Journal of Modern Mathematics and Statistics 7 (4): 47-54, 2013
ISSN: 1994-5388

An Application of Artificial Intelligent Neural Network and Discriminant Analyses on Credit Scoring

[1]M.A. Alabi, [2]S. Issa and [3]R.B. Afolayan
[1]Department of Mathematics and Statistics,
Akanu Ibiam Federal Polytechnic Unwana, Afikpo, Ebonyi, Nigeria
[2]Usmanu Danfodiyo University, Sokoto, Nigeria
[3]Department of Statistics, University of Ilorin, Ilorin, Nigeria

Abstract: The research paper deals with credit scoring in banking system which compares most commonly statistical predictive model for credit scoring, Artificial intelligent Neural Network (ANN) and discriminant analyses. It is very clear from the classification outcomes of this research that neural network compares well with linear discriminant model. It gives slight better results than discriminant analysis. However, it is noteworthy that a bad accepted generates much high costs than a good rejected and neural network acquires less amount of bad accepted than the discriminant predictive model. It achieves less cost of misclassification for the data set use in the research. Furthermore, if the final section of this research, an optimization algorithm (genetic algorithm) is proposed in order to obtain better classification accuracy through the configuration of the neural network architecture. On the contrary, it is important to note that the success of the predictive model largely depends on the predictor variables selection to be used as inputs of the model.

Key words: Credit scoring, artificial neural network, discriminant analysis, linear discrimination model, optimization algorithm

INTRODUCTION

The objective of credit scoring models is to help the banks to find good credit applications that are likely to observe obligation according to their age, credit limit, income and marital condition. Many different credit scoring models have been developed by the banks and researchers in order to solve the classification problems, such as Linear Discriminant Analysis (LDA), Logistic Regression (LR), Multivariate Adaptive Regression Splines (MARS), Classification and Regression Tree (CART), Case Based Reasoning (CBR) and Artificial Neural Networks (ANNs).

LDA and ANNs are generally used as methods to construct credit scoring models. LDA is the earliest one used for the credit scoring model. However, the utilization of LDA has often been criticized due to the assumptions of linear relationship between input and output variables which seldom holds and it is sensitive to deviations from the multivariate normality assumption (West, 2000).

Aim and objectives: The aim is to compare the predictive ability of artificial neural network with the linear discriminant models for credit scoring. This aim can be achieved through the following objectives:

- To build a linear discriminant model capable of predicting an applicant credit status
- To build an artificial neural network model capable of identifying an applicant credit status
- To compare and contrast the predictive powers of artificial neural network and linear discriminant models

Predictor variables selection: Credit scoring is performed through credit risk assessment which has mainly three purposes (Colquitt, 2007). First of all and most importantly, it goes through the borrower's probability of repaying the debt by appraising his income, character, capacity and capital adequacy, etc. In addition, it tries to identify borrower's primary source of repayment, especially in the case of extended debt. And finally, it tries to evaluate borrower's secondary source of repayment if the primary source of repayment becomes not available.

Although, credit risk assessment is one of the most successful applications of applied statistics, the best statistical models do not promise credit scoring success, it depends on the experienced risk management practices, the way models are developed and applied and proper use

Corresponding Author: M.A. Alabi, Department of Mathematics and Statistics, Akanu Ibiam Federal Polytechnic Unwana, Afikpo, Ebonyi, Nigeria

of the management information systems (Mays, 1998). And at the same time, selection of the independent variables are very essential in the model development phase because they determine the attributes that decide the value of the credit score (Eq. 1) and the value of the independent variables are normally collected from the application form. It is very significant to identify which variables will be selected and included in the final scoring model.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... \beta_n x_n \quad (1)$$

Where:
Y = The dependent variable
$x_{i,s}$ = The independent variables

Selection of the predictor variables should not be based only on statistical analysis other points have to be noted also. For example, variables those are expensive or time consuming to obtain like debt burden should be excluded. Moreover, variables that are influenced by the organization itself like utilization of advanced cash, should also be excluded (Mays, 1998).

MATERIALS AND METHODS

The data for this write-up was collected from a sample of 200 applicant on credit scoring, extracted from the application form of First Bank of Nigeria plc, Kontagora. The methods used for credit worthy are linear discriminant analysis and artificial neural network.

Linear discriminant analysis: LDA was first proposed by Fisher (1936) as a classification technique. It has been reported so far as the most commonly used technique in handling classification problems (Lee and Jung, 2000). In the simplest type of LDA, two group LDA, a Linear Discriminant Function (LDF) that passes through the centroids (geometric centres) of the two groups can be used to discriminate between the two groups. The LDF is represented by Eq. 2:

$$LDF = a + b_1 x_1 + b_2 x_2 + L + b_p x_p \quad (2)$$

Where, a is a constant and b_1 to b_p are the regression coefficients for p variables. LDA can also be applied in other areas, such as business investment, bankruptcy prediction and market segment (Kim *et al.*, 2000).

Neural network: A Neural Network (NNW) is a mathematical representation inspired by the human brain and its ability to adapt on the basis of the inflow of new information. Mathematically, NNW is a non-linear optimization tool. Many various types of NNW have been specified in the literature.

The NNW design called Multilayer Perceptron (MLP) is especially suitable for classification and is widely used in practice. The network consists of one input layer, one or more hidden layers and one output layer, each consisting of several neurons. Each neuron processes its inputs and generates one output value that is transmitted to the neurons in the subsequent layer. Each neuron in the input layer (indexed i = 1,…, n) delivers the value of one predictor (or the characteristics) from vector x. When considering default/non-default discrimination, one output neuron is satisfactory. In each layer, the signal propagation is accomplished as follows. First, a weighted sum of inputs is calculated at each neuron: The output value of each neuron in the proceeding network layer times the respective weight of the connection with that neuron. There are 2 stages of optimization. First, weights have to be initialized and second, a nonlinear optimization scheme is implemented. In the first stage, the weights are usually initialized with some small random number. The second stage is called learning or training of NNW. The most popular algorithm for training multilayer perceptrons is the back-propagation algorithm. As the name suggests, the error computed from the output layer is back-propagated through the network and the weights are modified according to their contribution to the error function. Essentially, back-propagation performs a local gradient search and hence its implementation, although not computationally demanding, it does not guarantee reaching a global minimum. For each individual, weights are modified in such a way that the error computed from the output layer is minimized. The BP network used herein has an input layer, an intermediate hidden layer and an output layer. The BP-based credit scoring method is succinctly illustrated in Fig. 1. The input nodes represent

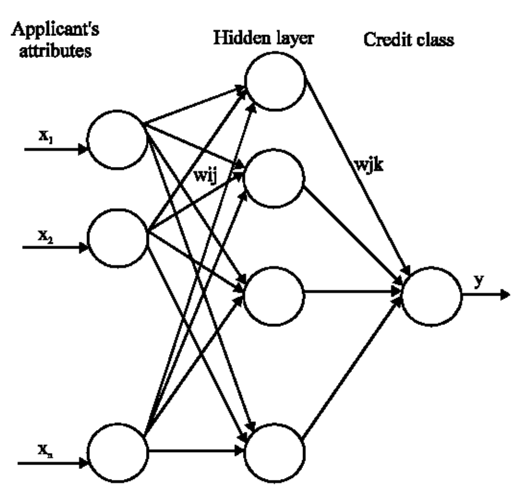

Fig. 1: The BP-based credit scoring architecture

the applicant's characteristics and output node represents the identified class (say 1 for rejected and 2 for accepted). The BP learning involves three stages: The feed-forward of the input training pattern, the calculation of the associated error and the adjustment of the weights. After the network reaches a satisfactory level of performance, it learns the relationships between independent variables (applicant's attributes) and dependent variable (credit class). The trained BP network can then be adopted as a scoring model to classify the credit as either good or bad by inserting the values of applicant's attributes.

Wilks' lambda test for significance of canonical correlation

Hypothesis canonical correlation: H_0 is; there is no linear relationship between the 2 sets of variables. H_1 is; there is linear relationship between the two sets of variables. Test statistic:

$$\lambda = \frac{|W|}{|W+H|}$$

Where:

W = Residual variance
H = Variance due to linear relationship
W+H = Total variance

Decision rule: Reject H_0 if $p<0.05$ otherwise accept H_0 at the 5% level of significance.

Chi-square test

Hypothesis for Chi-square test: H_0 is; the 2 variables are independent. H_1 is; the 2 variables are not independent. Test statistic:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij} - e_{ij}\right)^2}{e_{ij}}$$

Where:

O_{ij} = Observed value
e_{ij} = Expected value

Table 1: Credit dataset description

Variables	Type	Scale	Description
Attribute1	Input variable	Scale	Age of the applicant
Attribute2	Input variable	Nominal	Sex of the applicant
Attribute3	Input variable	Nominal	Ownership of residence
Attribute4	Input variable	Nominal	Marital status
Attribute5	Input variable	Nominal	Qualification
Attribute6	Input variable	Nominal	Employment status
Attribute7	Input variable	Nominal	Employment classification
Attribute8	Input variable	Scale	Length of service
Attribute9	Input variable	Scale	Salary
Attribute10	Input variable	Nominal	Application request
Attribute11	Input variable	Scale	Amount request
Attribute12	Input variable	Scale	Credit amount
Attribute13	Input variable	Scale	Proposed tenor in month
Attribute14	Input variable	Nominal	Other borrowing
Attribute15	Output variable	Nominal	Status of the credit applicant

Decision rule: H_0 reject if $p<0.05$ otherwise accept H_0 at the 5% level of significance (Table 1).

RESULTS AND DISCUSSION

Data collection and analysis: The dataset contains 200 cases, 163 applicants are considered as creditworthy and the rest 37 are treated as non-creditworthy. Data preparation allows identifying unusual cases, invalid cases, erroneous variables and incorrect data values in dataset (Table 2).

A real world credit dataset is used in this research. The dataset is extracted from the application forms of First Bank of Nigeria, plc Kontagora. It is referred to as credit dataset. After preparing the dataset, it is used in the subsequent sections for conducting the analysis with neural network and discriminant analyses.

The dataset contains 200 cases, 163 applicants are considered as creditworthy and the rest 37 applicants are treated as non-creditworthy. The dataset holds 15 variables altogether. Among the variables, 9 variables are categorical and the rest 6 variables are numerical. Moreover, there are 14 independent variables (input variables) and 1 dependent variable (output variable) in the dataset.

Measurement of model performance in discriminant analysis: Discriminant analysis is a statistical technique to classify the target population (in this study, credit card applicants) into the specific categories or groups (here, either creditworthy applicant or non-creditworthy applicant) based on the certain attributes (predictor variables or independent variables) (Plewa and Friedlob, 1995). Discriminant analysis requires fulfilling definite assumptions, for example assumption of normality, assumption of linearity, assumption of homoscedasticity, absence of multicollinearity and outlier but this method is fairly robust to the violation of these assumptions (Meyers *et al*., 2005). Here in this study, it is assumed that all required assumptions are fulfilled to use the predictive power of the discriminant analysis for classification of the applicants. At this point, creditworthiness is the dependent variable (or grouping variable) and the rest 14 variables are the independent variables (or input variables).

A discriminant model is identified as useful if there is at least 25% more improvement achievable over by chance accuracy rate alone. By chance accuracy means that if there is no relationship between the dependent variable and the independent variables, it is still possible to

Table 2: SPSS output; model test Wilks' Lambda

Test of function (s)	Wilks' Lambda	χ^2	Df	Sig.
1	0.501	131.995	14	0.000

Table 3: Canonical correlation

Function	Eigen value	Percentage of variance	Cumulative (%)	Canonical correlation
1	0.996[a]	100	100	0.706

[a]First 1 canonical discriminant functions were used in the analysis

achieve some percentage of correct group membership. Here by chance, accuracy rate is 70% ($0.815^2+0.185^2$) and 25% increase of this value equals to 87.5% (1.25×70%) and the cross validated accuracy rate is 88.5%. Hence, cross validated accuracy rate is greater than or equal to the proportional by chance accuracy rate, it is possible to declare that the discriminant model is useful for the classification goal. Moreover, Wilks' lambda is a measure of the usefulness of the model. The smaller significance value indicates that the discriminant function does better than chance at separating the groups (Table 3).

Here, Wilks' lambda test has a probability of <0.000 which is less than the level of significance of 0.05 means that predictors (i.e., independents variables, such as age of applicant, sex, salary and length of service, etc.) significantly discriminate the groups. This provides the proportion of total variability not explained, i.e., 50.1% unexplained.

A canonical correlation of 0.706 suggests that the model explains 49.84% (i.e., 0.706^2) of the variation in the grouping variable, i.e., whether an applicant is a creditworthy or not.

Importance of independent variables in discriminant analysis: One of the most important tasks is to identify the independent variables that are important in the predictive model development (Table 4).

It can be identified from the structure matrix that the predictor variables strongly associated with the discriminant model are the age of the applicant, length of service and other borrowing. The structure matrix shows the correlations of each variable with each discriminant function. These pearson coefficients are structure coefficients or discriminant loadings. They serve like factor loadings in factor analysis. Generally just like factor loadings, 0.30 is seen as the cut-off between important and less important variables. The interpretation of the standardized discriminant function coefficients is like that in multiple regressions. It provides an index of the importance of each predictor like the standardized regression coefficients did in multiple regression. The sign indicates the direction of the relationship. Other borrowing score is the highest predictor follow by age of applicant and length of service is the least importance of a predictor. The canonical discriminant function coefficients (unstandardized coefficients) are used to create the discriminant function (equation). It operates just like regression equation.

Table 4: Discriminant model

	Functions		
Variables	Structure matrix	Standardized discriminant coefficients	Canonical discriminant coefficients
Age	0.473	0.434	0.062
Sex	-0.041	-0.025	-0.055
Owner ship	-0.058	-0.033	-0.067
Marital status	-0.095	-0.131	-0.173
Qualification	0.051	0.015	0.011
Employment status	-0.010	0.012	0.029
Employment classification	-0.119	-0.267	-0.320
Length of service	0.467	0.158	0.020
Salary	-0.148	-0.412	0.000
Application request	-0.061	-0.163	-0.380
Amount request	0.040	1.253	0.000
Credit amount	-0.153	-1.031	0.000
Propose tenure	-0.092	-0.137	-0.023
Other borrowing	0.554	0.688	1.494
Constant	-	-	-3.173

Table 5: SPSS output; group statistics

Status of the credit applicant	Mean	SD	Valid N (listwise) Unweighted	Weight
Good group				
Age of applicant	45.9141	6.90267	163	163
Length of service	20.1166	7.80186	163	163
Other borrowing	1.1043	0.37865	163	163
Bad group				
Age of applicant	54.4054	7.51456	37	37
Lenght of service	29.6757	8.85078	37	37
Other borrowing	1.7568	0.72286	37	37

There are specific characteristics determined by the discriminant model for the two groups (creditworthy applicants and non-creditworthy applicants). Based on these given characteristics, an applicant is awarded good and another one is the bad. These characteristics differ between the two groups. For example on an average, a good customer has the age of at most 45 years; spend at most 20 years in service and not borrowing from cooperative, employers and other bank. The most important variables, identified in structure matrix before are in Table 5.

Characteristics of the creditworthy applicants: On the other hand, the bad group holds some certain characteristics in contradictory with the good group. For example on an average, non-creditworthy applicants possess at least 54 years of the age; spend atleast 29 years in service and having borrowing from employer. Only the most important characteristics are shown in Table 5.

Characteristics of the non-creditworthy applicant

Neural network structure (architecture): Neural network model is constructed with the multilayer perceptron algorithm. In the architectural point of view, it is a 14-10-1 neural network means that there are total 14 independent variables, 2 neurons in the hidden layer and 1 dependent (output) variable. SPSS software is used. SPSS procedure can choose the best architecture automatically and it builds the network with one hidden layer. It is also possible to specify the minimum (by default 1) and maximum (by default 50) number of units allowed in the hidden layer and the automatic architecture selection procedure finds out the best number of units (6 units are selected for this analysis) in the hidden layer. Automatic architecture selection uses the default activation functions for the hidden layer (hyperbolic tangent) and output layers (softmax) (Fig. 2). Predictor variables consist of factors and covariates. Factors are the categorical dependent variables (8 nominal variables) and the covariates are the scale dependent variables (6 continuous variables). Moreover, standardized method is chosen for the rescaling of the scale dependent variables to improve the network training. Further, 70% of the data is allocated for the training (training sample) of the network and to obtain a model and 30% is assigned as testing sample to keep tracks of the errors and to protect from the overtraining. Different types of training methods are available like batch, online and minibatch. Here, batch training is chosen because it directly minimizes the total error and it is most useful for smaller datasets. Moreover, optimization algorithm is used to estimate the synaptic weights and scaled conjugate gradient optimization algorithm is assigned because of the selection of the batch training method. Batch training method supports only this algorithm. Additionally, stopping rules are used to determine the stopping criteria for the network training. According to the rule definitions, a step corresponds to iteration for the batch training method. Here, maximum step is allowed if the error is not decreased further. Here, it is important to note that to replicate the neural network results exactly, data analyzer needs to use the same initialization value for the random number generator, the same data order and the same variable order in addition to using the same procedure settings.

Measurement of model performance in neural network structure (architecture): The following model summary table displays information about the results of the neural network training. Here, cross entropy error is displayed because the output layer uses the softmax activation function. This is the error function that the network tries to minimize during training. Moreover, the percentage of incorrect prediction is equivalent to 18.88% in the training samples. So percentage of correct prediction is nearer to 94.4% that is quite high. If any dependent variable has scale measurement level, then the average overall relative error (relative to the mean model) is displayed. On the other hand, if the defined dependent variables are categorical, then the average percentage of incorrect predictions is displayed in Table 6.

Importance of independent variables: Table 7 performs an analysis which computes the importance and the normalized importance of each predictor in determining the neural network. The analysis is based on the training and testing samples. The importance of an independent variable is a measure of how much the network's model-predicted value changes for different values of the independent variable. Moreover, the normalized importance is simply the importance values divided by the largest importance values and expressed as percentages. From Table 7, it is evident that amount request contributes most in the neural network model construction followed by credit amount, annual estimated earnings of the applicants, ages in years of the applicants, other borrowing, length of service, propose tenor in month, etc.

Table 7 shows importance on how the network classifies the prospective applicants. But, it is not possible to identify the direction of the relationship

Table 6: SPSS output; model summary

Variables	Types	Values
Training	Cross entropy error	18.882
	Percent incorrect predictions	5.6%
	Stopping rule used	1 consecutive step (s) with no decrease in error
	Training time	0:00:00:265
Testing	Cross entropy error	5.383
	Percent incorrect predictions	1.8%

Dependent variable: Applicant status; error computations are based on the testing sample

Table 7: SPSS output; independent variable importance

Variables	Importance	Normalized importance (%)
Gender of applicant	0.018	8.1
Ownership of residence	0.004	2.0
Marital status	0.027	12.6
Qualification of applicant	0.031	14.2
Employment status	0.019	8.9
Employment classification	0.021	9.4
Application request	0.029	13.5
Other borrowing	0.088	40.4
Age of applicant	0.115	53.1
Length of service	0.079	36.4
Applicant salary	0.128	58.7
Amount request	0.218	100.0
Credit amount	0.151	69.2
Proposed tenor in month	0.072	33.0

Important variables identified by the neural network model

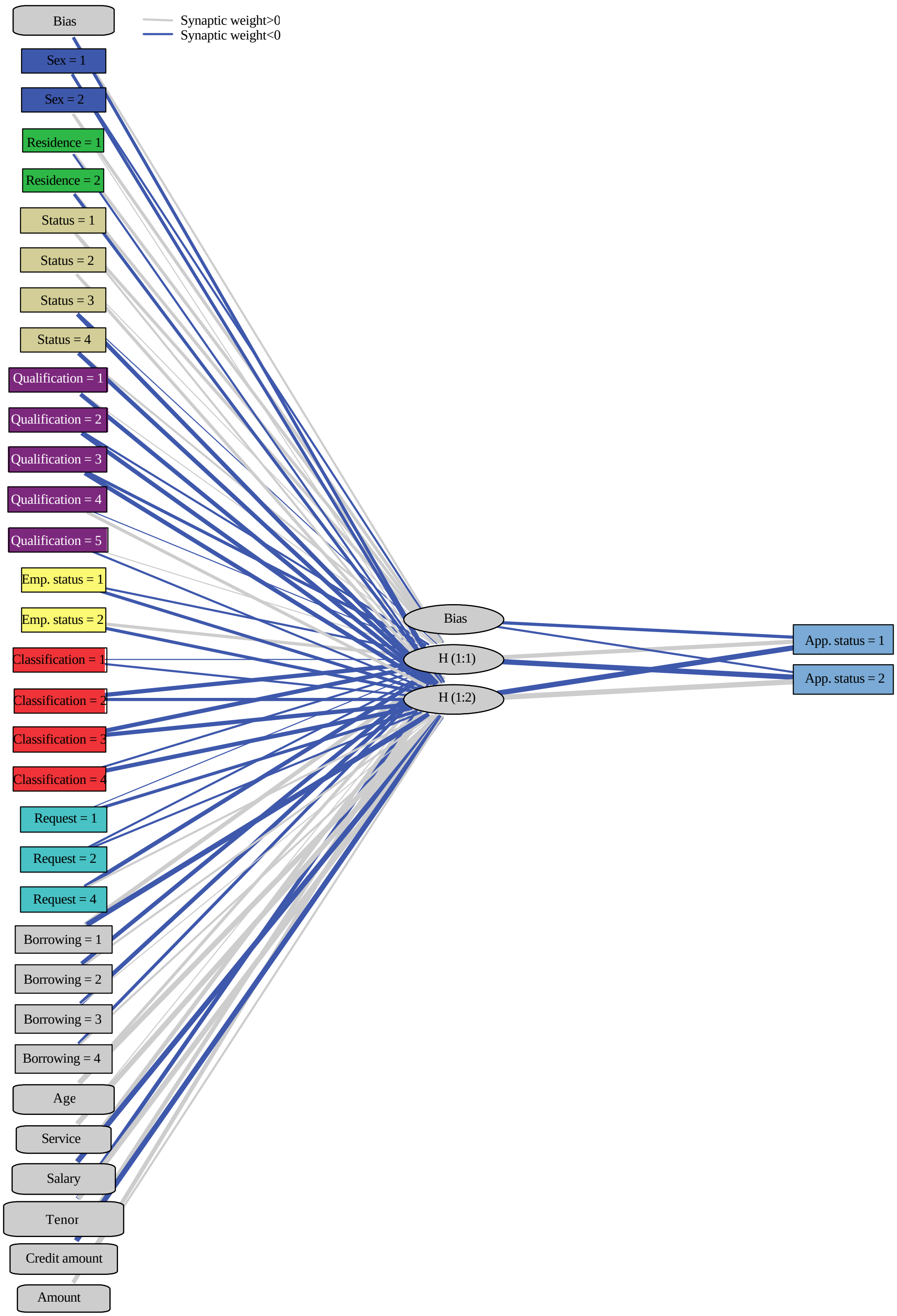

Fig. 2: Hidden layer activation function; hyperbolic tangent output layer activation function, softmax

between these variables and the predicted probability of default. This is one of the most prominent limitations of the neural network. So, statistical models will help in this situation.

Comparison of the model's predictive ability

Discriminant analysis: In the discriminant analysis model development phase, a statistically significant model is derived which possess a very good classification accuracy capability. In Table 8, it is shown that the discriminant model is able to classify 152 good applicants as good group out of 163 good applicants. Thus, it holds 93.3% classification accuracy for the good group. On the other hand, the same discriminant model is able to classify 30 bad applicants as bad group out of 37 bad applicants. Thus, it holds 81.1% classification accuracy for the bad group. Thus, the model is able to generate 91.0% classification accuracy in combined groups.

Artificial neural network: In the artificial neural network model development stage, a predictive model is derived which enjoys a very good classification accuracy capability. In Table 9, it is shown that the neural network model is able to classify 115 good applicants as good group out of 116 good applicants. Thus, it holds 99.1% classification accuracy for the good group. On the other hand, the same neural network model is able to classify 21 bad applicants as bad group out of 28 bad applicants. Thus, it holds 75.0% classification accuracy for the bad group. Thus, the model is able to generate 94.4% classification accuracy for the both groups. Here, the training sample is taken into account because statistical models do not use testing sample.

Table 8: SPSS output; classification results predictive ability of the discriminant model

Classification	Attribute 15	Predicted group membership 1	2	Total
Original[a]				
Count	1.00	152.0	11.0	163
	2.00	7.0	30.0	37
Percentage	1.00	93.3	6.7	100
	2.00	18.9	81.1	100
Cross-validated[b]				
Count	1.00	150.0	13.0	163
	2.00	10.0	27.0	37
Percentage	1.00	92.0	8.0	100
	2.00	27.0	73.0	100

[a]0.91.0% of original grouped cases correctly classified; [b]0.88.5% of cross-validated grouped cases correctly classified

Table 9: SPSS output; classification results predictive ability of the artificial neural network

Samples	Observed	Predicted 1	2	Correct (%)
Training	1	115.0	1.0	99.1
	2	7.0	21.0	75.0
	Overall (%)	84.7	15.3	94.4
Testing	1	47.0	0.0	100.0
	2	1.0	8.0	88.9
	Overall (%)	85.7	14.3	98.2

Dependent variable: Applicant status

CONCLUSION

Appropriate predictor variables selection is one of the conditions for successful credit scoring models development. This study reviews several considerations regarding the selection of the predictor variables. Moreover using the multilayer perceptron algorithm of neural network, network architecture is constructed for predicting the probability that a given customer will default on a loan. The model results are comparable to those obtained using commonly used techniques like discriminant analysis and artificial neutral network as described in Table 10.

There are two noteworthy and interesting points about Table 10. First of all, it shows the predictive ability of each model. Here, the column 2 and 5 (good accepted and bad rejected) are the applicants that are classified correctly. Moreover, the column 3 and 4 (good rejected and bad accepted) are the applicants that are classified incorrectly. Furthermore, it shows that neural network gives slightly better results than discriminant analysis. It should be noted that it is not possible to draw a general conclusion that neural network holds better predictive ability than discriminant analysis because this study covers only one dataset. On the other hand, statistical models can be used to further explore the nature of the relationship between the dependent and each independent variable.

Secondly, Table 10 gives an idea about the cost of misclassification which assumed that a bad accepted generates much higher costs than a good rejected because there is a chance to lose the whole amount of credit while accepting a bad and only losing the interest payments while rejecting a good. In this analysis, it is apparent that neural network (equals to 7) acquired less amount of bad accepted than discriminant analysis (equals to 30). So, neural network achieves less cost of misclassification.

Table 10: Predictive models comparison

Models	Dataset: Good accepted	Good rejected	Bad accepted	Bad rejected	Success rate (%)
Discriminant analysis	152	11	30	7	91.0
Neural network	115	1	7	21	94.4

REFERENCES

Colquitt, J., 2007. Credit Risk Management: How to Avoid Lending Disasters and Maximize Earnings. 3rd Edn., McGraw Hill Professional, New York, USA., ISBN-13: 9780071510530, Pages: 372.

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen., 7: 179-188.

Kim, J.C., D.H. Kim, J.J. Kim, J.S. Ye and H.S. Lee, 2000. Segmenting the Korean housing market using multiple discriminant analysis. Constr. Manage. Econ., 18: 45-54.

Lee, T.H. and S.H. Jung, 2000. Forecasting creditworthiness: Logistic vs. artificial neural net. J. Bus. Forecasting Methods Syst., 18: 28-30.

Mays, E., 1998. Credit Risk Modeling: Design and Application. Global Professional Publ., Chicago, IL., USA., ISBN-13: 9781888998382, Pages: 257.

Meyers, L.S., G.C. Gamst and A.J. Guarino, 2005. Applied Multivariate Research: Design and Interpretation. Sage Publications, USA., ISBN-13: 978-1412904124, Pages: 760.

Plewa Jr., F.J. and G.T. Friedlob, 1995. Understanding Cash Flow. John Wiley and Sons, New York, USA., ISBN-13: 9780471103868, Pages: 246.

West, D., 2000. Neural network credit scoring models. Comput. Oper. Res., 27: 1131-1152.