: AN ASSESSMENT OF THE PERFORMANCE OF SIX MACHINE LEARNING TECHNIQUES IN CREDIT SCORING MODELLING

[1]Abba Bello Muhammad, [1]Ishaq O. Olawoyin, [2]Abubakar Yahaya, [3]Gulumbe, S. U., [1]Muhammad Abdullahi. A. and [1]Iliyasu Abubakar Salisu

[1]Department of Statistics, Aliko Dangote University of Science and Technology, Wudil, Nigeria.
[2]Department of Statistics, Ahmadu Bello University, Zaria, Nigeria.
[3]Department of Statistics, Usman Danfodio University, Sokoto, Nigeria.

*Corresponding authors' email: mrwudil@gmail.com

ABSTRACT

This study checked the credit risk analysis domain, concentrating on assessing the efficacy of six distinct credit scoring methodologies: linear discriminant analysis, logistic regression, artificial neural networks, support vector machine, decision tree and, K-nearest neighbour on microcredit applicant's data. Two performance metrics were used: Area under the receiver operative characteristic curve and, Precision. The results obtained from the experimentation phase reveal distinct performance levels for each technique. Specifically, K-nearest neighbour and artificial neural networks showcase exceptional prowess, yielding an AUC of 0.9833 and 0.9062 and, an impressive precision score of 0.8065 and 1 respectively. In contrast, logistic regression and support vector machine demonstrate a good performance with an area under the curve value of 0.8537 and 0.8532 respectively, on precision metric score, support vector machine showed impressive high performance while logistic regression performed poorly. Linear discriminant analysis and Decision tree exhibit comparatively moderate accuracy scores and achieved an AUC of 0.8494318 and 0.7524 respectively. Thus, we underscore the potential of K-nearest neighbour and Artificial neural networks as a superior method for credit risk analysis, supported by robust performance metrics. Although, all techniques achieve significantly good discriminative power and good precision. The findings advocate for the adoption of modern techniques in credit scoring modelling, positioning K-nearest neighbour and Artificial neural networks as a valuable tool in financial institutions' risk assessment processes.

Keywords: Metrics, Credit scoring, Precision, Techniques, Variables

INTRODUCTION

Microfinance banks and microfinance institutions' activities have grown significantly over time, and they play a crucial role in the economy. Like other companies, the bank faces some risks in its businesses, like credit, operational, and market risks. This research looks into the assessment of credit risk models (i.e., credit scoring), which is one of the most vital areas of study in the world. Given its role in the global financial crisis of 2008 and the subprime mortgage crisis of 2007, credit risk analysis has become more crucial than ever before. Basel II accord of the Basel Committee on Banking Supervision required the establishment of internal rating measures to assess the risk exposure of a financial institution. This results in banks improving their method of credit risk analysis (Bank for International Settlement 2010). In addition, Hand and Henley(1997) assert that banks and Financial institutions should enhance their credit scoring system not only due to the policy but rather the profit that might be obtained due to small improvements in the system.

Microcredit institutions and programs have been developed over the past year to cover a deficit in adequate saving and credit services for poor and small-scale entrepreneurs. Micro credits are recognized as a strategy for resource transfer to impoverished people to promote self-employment, income generation, poverty alleviation or eradication, and reducing the disparity between rich and poor as objective number 10 of the Sustainable Development Goals (United Nations).

Microcredit is a small amount of money lent to a person or group of persons (household or micro-enterprises) usually with zero collateral. Credit risk is a potential loss arising from the failure of some client to meet the obligation of the loan (I.e. loan or bonds will not be repaid either fully or partially). Most micro lending is unsecured (i.e. the loan usually has zero collateral). Credit risk is the single largest risk most financial institutions battle with and it's a result of the possibility that the loan or bond will not be repaid either fully or partially. It can be represented by the factor credit/default risk, loss, and exposure risk.

When a financial institution decides whether to issue a loan to a customer, typically, those customers are labelled as either "Good" or "Bad". Good credit means the one that is likely to meet his financial obligation as and when due. While bad credit refers to the one that has a high likelihood of defaulting (Yap et al., 2011). All the pertinent information regarding the applicants such as economic conditions, marital status, and intentions, is considered when making those decisions.

Credit risk assessment is a process that provides a lender with the necessary tools that would help in making a decision on whether to grant credit to a new applicant or not and how to deal with existing applicants; whether or not to raise their credit limits. Credit risk decisions are a major factor in determining the success of financial institutions due to the enormous losses that result from wrong decisions (Lahsasna et al., 2010). In financial institutions, credit risk assessment is the cornerstone of credit risk management and the decision-making process for loans, (Wu et al., 2010). One broadly used method for dealing with this classification challenge is the Credit Scoring technique.

Credit scoring is the set of techniques and decision models that help lending institutions or bodies in granting credit to their clients with the minimum possible risk. It was first proposed by Fisher (1936) and the only method used was the discriminant and classification method.

Credit scoring techniques were one of the first areas of application of machine learning techniques in economics. Some examples are *k* -nearest neighbours (Henley and Hand, 1996; 1997), support vector machines (SVMs) (Baesens et al., 2003), decision trees (Coffman, 1986; Makowski, 1985; Srinivasan & Kim, 1987), and neural networks (NN) (Desai, Crook, & Overstreet Jr, 1996; Tam & Kiang, 1992; West, 2000; Yobas, Crook, & Ross, 2000). Other credit scoring techniques are: logistic regression, linear regression, and genetic algorithms are largely used in credit scoring given their ability to model extremely complex problems with good results.

The commonality of all these techniques is that they are all data-driven and they have two phases learning/training phase and the testing phase. To this end, this study aims at assessing the efficacy of six credit scoring methodologies: Linear Discriminant Analysis, Logistic Regression, Artificial neural networks, Support Vector Machine, Decision Tree and, K-Nearest Neighbour to identify the most efficient scoring model that will help in minimizing the credit risk in microcredit banks and institutions.

MATERIALS AND METHODS

This study applied linear discriminant analysis, logistic regression, artificial neural network, support vector machine, decision tree and K-nearest neighbour techniques on some financial and nonfinancial factors to identify the best credit scoring model that aids in investment decisions and predicting the creditworthiness of the new client. This research used secondary data obtained from Gombe Microfinance Bank Ltd. The data analysis was done using R-statistical software.

Description of Variable

The variables under consideration in this study are extracted from the client loan application form, creditworthiness appraisal form, and client credit bureau report. The variables are: age, marital status (Categorical), Gender (Categorical), experience in present business (Month), Amount applied, Number of households, Number of dependents, Amount approved, return over investment per annum (ROI PA), Capital assessed, Status of previous loan (Categorical) and current loan status (Categorical).

Preliminary data analysis

Preliminary data analysis has been employed to get the data ready for further analysis; this includes data cleaning, exploratory data analysis, feature selection, multicollinearity testing, and dimensionality reduction (Han and Kamber, 2006).

Data Partition

The data used in this research was divided into training and, testing subsets. The training set is used to learn the pattern in the data, and the test set is used to assess the generalization ability of the model. The training and, testing was carried out using a similar dataset across all the methodologies; linear discriminant analysis, logistic regression, artificial neural network, support vector machine, decision tree and, K-nearest neighbour for model performance comparison as recommended by (Jha, 2007).

Linear Discriminant Analysis Methodology

Consider a linear score function of linear discriminant analysis

$$d_i^l(x) = -\frac{1}{2}\mu_i^1\Sigma^{-1}\mu_i + \mu_i^1\Sigma^{-1}x + log p_i \qquad (1)$$

$$d_i^l(x) = d_{i0} + \sum_{j=1}^{p} d_{ij}x_j + log p_i \qquad (2)$$

$$where\ d_{i0} = -\frac{1}{2}\mu_i^1\Sigma^{-1}\mu_i \qquad (3)$$

Given a sample unit of credit client features $x_1, x_2, \dots. x_p$. The sample unit would be classified into the population that has the largest Linear Score Function.

Prior probabilities: $p_i = p_r(\pi_i); i = 1,2, \dots., k$

Population Means: these will be estimated by the sample mean vectors: $\mu_i = E(X/\pi_i);\ i = 1,2, \dots., k$

Variance-covariance matrix: this is going to be estimated by using the pooled variance-covariance matrix

$\Sigma = Var\left(\frac{X}{\pi_i}\right);\ i = 1,2, \dots., k$

These parameters will be estimated from training data, in which the population membership is known priori.

Logistic Regression Methodology

The logistic regression model according to Park (2013) is given by:

$$y_{ij} = logit\left(\frac{\pi}{1-\pi}\right) = \beta_\circ + \sum_{ij}\beta_i x_{ij} + \varepsilon_{ij} \qquad (4)$$

Where y_{ij} is an indicator variable that takes a value of one if the condition is satisfied, else it takes zero. It represents the loan status of the client i of category j.

$$y_{ij} = \begin{cases} 1 & (Good\ loan).\ if\ \pi \le 0.5 \\ 0 & (Bad\ Loan(Default))\ elsewhere. \end{cases}$$

$where\ \pi\epsilon(0,1)$

ε_{ij} Is the random error term, $\beta_\circ$ is an interception term and β_i is the coefficient of explanatory variable x_{ij}. $\pi = p(y_{ij} = 1)$ Is the probability of Bad/default loan and the term $\left(\frac{\pi}{1-\pi}\right)$ is defined as odds and has formular $\pi = \frac{odds}{1+odds}$ where

$$\pi = e^{\beta X} - \pi e^{\beta X} \qquad (5)$$

$$\pi = \frac{e^{\beta X}}{1+e^{\beta X}} \qquad (6)$$

Also, $X = (1, x_1, x_2 \dots \dots x_n)$ and

$$\beta = (\beta_0, \beta_1 \dots \dots \beta_p), \qquad (7)$$

$i = 1,2, \dots. p\ j = 1,2, \dots. k$.

The probability that client i of category j will default on the loan is:

$p(y_{ij} = 1/XY) = \frac{e^{\beta X}}{1+e^{\beta X}}$ And the probability that a client will fulfilled his loan obligation in due time is $p(y_{ij} = 0/XY) = \frac{e^{\beta X}}{1+e^{\beta X}}$. (Tabachnick 1996).

Logistic Regression R^2

(Hosmer and Lemeshow 1989) described logistic regression R^2as a good analogue to the R^2 in linear regression. The R^2 statistic is the proportion of the variance in the dependent variable that is explained by the prediction variable, the larger R^2 values indicates that more variation is explained in the model. In logistic regression R^2 is estimated by the cox and Snell's R^2, the Nagelkerke's R^2, the Cohen's R^2 or the Mcfdden's R^2 (Veall and Zimmermann 1996).

This work used all the three method: cox and Snell's R^2, the Nagelkerke's R^2 and, the Cohen's R^2 to check the goodness of fit of the logistic regression model.

Artificial Neural Network

An artificial neural network is a non-parametric technique with application in classification, forecasting, pattern recognition and multi-factorial analysis. The network structure was inspired by the human brain and it's adaptive to different environments by learning from experience (Bekesiene et al., 2021).

Determining the Artificial Neural Network Architecture

This research will use a feed-forward Artificial neural network which consists of multiple inputs $(x_1, x_2, \dots, x_n)$ and a binary output $y = 0\ or\ 1$(McCulloch-Pitts neuron). One hidden layer and the number of nodes are determined by the network with a set of nodes that gives the highest AUC value (discriminative ability). The sigmoid activation function and uniform random weight initialization method will also be used, the weights are initialized by drawing random values from a uniform distribution within a specified range.

Support Vector Machine

Support vector machine is a machine learning model used for classification, this work applies Support vector machine on credit data as a single classifier classification model. The Support Vector Machine algorithm is based on the idea of finding the optimal separating hyperplane between classes by maximising the class margin for a given data $[x_i, y_i]_{i=1}^{n}$ where the input is x_i and y_i is the corresponding observed binary class (client creditworthiness).

The maximum margin is a linear classifier that aims at finding the optimal separating hyperplane that divides the data points with the best possible margin, and the margin is the distance between the support vectors (nearest data points) of each class and the hyperplane. The high margin means a low misclassification probability on new data points.

In the case of nonlinear data, the maximum margin classifier may not exist, a kernel transformation is used to produce a dimension.

Suppose $\psi(\cdot)$ is a nonlinear function that maps the input space into a higher dimensional feature space. The separating hyperplane can be represented as:

$$g(x) = \omega^T \psi(x_i) + b = 0 \qquad (8)$$

Where $\omega\ and\ b$ are the normal vector of the hyperplane and the bias which is scalar respectively. The classifier for a linearly separable set in the feature space is as follows.

$$\omega^T \psi(x_i) + b \geq 1 \qquad if\ \ y_i = 1 \qquad (9)$$

$$\omega^T \psi(x_i) + b \leq -1 \qquad if\ \ y_i = -1 \qquad (10)$$

$$y_i(\omega^T \psi(x_i) + b) \geq -1 \qquad for\ i = 1, \ldots., N \qquad (11)$$

To deal with data that are not linearly separable, the equation can be generalised by putting a nonnegative variable $\xi_i \geq 0$.

$$y_i(\omega^T \psi(x_i) + b) \geq 1 - \xi_i \qquad (12)$$

Where the sum of ξ_i can be considered as a misclassification measurement.

According to the structural risk minimization principle, the minimization can be done by the optimization problem below.

$$\text{minimize } \psi(\omega, b, \xi_i) = \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{N}\xi_i \qquad (13)$$

C is a free regularization parameter controlling the trade-off between margin maximisation and tolerable classification error.

Subject to

$$y_i(\omega^T \psi(x_i) + b) \geq 1 - \xi_i \qquad for\ i = 1, \ldots, N; \qquad \xi_i \geq 0 \qquad (14)$$

The support vector machine decision function can be written as:

$$g(x) = sign(\sum_{i=1}^{m} \alpha_i y_i K(x_i, x_j) + b) \qquad (15)$$

Where: $\alpha_i\ and\ \beta_i$ are set of Lagrangian multipliers and also the primal function can be written as

$$\mathrm{L}(\omega, b, \xi_i, \alpha_i, \beta_i) = \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}[y_i\alpha_i(\omega^T\psi(x_i) + b) - 1 - \xi_i] - \sum_{i=1}^{N}\xi_i\,\beta_i \qquad (16)$$

The weight vector ω optimal solution is

$$\omega = \sum_{i=1}^{m} \alpha_i y_i \psi(x_i) \qquad (17)$$

$K(x_i, x_j)$ is the kernel function in the input space that satisfy $K(x_i, x_j) = \psi(x_i) \cdot \psi(x_j)$.

Decision Tree

Decision trees are non-parametric techniques that classify observations by recursively partition the feature space into regions, each region corresponding to a distinct class label using a split criterion (Gini criterion) (Song and Ying, 2015). It's commonly used in credit scoring to fit the data and predict default. The goal is to create a final subset that is homogeneous concerning the group or class variable.

Structure of Decision Tree

The decision tree has nodes, branches and leaves which represent the feature test, the decision tree outcomes and the class label respectively (Lewis, 2000). At each node, the algorithm selects the feature that best splits the data based on a given criteria such as Gini impurity, Entropy or information gain. The process continues until the stopping (Halt) criterion is achieved. (Sandri and Zuccolotto, 2010)

***k* -Nearest Neighbour**

k-nearest neighbour is a nonparametric supervised learning classification with application in both regression and classification. It remains the simplest and most widely used family of lazy learning algorithms, it operates based on the similarity principle (principal of proximity) Peterson (2009), and it classifies a new instance by a majority vote of its k-nearest neighbour in the training test.

***k*-hyperparameter**

The *k*-hyperparameter defines the number of nearest neighbours considered when predicting the class of a given instance. The small *k* makes the model more vulnerable to noise from nearby points as a result of high variance and overfitting. If the *k* is large the model becomes less sensitive to local variations in the data which leads to potential misclassification of instances as a result of high bias and underfitting.

Model Performance

Due to the imbalanced nature of our dataset, this study employed AUC-ROC and precision in evaluating the model's performance as the most suitable metric.

Precision:

$$Precision\ = \frac{TP}{TP+FP} \qquad (18)$$

The area under the receiver operative characteristic curves: Plot TPR against FPR at the different threshold, then join the dots with the line. The area covered below the line is called AUC. The higher the AUC the better the discriminative ability of the model. Hence, the model with the highest AUC will be considered the best. The variable with the highest coefficient (weight) will be considered an important factor in investment and loan decisions.

RESULTS AND DISCUSSION

Descriptive Statistic

Table 1: Descriptive Statistics

	X1	X4	X5	X6	X7	X8	X9	X10
Median	37	120	250,000	5	1	20,000	642.45	35,000
Mean	37.81	131.84	2,7946.97	4.72	2.13	23,322.73	764.02	42615.15
Variance	88.27	8063.57	136636030.26	9.05	10.54	97209042.63	306582.8	897752273.88
Std. Deviation	9.4	89.80	11,689.14	3.01	3.25	9859.46	553.70	29962.51
Coef. Variance	0.25	0.68	0.42	0.64	1.52	0.42	0.72	0.70
Skewness	0.48	0.95	3.00	1.58	2.60	9.96	3.56	6.43
Skewness S.E	2.55	5.01	15.76	8.33	13.66	52.34	18.73	33.82
Kurtosis	0.22	0.56	20.34	8.61	8.92	161.14	22.18	83.53
Kurtosis S.E	0.59	1.47	53.52	22.65	23.47	424.11	58.38	219.83
Normal Test W	0.98	0.91	0.77	0.90	0.68	0.48	0.73	0.64
Normal Test P-Value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

From the above table 1, there is a wide merging between the min and max values of the variables, indicating that the data needs to be scaled before the main data analysis. The values of kurtosis and skewness show that the data is largely skewed and Shapiro Wilk's Test P-value indicates the non-normality of the variables in the dataset.

Linear Discriminant Analysis

The result obtained from the linear discriminant analysis is summarized in the tables below.

Table 2: Linear Discriminant Analysis Confusion Matrix

	Bad	Good
Bad	25	06
Good	13	90

Table 3: Linear Discriminant Analysis Measures of Discrimination and Precision

Precision	0.1389
AUC	0.849

From Table 3 above, a precision of 0.1389 indicates that the linear discriminant analysis model correctly identified approximately 13.89% of the positive instances out of all instances classified as positive. AUC (Area Under the Curve): is a metric commonly used in binary classification tasks to assess the model's ability to distinguish between the two classes. A value of 0.8494318 for AUC indicates that the linear discriminant analysis model has a good level of discrimination, with a high probability of ranking a randomly chosen "Good" instance higher than a randomly chosen "Bad" instance.

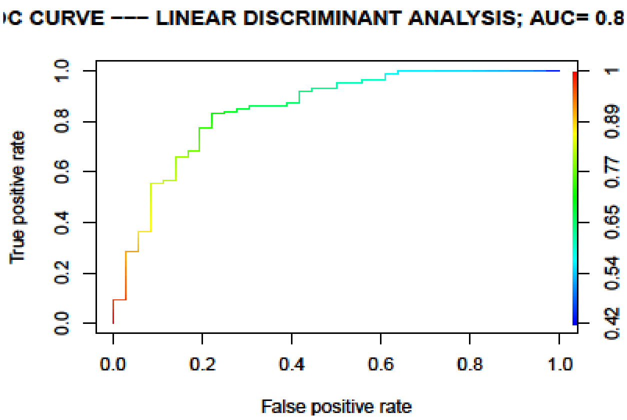

Figure 1: Linear Discriminant Analysis ROC

Therefore, the linear discriminant analysis model shows good discrimination ability (AUC = 0.8494318). However, the precision of 0.1389 suggests that the model's precision needs to be further enhanced.

Logistic Regression

Table 4: Logistic Regression Chi-Square Summary

ModelChiSqr(χ2)	ChiSqrDF	ChiSqrProb
111.2639	5	0.00

The differences between null deviance and residual deviance are the model chi-square. The chi-square value represents the logistic regression model's goodness of fit. It measures the discrepancy between the expected and observed values based on the model. In the case of this work, a chi-square value of 111.26 and a p-value of 0.00 (assuming it's very close to zero) indicate extremely strong evidence that the logistics regression model is highly significant. It implies that the explanatory variables included in the model have a significant impact on the predicted outcomes.

Table 5: Logistic Regression R^2 Summary

Hosmer-Lemeshow	Cox and Snell	Negelkerke
0.2216975	0.5893899	0.6002199

Hosmer-Lemeshow test: The Hosmer-Lemeshow statistic is used to assess the model's calibration (how well the predicted probabilities match the observed probabilities). A higher value indicates better calibration, suggesting that the model's predicted probabilities align well with the observed outcomes. Thus, the Hosmer-Lemeshow statistic of 0.2216975 means a fairly good fit to the data.
Cox and Snell's R^2: The Cox and Snell's R^2 value of 0.5893899 is a measure of the percentage of the variability in the response variable that is explained by the model. A value closer to 1 indicates a better fit of the model. The Cox and Snell's R^2 value of 0.5893899 suggests an acceptable moderate-fit to the data.
Nagelkerke's R^2: The Nagelkerke's R^2 value of 0.6002199 is another measure of the proportion of the variability in the dependent variable explained by the model. Similar to Cox and Snell's R^2, a higher value suggests a better fit of the model. The Nagelkerke's R^2 value of 0.6002199 suggests an acceptable moderate fit to the data.

Table 6: Logistic Regression Confusion Matrix

	Bad	Good
Bad	16	91
Good	14	4

Table 7: Logistic Regression Measures of Discrimination and Precision

AUC	0.8537
Precision	0.1495

AUC (Area Under the ROC curve): the AUC value of 0.8537 represents the discriminative power of the logistic regression model. A higher AUC indicates better discrimination between the positive and negative outcomes. The precision of 0.1495 suggests that the logistic regression model correctly identified approximately 14.95% of the positive cases from all the cases classified as positive.

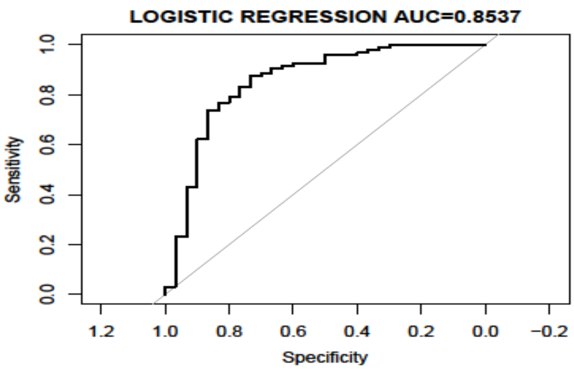

Figure 2: Logistic Regression ROC

Therefore, the logistic regression model has a statistically significant association with the outcome variable (as indicated by the chi-square values and probabilities). The model's goodness of fit (Hosmer and Lemeshow, Cox and Snell, Nagelkerke) suggests an acceptable fit to the data. AUC value of 0.8537 indicates a reasonable level of discrimination power in distinguishing between Good and Bad loans.

Artificial Neural Networks

Variable selection

Variable selection is a preliminary data analysis, usually done before network training to save resources and time (May et al., 2011). Garson (1991) proposed the variable selection method to select fewer variables for the neural network model by discarding the variable with less or zero contribution toward the neural network models' accuracy. The results obtained from the variable selection are shown below:

Table 8: Results of variable selection

Variable	Relative important
X10	-0.1071
X7	-0.0506
X1	-0.0212
X4	-0.0180
X2	-0.0094
X3	0.0000
X11	0.0498
X6	0.0787
X9	0.3405
X5	0.7033
X8	1.0000

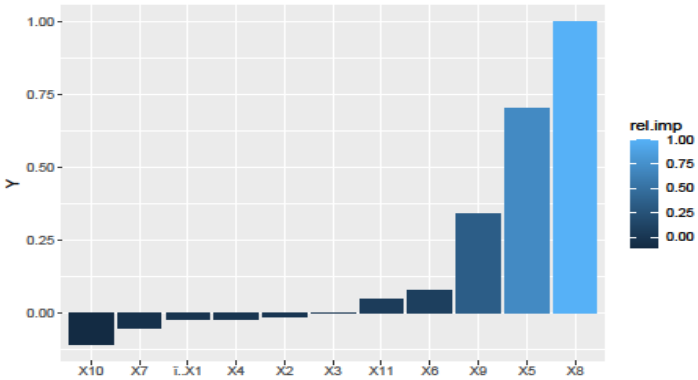

Figure 3: Relative importance of the explanatory variables used in neural network model development

Although Garson algorithms do not provide a fixed threshold for variable selection in artificial neural network models. The threshold for variable incursion depends on the user's discretion. A recommended top-down approach is used in this research, where we start with the most important variables and gradually include the subsequent ones until a satisfactory level of model performance is reached.

Variables with relative importance greater than or equal to 0.05 are considered to have the highest relative importance to the dependent variable (creditworthiness of the loan applicant). The variables are: Amount applied X5, Number of Children X6, Number of dependency X7, Amount Approved X8, ROI PA X9 and Capital Assess X10.

Results

Table 9: Artificial Neural Network Results Summary

Number of Nodes	Error	Steps	Area Under Curve (AUC)
1	51.31	6086	0.91
2	43.72	5825	0.79
3	41.63	56074	0.86
4	28.09	51527	0.82
5	38.22	66481	0.88
6	32.02	86727	0.75

The above table shows the results of an artificial neural network with different numbers of hidden nodes. The artificial neural network with one hidden node was selected as the best, with the highest AUC (discriminative ability) of 0.9062 and a confusion matrix below.

Table 10: Artificial Neural Network Confusion Matrix

	Bad	Good
Bad	25	6
Good	13	90

AUC (Area Under the Curve): A value of 0.9062 for AUC indicates that the Artificial neural network model has a high level of discrimination, with a high probability of ranking a randomly chosen positive instance (belonging to one of the output classes) higher than a randomly chosen negative instance. The Artificial neural network model achieved a high precision of 0.8065, suggesting that it accurately detected around 80.65% of positive occurrences out of all positive instances classified as positive.

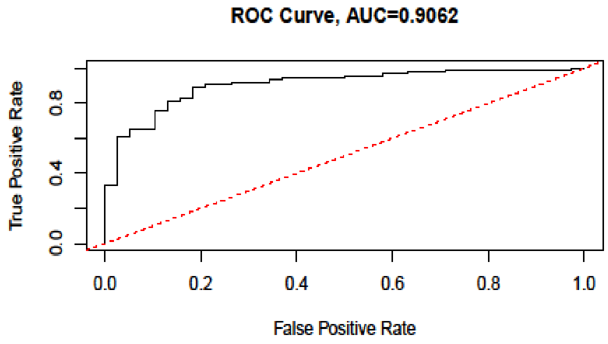

Figure 4: Artificial Neural Network ROC

Therefore, the artificial neural network model with 6 input nodes, 2 output nodes, 1 hidden layer, and 1 hidden node trained using the Rprop algorithm demonstrates high precision (0.8065), and good discriminative ability (AUC = 0.9062). These results indicate that the artificial neural network model is effective in classifying instances and has the potential to make accurate predictions on unseen data. Hence, the model can generalize its results.

The architecture of the final selected model is summarized in the table below.

Table 11: The Artificial Neural Network Architecture of the Final Selected Model

Number of hidden layer	1
Number of hidden nodes	1
Number of input variable	6 (X5,X6,X7,X8,X9 and X10)
Number of output	2 (Y= Good/Bad)
Activation function	Logistic
Algorithm	Rprop
Number of repetition	20
Threshold	0.01

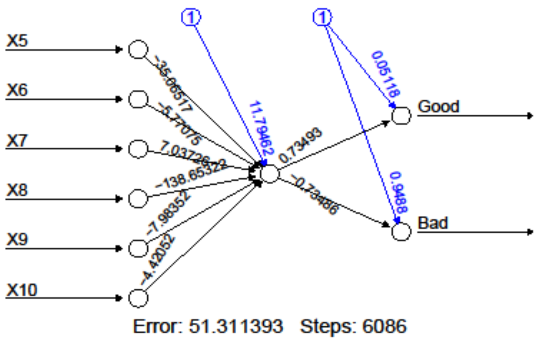

Figure 5: Final selected Artificial Neural Network model Structure

Support Vector Machines

Table 12: Support Vector Machine Confusion Matrix

	Bad	**Good**
Bad	87	19
Good	04	31

Table 13: Support Vector Machine Performance Metrics Scores

AUC	0.8532
Precision	0.8208

From Table 13 above, the AUC (Area Under the ROC curve) value of 0.8532 means the support vector machine has a strong discriminative power. A higher AUC means better discrimination between the creditworthy and non-creditworthy. The support vector machine achieved a high precision of 0.8208, suggesting that it accurately detected around 82.08% of creditworthy clients out of all creditworthy clients.

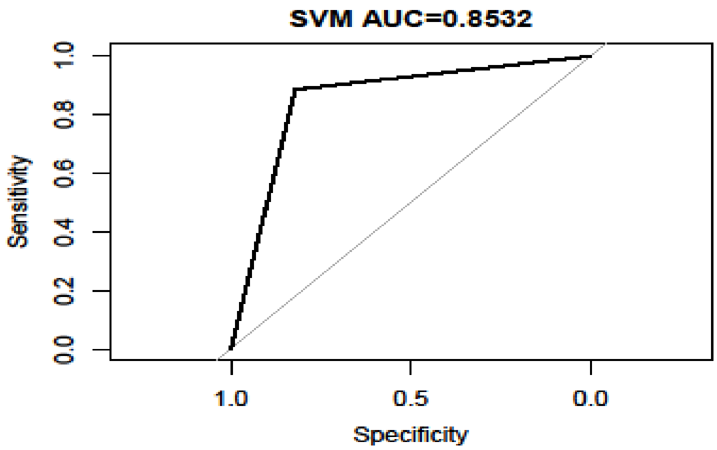

Figure 6: Support Vector Machine ROC

Decision Tree

Table 14: Decision Tree Confusion Matrix

	Bad	**Good**
Bad	94	21
Good	10	22

Table 15: Decision Tree Performance Metrics Scores

AUC	0.7524
Precision	0.8174

From Table 15 above, the AUC (Area Under the ROC curve) value of 0.7524 represents the discriminative power of the Decision tree technique. A higher AUC means better discrimination between the positive and negative outcomes. The precision of 0.8174 means the decision tree technique correctly identified approximately 81.74% of the applicants that are creditworthy as creditworthy.

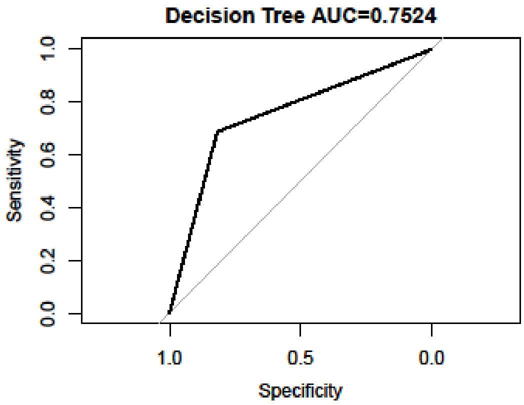

Figure 7: Support Vector Machine ROC

K-Nearest Neighbour

Table 16: Determining *k*-Hyperparameter

***k*-Hyperparameter**	**Accuracy**
k = 1	0.9492
k = 2	0.9322
k = 6	0.9492
k = 7	0.9576
k = 10	0.9746
k = 11	0.9661
k = 12	0.9661
k = 13	0.9576
k = 14	0.9492

The above table 16 shows the different accuracy results of the k-nearest neighbor under the different k-hyperparameter. The k -nearest neighbor with k =10 shows a slightly high accuracy score of 0.9746 compared to other k values. Thus, k=10 is used.

Table 17: *k*-Nearest Neighbors Confusion Matrix

	Bad	Good
Bad	87	0
Good	03	28

Table 18: K-Nearest Neighbors Performance Metrics Score

AUC	0.9833
Precision	1.00
Accuracy	0.9746

From Table 18 above, AUC (Area Under the Curve): A value of 0.9833 for AUC indicates that the K-Nearest Neighbors technique has very excellent discriminative power, with a high probability of ranking a randomly chosen positive instance higher than a randomly chosen negative instance. The K-Nearest Neighbors Performance model achieved a perfect precision of 1.00, suggesting that the technique accurately detected 100% of the credit applicants who are creditworthy as creditworthy.

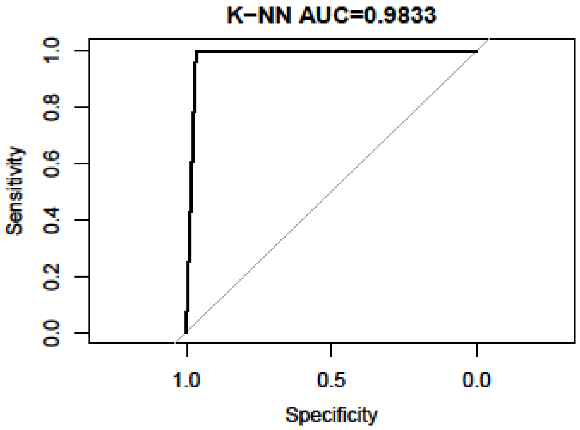

Figure 8: K-Nearest Neighbor ROC

Model Performance Assessment

Based on the results obtained and summarized in Table 19 below, we assess the performance of linear discriminant analysis, logistic regression, artificial neural network, support vector machine, decision tree and K-nearest neighbours in credit classification

Table 19: Models Performance Assessment Summary

Model	AUC	Precision
Linear Discriminant Analysis	0.8494318	0.1389
Logistic Regression	0.8537	0.1495
Artificial Neural Network	0.9062	0.8065
Support Vector Machine	0.8532	0.8208
Decision Tree	0.7524	0.8174
K-Nearest Neighbor	0.9833	1

Therefore, the performance of the models can be ranked as follows:

i. K-nearest neighbor: The K-nearest neighbor technique outperformed all the techniques, demonstrating outstanding performance on all the performance metrics. The perfect precision indicates its ability to perfectly identify creditworthy applicants among the applicants predicted as creditworthy. Additionally, K-Nearest Neighbor achieved the highest AUC value, indicating superior discriminative performance.

ii. Artificial neural network: The artificial neural network model also shows outstanding performance second to K-nearest neighbors, demonstrating high precision in comparison with other techniques (SVM, CART, LR and LDA), indicating a strong ability to correctly identify positive instances. Additionally, it achieved the highest AUC value, indicating superior discriminative performance.

iii. Logistic regression and Support vector machine: The logistic regression and support vector machine show a good AUC value suggesting a good level of discrimination in both techniques. However, on the other metrics (Precision, Accuracy, Recall and F1 score) Support vector machine records good performance while Logistic regression records a low score on all the metrics, suggesting its inability to detect a potential creditworthy applicant.

iv. Linear Discriminant Analysis: The Linear discriminant analysis model has good precision, indicating a strong ability to correctly identify creditworthy applicant. The AUC score showed a good level of discrimination.

v. Decision tree: The decision tree technique has good discriminative ability, precision and accuracy. However, it records the lowest AUC score compared to the other techniques.

CONCLUSION

This research assesses the performance of six credit scoring methodologies (linear discriminant analysis, logistic regression, artificial neural networks, support vector machines, decision tree and K-nearest neighbour) in classifying credit applicants into their correct classes based on the 660 credit client data.

Based on the results obtained, the K-nearest neighbor and artificial neural network techniques emerged as the best model with outstanding discriminative performance (outstanding high accuracy, precision and excellent discriminative ability), followed by logistic regression, support vector machine, linear discriminant analysis and decision tree. Although, all techniques achieve significantly good accuracy except logistic regression. The advanced techniques show more robustness to the non-normality and imbalanced nature of the credit data over the traditional methods. However, the complex structures of modern techniques make it challenging to understand the individual variable contribution to model prediction.

The most contributing variables to default and delinquency are X8 (Amount Approved), X5 (Amount Applied), and X9 (ROI PA), according to Garson's measure of the relative importance of the independent variables. Therefore, K-nearest neighbor and artificial neural networks are robust and effective in classifying instances and have the potential to make accurate predictions on unseen data. Hence, the models can generalise their results.

Further Work

Further studies should aim at using an artificial neural network as their credit evaluation model, which is most suitable to the non-linear nature of the credit data. For generalisation and accuracy of the credit scoring model's result, it's recommended to use larger data and more variables.

The evolution of a borrower's credit quality (how can a client gradually change from a good credit quality to a poor credit quality), the evolution mechanism, and the performance status of a borrower's credit quality at various stages are good research questions recommended for future studies.

REFERENCES

Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003). "Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation". *Journal of Operational research and the management science*, 312-329.

Bank for International Settlement (2010). "Basel Committee on Banking Supervision Publication (BCBS Publication)" https://www.bis.org/bcbs/publications.htm

Bekesiene, S., Smaliukiene, R., & Vaicaitiene, R. (2021). Using artificial neural networks in predicting the level of stress among military conscripts. Mathematics, 9(6). https://doi.org/10.3390/math9060626

Coffman, J. (1986). The proper role of tree analysis in forecasting the risk behavior of borrowers. *Journal of Management Decision Systems, Atlanta, MDS Reports, 3* (4), 7.

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

Fisher, R.A. (1936). "The Use of Multiple Measurements in Taxonomic Problems". *Annals of Eugenics, 7,179-188.* http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x

Garson, G.D. (1991). "Interpreting neural network connection weights". *Articial Intelligence Expert*, Vol.6, No.4, pp.46-51.

Han, J. and Kamber, M. (2006). *Data Mining: Concepts and Techniques* (2nd ed.). San Franscisco: Elsevier, 770 pages.

Henley, W., & Hand, D. (1996). "A k-nearest neighbor classifier for assessing consumer credit risk". *The Statistician, 45* (1), 77–95.

Henley, W. E., & Hand, D. J. (1997). "Construction of a k-nearest neighbor credit-scoring system". *IMA Journal of Mathematics Applied in Business and Industry, 8*, 305–321.

Hosmer, D., & Lemeshow, S. (1989). *Applied Logistic Regression.* New York: John Wiley & Sons, Inc.

Jha, G.K. (2007). *Artificial neural networks and its application*. IARI, NewDehli, girish_iasri@rediffmail.com, pp. 1-10.

Lahsasna, A., Ainon, R.N., Wah, Y. T.,(2010): "Credit Scoring Models Using Soft Computing Methods: A Survey". International Arab Journal of Information Technology *7(2):115-123* *https://www.researchgate.net/journal/International-Arab-Journal-of-Information-Technology-1683-3198*

Lewis, R. J. (2000, May). An introduction to classification and regression tree (CART) analysis. In Annual meeting of the society for academic emergency medicine in San Francisco, California (Vol. 14). San Francisco, CA, USA: Department of Emergency Medicine Harbor-UCLA Medical Center Torrance.

Makowski, P. (1985). Credit scoring branches out. *Credit World, 75* (1), 30–37.

May, R., Dandy, G. and Maier, H. (2011). "Review of Input Variable Selection Methods for Artificial Neural Networks". *Artificial Neural Networks- Methodological Advances and Biomedical Applications, Prof. Kenji Suzuki (Ed.), InTech,* DOI: 10.5772/16004. [online] Available from: http://www.intechopen.com/books/artificial-neural-networks-methodological-advances-and-biomedical/-applications-review-of-input-variable-selection-methods/-for-artificial-neural-networks, [4 February 2017].

Park, H.A. (2013) An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain. *Journal of Korean Academy of Nursing. 43(2):154-164* . https://doi.org/10.4040/jkan.2013.43.2.154.

Peterson, L. E. (2009). K-nearest neighbor. Scholarpedia, 4(2), 1883.

Srinivasan, V., & Kim, Y. H. (1987). "Credit granting: A comparative analysis of classification procedures". *The Journal of Finance, 42* (3), 665–681.

Sandri, M., & Zuccolotto, P. (2010). Analysis and correction of bias in total decrease in node impurity measures for treebased algorithms. Statistics and Computing, 20, 393-407

Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. Shanghai archives of psychiatry, 27(2), 130.

Tabachnick, B.G., and Fidell , L. S. (1996), "*Using multivariate statistics*", (3rd ed.), New York: Harper Collins publishers.

Tam, K. Y., & Kiang, M. Y. (1992). "Managerial applications of neural networks: The case of bank failure predictions". *Journal of Management Science, 38* (7), 926–947.

United Nations Sustainable development goals, goal 10. https://sdgs.un.org/goals/goal10

Veall, M. and Zimmermann, K.(1996). "Evaluating Pseudo-R2's for Binary Probit Models". Quality & Quantity: International Journal of Methodology, *Springer, vol. 28(2), pages 151-164, May. DOI:* https://doi.org/10.1007/BF01102759

West, D. (2000). "Neural network credit scoring models". *Computers & Operations Re- search, 27* (11–12), 1131–1152.

Wu, J., Tennyson, R.D. and, Hsia, T. (2010). "A study of student satisfaction in a blended e-learning system environment". *Journal of* Computers & Education *55(1):155-164* http://dx.doi.org/10.1016/j.compedu.2009.12.012

Yap, B., Ong, S., and Husain, N. (2011). "Using data mining to improve assessment of credit worthiness via credit scoring models". *Expert Systems with Application*

Yobas, M. B., Crook, J. N., & Ross, P. (2000). "Credit scoring using neural and evolutionary techniques". *IMA Journal of Mathematics Applied in Business and Industry, 11*, 111–125.