Classification of Bank Customers Using the Neural Network Algorithm for Credit Risk

Reza Shah'savari[1,] Hassan Rashidi Heram Abadi[2*]

[1]Faculty of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran
[2] Department of Mathematics and Computer Science, Allameh Tabataba'i University, Tehran, Iran

Email: rezashahsavari518@gmail.com

Abstract:

One of the most important issues always facing banks and financial institutes is the issue of credit risk or the possibility of failure in the fulfillment of obligations by applicants who are receiving credit facilities. The considerable number of banks' delayed loan payments all around the world shows the importance of this issue and the necessary consideration of this topic. Accordingly, many efforts have been made for providing an efficient model for more accurate evaluation and classification of applicants receiving credit facilities for valid decision making about granting or not granting these facilities to them. Granting of banking facilities are great importance economically. Because increasing the quantity of capital cause the growth and development of economic. But in granting facilities, banks are faced to great danger which called bank credit. This is due to the exposures of banks to the financial crisis. The credit risk can take into account the possibility of non-repayment of loans by the applicant. Using data mining is significant to detect customers that will increase your credit risk. In this paper we intend to reduce bank customers using random forest algorithm in order to reduction credit risk.

Keywords: Credit risk, Data mining, Forecasting, Neural networks, Decision tree, Support vector machine, Random forest.

1. Introduction

Over the last few years organizations and especially financial institutions in our country are concerned about the issue of risk and damages caused by it but despite its importance, a coordinated framework for implementing risk management and also accurate indicators for determining credit risk are not available. In addition, rating industry has not found its own good place in our country where the main reasons for it include cultural, economic and educational issues, the lack of a centralized databank, the lack of a strong and effective information exchange network, the lack of

adequate laws and regulations, and political issues. Therefore, a strategy shall be advised in order to provide the financial resources required for applicants and banks perform their main duty which is lending with the least possible risk; because in current changing conditions, the success of any firm depends on the risks and their risk management methods. The most important risk facing banks is credit risk which includes loans that have been paid in the past. Overall, credit risk for a bank is the possibility of losing time or generally obligations being neglected by debtors because of their inability to fulfill their obligations to the bank. These obligations usually involve repayment of the debts and their interest to the bank on the specified date. Credit loans are the basis for banking industry. The performance of credit section in a good situation guarantees the profitability and stability of a bank. Therefore, securing the financial background history of customers is a very important factor before making any decisions regarding credits and also a key determinant in reducing credit risk. Credit risk is one of the most critical and biggest challenges facing banks. In fact, the estimation of a risk is an important factor for any decisions regarding credits and inability to determine the accurate risk has a reversed effect on credit management. In addition, risks can affect approved and non-approved investment decisions. When the credit manager approves a loan, he runs the potential risk of client being unable to repay it. On the contrary, when a loan is turned down, the potential risk of losing customers to competitors arises. Hence, assessing credit risk before making a decision to lend is important. Consequently, banks are considering lending customers with returns adequate to the profit while having a low risk of return. It happens when banks are able to identify credit customers whether natural or legal and classify them according to their ability and willingness to repay their obligations fully and on time using appropriate financial and non-financial criteria.; because in such a system, facilities are granted to those applicants that have a low risk credit and the possibility of repaying their debts on due date is much higher. Since these funds can be used as a source of financing for future lending, they have an important role in increasing investment, growth and economic development in the country. [4] Credit risk is important because by its measurement and estimation, managers in financial institutions can properly evaluate the loans or bonds and accordingly impose restrictions on credit allocation to different customers; therefore, they can support financial institutions against the risk of non-repayment of financial obligations. [5] Granting of banking facilities are great importance economically. Because increasing the

quantity of capital cause the growth and development of economic. But in granting facilities, banks are faced to great danger which called bank credit. This is due to the exposures of banks to the financial crisis. The credit risk can take into account the possibility of non-repayment of loans by the applicant that must be managed. Different techniques can be used for credit risk management; one of the methods is to design of credit ratings determination system for the recipients of facilities. The purpose of this study is that in credit markets by designing and implementing customer validation try to identify of customer behavior patterns and therefore possibility of behavior prediction. The customer validation is a very complex field in activities. The number of factors and complexity of financial, economic and behavioral relations make difficult the validation. The other hand evaluation process often should be done in a short period because prolonging of evaluation process can cause delay in operation and finally the costs will increase. On the other hand, possible imprecision in the evaluation can lead to incorrect decisions and as a result, the high drawbacks. Time constraints and the necessity of accuracy in evaluation will double the complexity of the issue. Data mining exploits of developments that occur in the field of artificial intelligence and statistics. Both of these fields work in pattern recognition issues and data

classification and consequently will be directly used in data mining and both groups are active in recognition and use of neural networks and decision trees. Data mining is not a substitute for former statistic techniques but is their heritor and in fact is the change and development of former techniques for fitting them with the volume of data and today s issues. Classical techniques were suitable for limited data and simple issues while with getting complex the issues and increasing growth of data, the need for their change is perfectly normal. On the other hand data mining is the combination of classical techniques with new algorithms such as neural networks and decision tree. The most important point is that the data mining is a strategy for today s business issues with the help of statistical techniques and artificial intelligence for professionals who intend to create a prediction model. In this paper, we want to classify the bank customer by using the neural network algorithm. The random algorithm is one of the algorithms based on the assembly. Assembly-based algorithms are considered in recognition problems of patterns due to their high generalization capabilities. In this paper, we first review the previous methods.

2. Review of Literature

Designing a model for measuring and rating the credit risk was conducted for the first time on the

bonds in 1909 by John Murray. Similarities between bank credit facilities and bonds led researchers to consider the rating of bank credit risk, i.e. measuring the risk of not repaying the loan and its interest. The first model used to determine the bankruptcy of firms was the multivariate logistic regression model which was presented by Beaver in 1966. Later, this model was used for measuring credit risk of bonds issued by the companies. One of the first studies in the field of measuring credit risk of corporate bonds using multivariate scoring model was done in 1968 by Altman which is known as Z-score model. Altman Z-score model is a discriminant analysis model that uses significant amounts of financial ratios and tries to differentiate between companies that are in financial distress (i.e. bankruptcy) and those that are not in financial distress. Due to the fact that non-repayment of loans mainly belongs to companies that will suffer from financial distress in the future, so the ability of predicting credit risk by using this model will be possible. Thus, in 2001, Sanders and Alan used this model for predicting credit risk of companies that received loans from the banks; their investigations showed that this model was powerful for predicting credit risk. [7]

In recent years, several studies at national and international levels have been done on the issue of credit risk, scoring and rating of customers of banks. These studies used different statistical and economic methods and techniques and tried to quantify the credit risk of banks and financial institutions' customers. In the next sections, some of these studies and researches will be discussed. In their research, Fallah-Shams and Tehrani evaluated the effectiveness of possible linear models, logistic and artificial neural networks to predict the credit risk of customers of banking system in our country. The results suggest that the relationship between the variables in the model of predicting credit risk is not linear and exponential and sigmoid functions are the best models in order to predict credit risk. Artificial neural networks and logistic models respectively have the maximum performance to predict credit risk. [8] In their research, Nilsaz et al used a neural network model for classifying loan applicants of installment sale and compared this model with two statistical models i.e. discriminant analysis and logistic regression. The results of this comparison show that the neural network compared to other models have higher efficiency and accuracy. [9] Ebrahimi and Daryabar identified factors affecting the credit risk and presented a model to predict credit risk and credit rating of legal clients seeking a commercial bank facility; they have used data envelopment analysis , logistic regression, and neural network to compare these three models. The results of the models used in the estimation

of credit risk and credit rating comparing with the actual results show that neural network model is more effective in order to predict credit risk of corporate customers and the credit rating. [4] Boguslauskas and Mileris in their study named, "Evaluation of credit risk by artificial neural network models, reported that artificial neural network and logistic regression are the most effective models and are widely used in the management of credit risk. They have analyzed the credit risk of companies by using credit risk assessment models. [10] In a study titled, "credit risk evaluation using neural networks: studying various neurological models and learning program", Khashman used neural networks to assess credit risk using the German database. Three neural networks were developed with 9 learning programs and then the results of different functions were compared. The results showed that one of the learning programs achieved a high performance with a ratio of 83/6. [11] Jagric et al. (2011) stressed that a major challenge for banks is designing new credit risk models with high accuracy prediction. They emphasized on using artificial neural networks to develop a model of credit marking because of their ability to attract non-linear financial data. They developed a credit decision model using a learning vector quantization (LVQ) neural network and logistic regression model. The results showed that LVQ model performed better

and more accurate than logistic model. [12] Blanco et al. (2013) compared the performance of multilayered neural network (MLP) with three statistical techniques: logistic regression, linear discriminant analysis, and quadratic discriminant analysis. The findings of their study confirmed that MLP is superior to other parametric statistical techniques. [13] In a study titled, "credit risk assessment model for Jordan's commercial banks", reviewed the credit decision with the help of logistic regression and neural networks. The results indicated the better performance of artificial neural network model in identifying customers who may not pay their debts [3].

3. Classification Algorithm

We use from three algorithm of decision tree, support vector machine and multi-layer perceptron network for comparison of the performance of the proposed method in order to the classification f bank customer. We use from k-cross validation for a more accurate comparison. Here k is considered to 5, so that the dataset is divided to 5 equal parts. Then the 4-part is used for training and one part for testing. This process is repeated 5 time-sand in each stage is used from a different part for test. Therefore the average accuracy of these 5 runs is considered as final accuracy. The result were shown in figure 1.
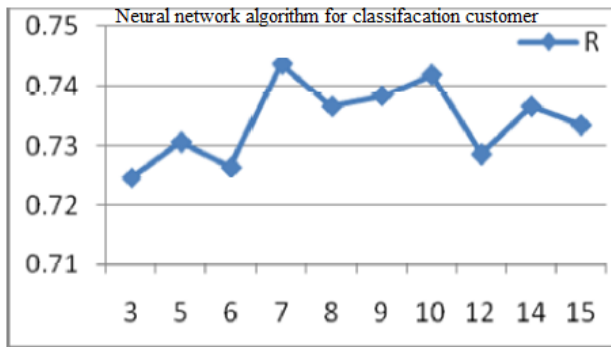
Figure 1. Implementation of the random neural network for classification customer [8]

4. Neural Network Algorithm

Artificial Neural Networks (ANNs) are non-linear techniques that imitate the human brains functionality. They are used broadly in classification, clustering and optimization problems [10]. ANNs are able to recognize the complex and non-linear patterns between input and output variables in credit scoring which then predict the creditworthiness of a new applicant. They can also use for clustering applicants. Artificial neural networks are considered a powerful alternative to conventional forecasting and classification methods due to their ability to capture nonlinear and complex relationships. These models have a biologically inspired capability that mimics processing capabilities of the human brain (Cao and Parry, 2009). They have been used successfully in financial applications, a good ability in classification (e.g., credit scoring, corporate failure prediction and bond ratings) as well as in modeling tasks such

as predicting share price movements and exchange rate fluctuations. The multi-layer perception (MLP) is the most popular feed-forward neural network (FFNN) model used in pattern recognition. Designing an artificial neural network model successfully relies on a clear understanding of the problem, and on deciding upon most influential input variables. A typical FFNN model is represented as some processing units called neurons cooperating across several linking layers [6]. The information flows from origin to destination strictly in one direction through a system of weighted connections, without interconnections between the output of a neuron and the input of another neuron in the same layer or in a preceding layer. The output of each neuron is the outcome after applying the transfer function to the weighted sum of all inputs to that neuron [7]. A typical FFNN model is usually comprised of a three-layered architecture: input, hidden, and output layers.

The input layer feeds the input variables (predictors) to the next layer. Each hidden neuron receives a weighted sum of all inputs in the input layer, applies a transfer function such as log sigmoid, hyperbolic tangent, soft-max to the weighted sum. Similarly, each hidden neuron transfers a weighted outcome to each neuron in the output layer i.e. each dependent variable neuron (Ong et al., 2005; Cao and Parry, 2009). The outcome of the output neuron is the solution

of the problem. The neural networks learn the desired relationship between the independent and dependent variables by training the net using a representative set of (input, target) pairs. A learning algorithm is used to find the values of the connection weights where the network preserves its knowledge. During training when an input pair is fed to the network, the net calculates a temporary output, Y. Next, the net compares the actual output, Y, with the desired output, T, and if not satisfied then it adjusts the connection weights in proportion to error which is equal to the difference between its output and the target in an iterative process until a desirable result is reached. Radial basis function neural networks are the most widely used basic functions.
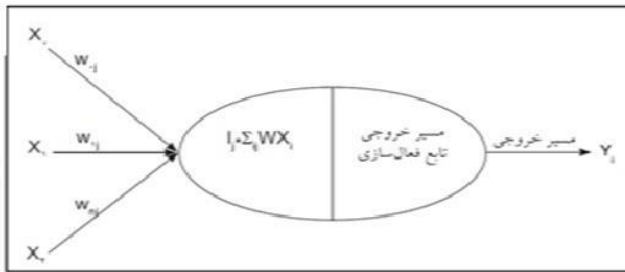

Figure 2. Schematic of Neural Network Algorithm [9]

The main structure, the ability of learning and the different applications are the features of this function. These functions are superior to multi-layered networks because the speed of their learning is considerably higher than that in other networks. Radial basis function neural networks

are usually used to approximation. If these functions are used to classify data, a sigmoid or limiting function should be placed on output nerves to output values appear as 0 or 1. [15] In the field of mathematical modeling, RBF is an artificial neural network which uses radial basis functions as excitation functions. The output of this network is a linear combination of radial basis functions for input parameters and neurons. These networks are used in approximation function, prediction of time series, classification and controlling of systems.

5. Neural network architecture

Neural networks usually consist of three layers: an input layer, a hidden layer with a nonlinear function of RBF stimulation and an output layer. This architecture is schematically shown in the figure below.
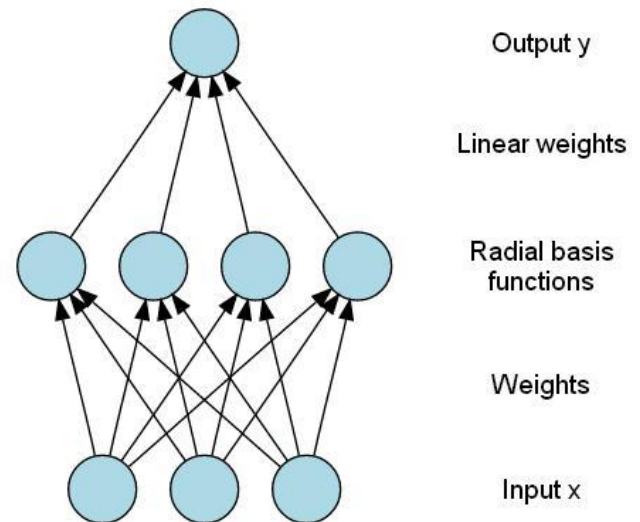

Figure 3. Schematic form of a neural network [10]

Input can be modeled as a vector of real numbers, and the output of this network is a

scalar function of input vector which is calculated as follows:

$$\varphi(x) = \sum_{i=1}^{N} a_i\, p\left(\left\| x - c_i \right\|\right)$$

In which N is the number of neurons in the hidden layer, c_i is the vector of central neuron i and a_i is the weight of neuron i in the linear output neuron [16].

6. Experimental Results

In this paper we studied the bank customer classification using data mining techniques and neural network algorithm in order to allocation of credit facilities for customers. The dataset have been collected that is related to the number of customer from one of bank branch in Iran. Data mining and classification techniques that have been used in references and sources for the customer classification include the support vector machine, decision tree and multi-layer neural network. We also used the neural network that is an algorithm based on the assembly. Results on criteria such as precision, accuracy and true classification rate of negative samples (True Positive Rate) indicate that neural network algorithm is much better than other previous methods.

7. Conclusions

Studies on the process of accreditation of banking system in our country indicate that the credit system is judgmental. Most banks in the country have neglected the issue of risk in general and credit risk in particular species. Implications of a judgmental system for credit granting in the country's banking system include: The high volume of reserves and doubtful debts and the subsequent deterioration of the bank's efficiency. Today, most of the banks lack a risk management in their organizational structure and if there is a unit for risk management, they have not done enough to control and manage risks.

The results indicate the fact that there is a possibility of predicting customers' credit risk when granting credit facilities to customers as predictor variables and using them in statistical models and neural networks. The results of this study show both logistic regression and neural network are effective; but in comparing the two models, neural network has higher efficiency than logistic regression. Overall, according to the results obtained from this study, the following suggestions for improving bank's credit system are recommended:

- Establishing databases and information systems that contain financial and management data belonging to bank customers. This data shall be converted to credit risk predictor variables. This database shall be updated based on new

data to estimate the amount of customers' credit risk at any time possible.

- Designing and establishing a software system for logistics and neural network models to predict the credit risk of customers.

- Combining quantitative and qualitative methods to predict the credit risk of customers in an expert system and establishing a credit rating of customers based on these models.

- Reviewing and revising the credit risk prediction model based on the continuous feedback of results.

- Determining the credit capacity of each customer based on the determined amount of credit risk based on this model.

- Coordinating between the units involved in the credit process, such as management of project evaluation, credit management, credit risk management and claims administration.

References

1. Hesabi, S., (1393). Specifying the optimum portfolio model of facilities for reducing credit risk, master thesis. Vol. 3, pp. 45-69.

2. Yurdakul, F., (2014). Macroeconomic Modeling Of Credit Risk For Banks. Procedia Social and Behavioral Sciences, vol. 109, pp. 784-793.

3. Bekhet, H. A., Fathi Kamel Eletter, S., (2014). Credit risk assessment model for Jordanian commercial banks : Neural scoring approach. Review of Development Finance, vol. 4, pp. 20- 28.

4. Ebrahimi, M., Daryabar, A., (1391). Credit risk management in banking system, data envelopment analysis approach, logistic regression and neural network. Quarterly Journal of knowledge investment. Vol. 6, First year. second number. Pages 35-62

5. Rostami, A., Ranjbar Fallah, M. R., Khademolhosseini Ardakani, M., Habibi, M. H., Jamali Neishaboor, A., Ali Mohammadi, M., (1393). Credit bonds: Legal, economic and practical feasibility level for microeconomic participation in improving the bank's credit risk. Journal of knowledge and research investment. Third year. vol. 9,. Pages 187 -205

6. Pooyanfar, A., Fallahpoor, S., Azizi, M. R., (1392). Least-squares support vector machine approach based on genetic algorithm to estimate the credit rating of the bank's clients. Journal of financial engineering and management of securities. Vol. 17. Pp. 133-158.

7. Mirzaei, H., Nazarian, R., Baegheri, R., (1390). Surveying factors affecting credit risk of banks legal entities. Quarterly Journal of Economic Research. Nineteenth year. Vol. 59. Pp. 67-98.

8. Tehrani, R., Fallah Shams, M., (1384). Designing and specifying the credit risk model in banking system. Journal of social and human sciences, Shiraz University. Twenty-second period. Vol. 6, Pp. 45-60.

9. Nilsaz, H., Rasekh, A., Assareh, A. R., Sinayee, H. A., (1386). Application of neural networks in credit rating of installment sales for loan applicants. Journal of Economic Research. vol. 32. Pp. 85-110.

10. Boguslauskas, V., Mileris, R., (2009). Estimation of credit risk by Artificial Neural Networks models. Eng. Econ. Vol. 4, pp. 7-14.

11. Khashman, A., (2010). Neural network for credit risk evaluation : investigation of different neural models and learning schemes. Exp. Syst. Appl. Vol. 37 (9), pp. 6233-6239.

12. Jagric, V. Kracun, D., Jagric, T., (2011). Does non-linearity matter in retail credit risk modeling ? Finance a uver – Czech J. Econ. Fin. Vol. 61 (4), pp. 384-402.

13. Blanco, A., Mejias, R., Lara, J., Rayo, S., (2013). Credit scoring models for the microfinance industry using neural networks: evidence from Peru. Exp. Syst. Appl. Vol. 40(1), pp. 356-364.

14. Fallah Shams, M., Mahdavi Raad, H., (1391). Designing a model for validating and predicting the credit risk of customers for leasing facilities. Journal of Economic Research. (Islamic-Iranian approach) twelfth year. Vol. 44. Pp. 213-234.

15. Sajadi, S., (2012). Design of radial basis function neural networks based on Fuzzy K nearest neighbors. Acta Compuare, vol. 1, pp. 100-105.

16. Liu, J., (2013). Radial Basis Function (RBF) Neural Network Control for Mechanical Systems: Design. Analysis and Matlab Simulation. Springer Science & Business Media, vol. 56, pp. 46-52.