View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Lumbung Pustaka UNY (UNY Repository)

Risk Analysis of Credit Default on Rural Bank by Using Back Propagation Neural Networks Approach

Sukono[1*], Akik Hidayat[2], Wulan Marselly Firdaus[3]
[1*]*Email: fsukono@yahoo.com*
[1,2,3]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia*

Abstract

In this paper we discuss the classification of credit risk by a rural bank. It is assumed that the risk of non-performing loans by the debtor is influenced by several factors, among others: the age of the debtor, the number of family dependents, the amount of savings, the value of collateral, income per month, ceiling credit filed, the take home pay, and the loan repayment period. Risk analysis of credit default here performed using the method of back propagation neural networks approaches. Method of back propagation neural networks is formed by generalizing the training rules, namely a way adding a hidden layer, learning rate, the target error, and test data the same. Based on the analysis indicates that the back propagation neural networks has a success rate of 88.3%, to recognize patterns of problematic credit risk diagnosis. The results can be used as a basis for decision making in the provision of credit by the creditor to the debtor.

Keywords: Credit risk, risk factors, neural network, back propagations, decision making.

1. Introduction

The provision of credit is a business activity that has a high risk and the effect on health, as well as the sustainability of the banking business. In the case of a credit application by the customer, a decision maker in a bank must be able to take the right policy, to accept or reject the loan application [4]. In the decision-making up to now, banks usually ask the applicant credit (debtors) fill out a form in the form of a list of questions, and applicants are asked to complete the paperwork required by the bank, and then an assessment of the credit application. When credit analyst is able to make the right decision, then the debtor will not have problems with credit. So that bank will not suffer losses caused by non-performing loans [2]. Conversely, when wrong the decision there will be non-performing loans that could harm the bank. It thus has the potential debtor dikarena the risk of failure to repay the loan in the future [8]. Therefore, we need a technique that can be used to perform a risk analysis of potential failure repayment of debtors [10].

In recent years, a number of credit scoring techniques used for the analysis of the risk of failure of repayment. Among other things, Yoba et al. [9] conducted an analysis of credit scoring using neural and evolutionary techniques. One of the conclusions expressed by the value of the accuracy of the neural network, at least is that their performance depends on the values of certain parameters that are under the control of the investigor. Lay et al. [6], credit risk analysis using a reliability model based on neural networks esemble. Analyses revealed that the triple-phase neural network ensemble technique proposed may provide a promising solution for credit risk analysis. Zekic-Susac et al. [10], the analysis of credit scoring on small business, to make a comparison between logistic regression techniques, neural networks, and decision tree. Based on the analyzed data and statistical values the results show that the model of neural networks is better than the logistic regression model and decision tree models. The best model of neural networks outperfomed significant logistic regression models, and extracted personal and business characteristics of entrepreneurs, as well as the characteristics of the credit program as an important feature. Al-Doori and Beyrouti [1], the analysis of credit scoring is based on back propagation neural networks by using multiple activation and error functions. When properly and adequately trained, apply the fault activation and proper function, Neural Networks perform much better than other statistical approaches, such as logistic regression or discriminant analysis.

Referring to the analysis carried out [1], [6], [9] and [10], in this paper analyzed the data credir scoring on lending by a rural bank. Analysis of credit scoring on a rural banks conducted by using a model of back propagation neural networks. The purpose of the analysis here is to predict the credit risk rating, which in turn will be considered in making policy of giving credit to the debtor.

2. Metod of Analysis

In this analysis method discussed include: problem analysis, design of procedural neural network, the credit assessment process scheme, and processing programs.

2.1 Problem Analysis

Needed a system that could be used as a support assessment of credit risk classification that aims to: (a) Receive input of a data consumer credit; (b) Create a neural network with backpropagation method; (c) conditioning the data through classification and pattern recognition by using artificial neural networks; and (d) provide information in terms of output shows the classification of credit risk assessment.

Referring to Brown [3], neural network-based prediction requires: (i) Collecting consumer credit data as basis neural network implementation; (ii) The training process is done with consumer credit data which has been converted into a training data; (iii) Receive input in the form of consumer credit data which is then converted into data used for the test and test data; (iv) The testing process consumer credit data; and (v) Provide the output with the test results.

This system will produce 2 pieces of output: (1) Credit Accepted; or (2) Credit Rejected. The inputs that affect the output is 8, namely: age, number of family dependents, Total Savings, Value Assurance, Revenue per Month, Proposed Credit amount, Take Home Pay, and Credit Recovery Period.

2.2 Designing Procedural of Neural Network

 ISBN : 978-602-1037-00-3

Procedurally, the classification of credit risk assessment system using back propagation neural network is divided into two parts:

- Training of neural networks is used as the learning that will result in the value of the weights connecting neural networks.
- Testing of consumer credit data, used for the assessment of credit risk classification after training is completed.

Training of Neural Network

Neural network used in the system using this type of multilayer perceptron with back propagation algorithm. Input is required in this system is a consumer credit data, amounting to 8 inputs (X) consisting of: X_1 (Age); X_2 (Number of Dependent Family); X_3 (Amount Savings); X_4 (Value of Collateral); X_5 (Income per Month); X_6 (The amount of the Proposed Credit); X_7 (Take Home Pay); and X_8 (Refund Term of Loans). Output to be produced in this system is 2 classifications of credit risk assessment: O_1 (Received Credit) and O_2 (Rejected Credit).

In this case, for the training process in a multi-layer perceptron, backpropagation algorithm is used. The end result of training in the form of value weights liaison will be saved to a file. This file will be used by the classification of credit risk assessment system to perform the process.

In this training algorithm, there are 3 stages, namely:

- *Stages of Feedforward*
  Training data entered will be forwarded to each hidden unit is connected to it, the next each hidden unit weighted summing each input signal incoming to him. The results are used to calculate the output signal by using the activation function. Then sent to each output. At each unit of output summing each weighted input signal incoming to him. The results are used to menghituung output signal using the activation function [3], [5].
- *Stages of Backpropagation*
  Each unit receives a target output using the pattern, calculating the error factor of these units. Using the error factor, calculate the correction weights and bias correction. This error factor is then distributed back to all units in the previous layer (hidden layer). Each hidden unit delta summing input of the output units. The value obtained is then used to calculate the error factor. Then, by using the error factor, count the correction weights and bias correction, and so on [3].
- *Stages of Weight Adjustment*
  Change the weights and biases for each unit of input and hidden units. Adjustable weights for all layers, repeated cycles specified maximum length [3], [5].

Back Propagation Network Algorithms

According to Brown [3], a detailed network back propagation training algorithm can be described as follows:

Step 0: Initialize the weights, the training rate constant (α), error tolerance or weight value (when using the weights as a stopping condition) or set the maximum epoch (if using many epochs as a stop condition).

 ISBN : 978-602-1037-00-3

Step 1: During the stop condition is not reached, then do step 2 to step 9.
Step 2: For each pair of training patterns, do step 3 to step 8.

Stage I: Feed Forward

Step 3: Each input unit x_i (from unit 1 until n in the input layer) sends the input signal to each input that is at the hidden layer.

Step 4: Each unit in the hidden layer (from unit 1 until p, z_j $(j = 1,2,..., p)$) multiplied by the weights and summed with the bias is also added:

$$z\_net_j = v_{oj} + \sum_{i=1}^{n} x_i v_{ij} \tag{1}$$

$$z_j = f(z\_net_j) = \frac{1}{1+e^{-z\_net_j}} \tag{2}$$

Step 5: Each unit of output y_k $(k = 1,2,..., m)$ multiplied by weights and summed also added to the bias.

$$y\_net_k = w_{ok} + \sum_{j=1}^{p} z_j w_{jk} \tag{3}$$

$$y_k = f(y\_net_k) = \frac{1}{1+e^{-y\_net_k}} \tag{4}$$

Stage II: Backward Proagation

Step 6: Each unit of output y_k $(k = 1,2,..., m)$ receiving target pattern t_k accordance with the current training input pattern, and then the error information output layer (δ_k) calculated. Value δ_k sent to the bottom layer and is used to calculate the magnitude of the correction weights and biases (Δw_{jk} and Δw_{ok}) between the hidden layers to the output layer:

$$\delta_k = (t_k - y_k) f(y\_net_k) = (t_k - y_k) y_k (1 - y_k) \tag{5}$$

Calculate the weight change rate w_{jk} (which will be used later to change the weights w_{jk}) with the rate acceleration α

$$\Delta w_{kj} = \alpha \delta_k z_j ;\; k = 1,2,..., m;\; j = 0,1,..., p \tag{6}$$

Calculate the change in bias

$$\Delta w_{ok} = \alpha \delta_k \tag{7}$$

Step 7: At each units in the hidden layer (from the unit 1 until p; k=1, *2*, .., m) calculation of information error hidden layer (δ_j). Value of δ_j then used to calculate the weights and bias correction (Δv_{ji} and Δv_{jo}) between the input layer and the hidden layer.

$$\delta\_net_j = \sum_{k=1}^{m} \delta_k w_{kj} \tag{8}$$

$$\delta_j = \delta\_net_j f(y\_net_k) = \delta\_net_j z_j (1 - z_j) \tag{9}$$

Calculate the weight change rate v_{ji} (which will be used later to change the weights v_{ji}).

$$\Delta v_{ji} = \alpha \delta_j x_i \text{; } j=1,2,\ldots,p;\ i=0,1,\ldots,n \tag{10}$$

Calculate the change in bias (to correct v_{jo}).

$$\Delta v_{jo} = \alpha \delta_j \tag{11}$$

Stage III: Updates Weight and Bias

Step 8: Each unit of output y_k ($k = 1,2,...,m$) performed an update of the bias and the weights (j = 0, 1, 2,…, p) so as to produce new weights and biases:

$$w_{kj}(new) = w_{kj}(old) + \Delta w_{kj} \tag{12}$$

Similarly, for each hidden unit ranging from unit 1 until p conducted an update weights and bias:

$$v_{kj}(new) = v_{kj}(old) + \Delta v_{kj} \tag{13}$$

Step 9: Test of the condition stops (end of iteration)

Testing of Simulation Data

To know the ability of neural networks to predict the classification of credit risk assessment is then carried out the testing process. After the learning process, it is assumed that neural networks have been formed have enough knowledge, so as to produce the correct output if given a new entry [3], [7].

The testing process begins with a call back input unit weights and the weights of hidden units that have been made during the training. Next enter the test data as a signal input. In the testing process, simply run the feedforward phase of the training algorithm. Each unit sends a signal to enter all the hidden units connected to it. Each hidden unit then summing, each weighted unit entering to him. Results are used to calculate the output signal using the activation function, then sent to each unit of output. During each unit of output summing, the weighted inputs each incoming signal to him. Results are used to calculate the output signal the activation function [3], [7].

Furthermore, the output signal is compared with the target value has been determined previously to obtain the output of the testing process.

2.3 Credit Assessment Scheme Process

Scheme for the classification of credit risk assessment process is described as in Figure 1.

 ISBN : 978-602-1037-00-3

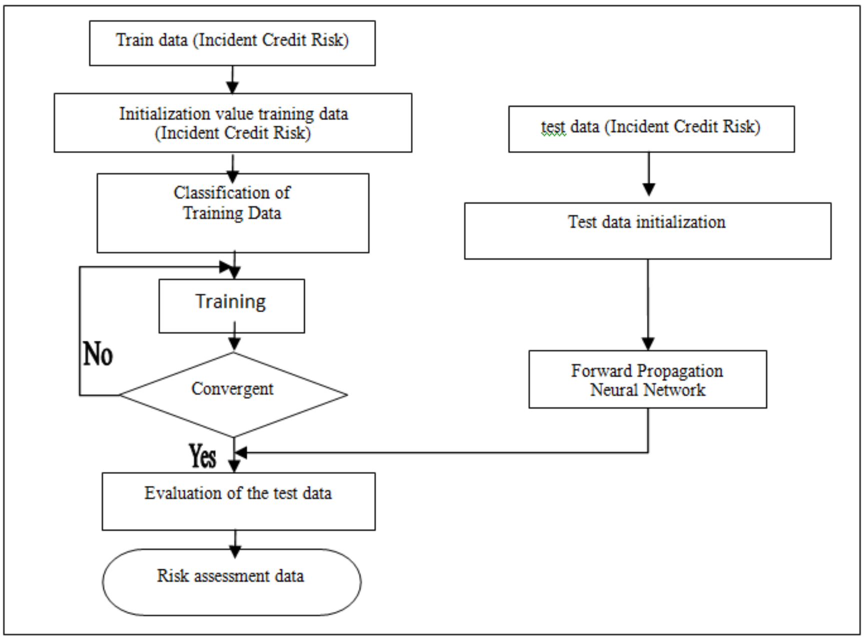

Figure-1. Scheme of Credit Risk Assessment Process Classification

2.4 Program Processing

Processing application program is divided into two stages, namely the training stage and testing stage. On the training phase, lasting two steps, namely the step forward (feedforward) and step back. Step feedforward, aims to calculate the input matrix and the calculation of the weight matrix. Meanwhile, the backwards aims to correct weighting matrix. In the training process, the maximum number of epochs or looping obtained from the input provided by the user program (user) [1], [9].

Unlike the training phase, the testing phase will only run the steps of feedforward backpropagation algorithm. Of the input is entered, the value will be output will be matched with the target database. The data that will be tested before the initialization is done first. After that, the input will be multiplied by the weighting matrix that has been obtained in the training phase. Multiplication between the input matrix and the weight matrix will yield an output value that has been determined in the training phase. The output obtained designate the targets that have been drilled previously [1], [9].

3. Results and Discussion

In this part of the discussion include: data analysis, program implementation, program evaluation, and implementation in making lending decisions. Starting with the data analyzed as follows.

3.1 Analyzed data

The data used in this study is secondary data, ie credit created simulated data similar to the circumstances of a particular bank. The data consists of 100 samples, which contained n_1=15 dengan kategori 1, meaning that the occurrence of credit risk, and n_0=85 with categories of 0, meaning there is always a credit risk. Most of the simulation data will be used to process credit training the rest will be used in the testing process. Simulation data of credit used as training data in the training process are not used in the testing process as the test data.

3.2 System Implementation

Credit risk assessment system consists of training and testing. The process of training and testing process are two user-selectable processes to be done in software. The training process is intended to train the neural network while the testing process is intended to test the neural network is trained and know the accuracy of the software.

The training process will stop when the value iteration has reached the maximum iteration, or an error value has been less than the maximum value of the error. The training process can also be stopped by the user by pressing the Stop button to view the training process. Once the training process is stopped, it will be shown that the weight values obtained in the program. Furthermore, the weight values are stored in a file that can then be used for the testing phase.

3.3 Evaluation

Testing of Credit Risk Assessment System

In this research, testing variations of the number of neurons and learning rate, to see its effect in achieving the maximum error value. Learning is done on the same input and serial, so it can be seen that the neural network performance comparison obtained. Neural network with back propagation method used to use one hidden layer and one output layer. On the output layer neuron number has been set used is 1 In contrast to the output layer, hidden layer makes it possible to change the value of the variation of the number of neurons in the hidden layer.

The variation of the number of neurons that will be tested are: 2, 5, and 10. Do also variations in learning rate, which is used in each variation of the number of neurons in the hidden layer before the learning rate, which is used are 0.1, 0.5 , and 0.9. The target value of the error made at 0 and the maximum value of 5000 made epoch, for each test variation is obtained as in Table-1.

Table-1. Results of Training with Variations in Network

No	Number of Hidden Layer Neurons	Learning Rate	Error Values
1	2	0.1	3.75861
2	2	0.5	3.93750
3	2	0.9	3.87186
4	5	0.1	0.81668
5	5	0.5	1.35932
6	5	0.9	2.88895
7	10	0.1	0.76243
8	10	0.5	1.31178
9	10	0.9	3.95959

From the Table-1 it can be seen that the pattern shows the absence of the greater number of neurons in the hidden layer, or the greater the value of learning rate affect the size of the error value. This is because making weight value by the system when testing is done randomly from [-1, 1], so for each results obtained training error rate fluctuation. Even if you try to repeat the training process with the number of neurons and learning rate values are the same, the value of different errors. So for the best training results epoch system needs maximum value greater error to achieve its target value. Or reduce the value of its error targets to achieve results that approach the training error value desired.

Accuracy Testing of Credit Risk Assessment System

To test the performance of the credit risk assessment system tested on existing test data. The testing process is done after weight training results obtained can then be seen in test results. The training process uses networks architecture with:

1. Learning rate : 0.1
2. Maximum of *epoch* : 100000
3. Target of error : 0.1
4. Neurons in the hidden layer : 2

The accuracy of the system can be calculated with the involvement of a number of test results in accordance with the value of its output targets and the amount of data tested.

$$\text{The level of accuracy} = \frac{\text{The appropreate of test result}}{\text{The number of data tested}} \times 100\%$$

From the 30 data used for testing, diperoleht 5 test results are not in accordance with the target output, and 25 test results in accordance with the target output. The level of accuracy obtained is someone of 88.3%.

3.4 Implementation of Decision Making

Case A. Debtor of A age (X_1) 40 years, the dependents of the family (X_2) as much as 4 people, apply for a loan (X_6) to Rural Bank Z of Rp 11.000.000, for family purposes (consumptive). Debtors of A guarantee (X_4) form BPKB motorcycle, which is estimated by the bank worth Rp 9.375.000. Based on the analysis conducted by the ability to pay the bank note that the Debtor of A has an income per month (X_5) Rp 3.500.000, and take home pay (X_7) as big as Rp 2.800.000, and have savings (X_3) in Rural Bank of Z at Rp 750.000. The repayment period (X_8) debtor of A is for 24 months (2 years). Based on the above information, whether debtor of A credit worthy?

Analysis. Based on the above data, then performed the testing process, and the results are as follows. Based on the backpropagation program there are 8 variables determinant of credit risk assessment results. Which each have a value which is then processed in the backpropagation program, and produces two criteria: if 0 then the loan rejected and 1 accepted. From the the analysis of the case, the debtor A gives 0, which means that the credit of the debtor A is rejected.

***Case* B.** Debtor of B age (X_1) 47 years, the dependents of the family (X_2) as much as 4 people, apply for a loan (X_6) to Rural Bank Z of Rp 75.000.000, family purposes (consumptive). Debtors of B guarantee (X_4) form BPKB motorcycle, which is estimated by the bank worth Rp 71.250.000. Based on the analysis conducted by the ability to pay the bank note that the Debtor of B has an income per month (X_5) Rp 5.000.000, and take home pay (X_7) as big as Rp 4.400.000, and have savings (X_3) in Rural Bank of Z at Rp 1.600.000. The repayment period (X_8) debtor of A is for 72 months (6 years). Based on the above information, whether debtor of A credit worthy?

Analysis. Using the above data, be tested and obtained the following results. Based on the backpropagation program there are 8 variables determinant of credit risk assessment results. Which each have a value which is then processed in the backpropagation program, and produces two criteria: if 0 then the loan rejected and 1 accepted. From the the analysis of the case, the debtor B produces 1, which means that the credit of the debtor B is accepted.

4. Conclusion

In this paper we analyzed the potential risk of failure classification repayment, using the approach of back propagation neural networks. Back propagation neural network is able to predict credit risk. Although using the learning rate, the target error, and training data same, yet surely result in an error and the test results same. This is because the value of the weights generated for each training process is different. The cause of the weights produced at each learning process is different, because the provision of initial weight value is randomly generated and different for each training process. Backpropagation neural networks are used in the analysis in a rural bank has a success rate of 88.3%, to recognize patterns of credit risk diagnoses (test data) as the assessment of credit risk provision. Based on the analysis of credit scoring can be used as consideration making lending decisions by a rural bank.

References

[1] Al Doori, M. and Beyrouti, B. (2014). Credit Scoring Model Based on Back Propagation Neural Network Using Various Activation and Error Function. *IJCSNS International Journal of Computer Science and Network Security*, VOL.14 No.3, March 2014.

[2] Baesens, B., Setiono, R., Christophe Mues, C., and Vanthienen, J. (2003). Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation. *Management Science © 2003 INFORMS,* Vol. 49, No. 3, March 2003 pp. 312–329.

[3] Brown, A.J. (2004). Analysing a Neural Network Trained by Backpropagation. *Report.* School of Computer Science, The University of Birmingham.

[4] Gouvêa, M.A. and Gonçalves, E.B. (2007). Credit Risk Analysis Applying Logistic Regression, Neural Networks and Genetic Algorithms Models. *Working Paper.* POMS 18th Annual Conference, Dallas, Texas, U.S.A. May 4 to May 7, 2007.

[5] Istook, E. and Martinez, T. (2003). Improved Backpropagation Learning in Neural Networks with Windowed Momentum. *International Journal of Neural Systems*, vol. 12, no.3&4, pp. 303-318.

[6] Lai, K.K., Yu, L., Wang, S., and Zhou, L. (2006). Credit Risk Analysis Using a Reliability-Based Neural Network Ensemble Model. *Working Paper*. Department of Management Sciences, City University of Hong Kong, College of Business Administration, Hunan University, Changsha 410082, China.

[7] Patidar, R. and Sharma, L. (2011). Credit Card Fraud Detection Using Neural Network. *International Journal of Soft Computing and Engineering (IJSCE)*, ISSN: 2231-2307, Volume-1, Issue-NCAI2011, June 2011.

[8] Sadatrasoul, S.M, M.R. Gholamian, M.R., Siami, M., and Hajimohammadi, Z. (2013). Credit scoring in banks and financial institutions via data mining techniques: A literature review. *Journal of AI and Data Mining*,Vol. 1, No. 2, 2013, 119-129.

[9] Yoba, M.B., Crook, J.N., and Roos, P. (2000). Credit Scoring Using Neural and Evolutionary Techniques. *IMA Journal of Mathematics Applied in Business and Industry* (2000) 11, 111-125.

[10] Zekic-Susac, M., Sarlija, N., and Bensic, M. (2005). Small Business Credit Scoring: A Comparison of Logistic Regression, Neural Network, and Decision Tree Models. *Working Paper*. University of J.J. Strossmayer in Osijek, Faculty of Economics in Osijek Gajev trg 7, Osijek, Croatia.