# Measuring Credit Risk of Bank Customers Using Artificial Neural Network

Mohsen Nazari

Department of Business Management, Faculty of Management, University of Tehran

E-mail: mohsen.nazari@ut.ac.ir

Mojtaba Alidadi (Corresponding author)

Department of Business Management, Faculty of Management, University of Tehran

E-mail: mojtaba.alidadi@ut.ac.ir

**Abstract**

In many studies, the relationship between development of financial markets and economic growth has been proved. Credit risk is one of problems which banks are faced with it while doing their tasks. Credit risk means the probability of non-repayment of bank financial facilities granted to investors. If the credit risk decreases, banks will be more successful in performing their duties and have greater effect on economic growth of the country. Credit rating of customers and identifying good and bad customers, help banks lend to their good payers and hereby, they reduce probability of non-repayment. This paper aims to identify classification criteria for good customers and bad customers in Iranian banks. This study can classified in applied studies group and the research strategy is descriptive. Artificial neural network technique is used for financial facilities applicants' credit risk measurement and the calculations have been done by using SPSS and MATLAB software. Number of samples was 497 and 18 variables were used to identify good customers from bad customers. The results showed that, individual loan frequency and amount of loan had most important effect and also status of customer's bank account, history of customer relationship with bank and received services had least important effect in identifying classification criteria of good and bad customers. The major contribution of this paper is specifying the most important determinants for rating of customers in Iran's banking sector.

**Keywords:** Credit risk, Credit rating, Artificial neural network, Bank, Iran

*JEL Classification:* G24; G32; E51; C45; H81

## 1. Introduction

The main role of commercial banks is to collect deposits of real and legal persons and also allocate them to borrowers and investors. Banks as one of the financial institutions are responsible for relationship between lenders and borrowers. Generally, evaluations of loan applications are based on a loan officers' subjective assessment. Such judgment is inefficient, inconsistent, and non-uniform. Therefore, a knowledge discovery tool is needed to assist in decision making regarding the application. Financial system efficiency and level of financial improvement play an important role in economic development of countries. The better the banks act, the more economic development occurs.

Banks face problems such as the probability of non-repayment of received loans at the due date or non-repayment that are called "credit risk". Credit risk is widely studied topic in bank lending decisions and profitability (Angelini, di Tollo, & Roli, 2008). Borrowers usually have better information about the projects to be financed, but lenders usually don't have sufficient information about those projects (Matoussi & Abdelmoula, 2009). If a bank faces good customers, it will definitely have more power in lending loans and thus, increasing in profit will occur. But if bank faces bad customers who don't repay loans in due date, it will likely go bankrupt.

Risk analysis in today's financial markets is one of important factor that could be applied with neural networks (Vasconcelos, Adeodato, & Monteiro, 1999). Artificial Neural Networks (ANN), play an increasingly important role in financial applications for such tasks as pattern recognition, classification, and time series forecasting. Factors affecting the presence and upsurge of credit risk are divided into two groups: within the organization and outside the organization.

  i.  Factors outside the organization cannot be controlled by bank management teams and are considered exogenous for bank. Political changes, earthquakes, war and etc. are included in this category.

  ii. Factors within the organization are those which bank management teams could take the matter under their control. These are called endogenous factors.

In this research we are trying to figure out the endogenous factors affecting credit risk within an organization. The main purpose of this paper is to identify and rank the factors that have effects on credit risk in one of the commercial banks in Iran. This research can estimate the credit risk of each customer and helps to make the right decision toward granting of loan to customers. ANN model is used in this research to achieve the above-mentioned. The ANN models use the same input and output parameters as in the linear models. These models have three primary components: the input data layer, the hidden layer(s) and the output measure(s) layer. Each of these layers contains nodes, and these nodes are connected to nodes at adjacent layer(s). The hidden layer(s) contain two processes: the weighted summation functions and the transformation function. Both of these functions relate the values from the input data to the output measures. The weighted summation function is typically used in a feedforward/back propagation neural network model. Figure 1 demonstrates a simplified
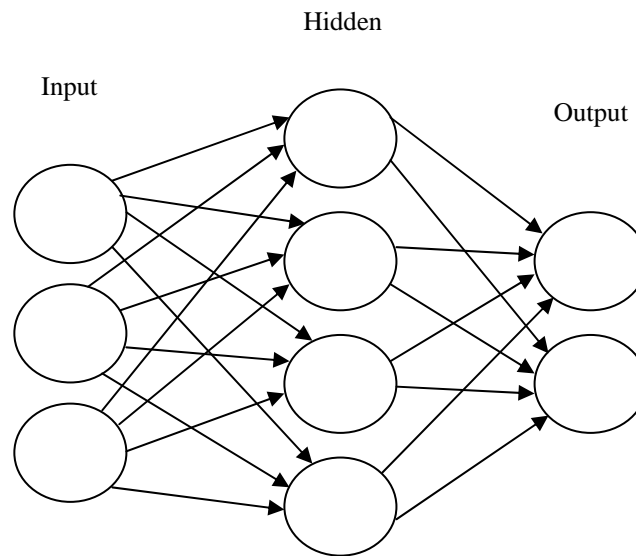
neural network.

Hidden

Input

Output

Figure 1. A simplified neural network model

## 2. Literature Review

Credit rating is one of technical factor in credit risk evaluation (Khashman, 2010). The aim of credit rating is to categorize the applicants into two groups; applicants with good credit and applicants with bad credit (Ghodselahi & Amirmadhi, 2011). Multilayer feedforward networks are a class of universal approximation (Hornik, Stinchcombe, & White, 1989). The concept of "universal approximation" is one of the fundamental and important features in ANN models. These models have a high predictive power (Steiner, Neto, Soma, Shimizu, & Nievola, 2006). This means that the networks are capable of adapting to arbitrary and unknown functional forms, with an arbitrarily specified degree of precision. These non-parametric models are well-known in many areas, especially in computing because of their highly-sophisticated pattern recognition (Hall, Muljawan, Suprayogi, & Moorena, 2009). Universal approximation leads us towards regarding neural networks generally as flexible non-linear statistical methods (Curry, Morgan, & Silver, 2002).

Extracting rules from a neural network are different. Steiner *et al.* (2006) analyse a credit risk data set by using the NeuroRule extraction technique. Angelini *et al.* (2008) indicate applicability of neural networks in credit risk applications, especially as black-box non-linear systems to be used in cohesion with classical rating and assortment systems. In order to forecast credit risk in banks, sometimes a combination of methods is used. Pacelli and Azzollini (2011) conclude that neural network models in combination with linear methods have further supported. Use of modern indicators in addition to traditional financial ratio indicators provides considerable improvement in forecast accuracy (Atiya, 2001). Salehi and Mansoury (2011) investigate the efficiency of neural network and logistic regression in forecasting customer credit risk. They state that both models have same efficiency. Credit

rating is also investigated with other methods of artificial intelligence. Ghodselahi and Amirmadhi (2011) Use a hybrid method for credit rating. They use Support Vector Machine, Neural Networks and Decision Tree as base classifiers. They found that accuracy of this hybrid model is more than other credit rating methods. Credit rating is also investigated with three methods including Logistic Regression, Neural Networks and Genetic Algorithms (Gouvêa & Gonçalves, 2007). According of this research's result, logistic regression and neural networks are good and similar. Although neural network is slightly better and genetic algorithms take third place. The important roles of ANN in financial application are pattern recognition, classification and time series forecasting (Eletter & Yaseen, 2010). Some researchers focus on learning approach in neural network models. In this field, Yu *et al.* (2008) use a multistage neural network ensemble learning approach. They state that this technique can provide a promising solution to credit risk analysis.

## 3. Methodology

### 3.1 Data and Variables

Data has been collected through stratified random sampling method over the period 2006-2011, based on documents and records of applicants for an Iranian commercial bank. Sample estimation has been done by a pretest sample size of 90 cases and according to the sample size formula, 497 are selected as the number of samples, which are derived from individual customers' profiles. The variables are as follows:

i. <u>Dependent variable</u>: good and bad customers; in this study we are aim to estimate the likelihood of good or bad customer being and also realize that how important are those factors. In this regard, good customer is a person who repays its loan plus the profit at the due date and in contrast, bad customer is a person who don't repay at the due date. To differentiate between good and bad customers in our neural network model calculations, we assign 0 to indicate good customers and 1 to indicate bad customers.

ii. <u>Independent variables</u>: In this study based on the previous researches and interview with bank experts, 18 variables are defined as independent variables:

1. Gender: data samples are divided to female and male according to their gender.

2. Age: data samples are divided to four groups; 18-25, 26-40, 41-60 and more than 60 years.

3. Education: data samples are divided to four groups; uneducated, primary education, high school and university degree holder.

4. Job: data samples are divided to following groups; self-employed, farmers and farm related jobs, Doctors, teachers, military personnel and office staff.

5. Work experience: data samples are divided into four groups; less than 5 years of job experience, 6-10 years, 11-20 years and 21-30 years.

6. Type of loan: data samples are divided into four groups according to their Islamic banking contracts.

7. Amount of loan: amount of money that is given to the customer.

8. Individual loan frequency: number of times that customer has received loans from bank.

9. Individual account turnover average: average of monthly account balance and turnover.

10. Time period of loan: that is divided to six groups; less than 7 months, 7-12 months, 13-24 months, 25-36 months, 37-60 months, more than 60 months.

11. Type of collateral: data samples are divided in two categories: physical assets like home and property; and financial assets like equity and long term deposit.

12. Interest rates: it expressed as percentage and it determine amount of bank's profit.

13. Penalty rates: The amount of money that customer has to pay for any delay in repayment of the loan (this is apart from the bank profit).

14. History of customer relationship with the bank: The time period that customer is in relation with the bank.

15. Received services: The service quantity that a customer receives from bank (bank's services that provided to customers).

16. Status of customer's bank account: it indicates whether a customer had any returned check or not.

17. Bank's branch ranking: it determines the rank of a certain bank's branch.

18. Value of collateral: it should be determined accordance with amount of loan.

*3.2 Model Estimation*

We apply data into neural network model to estimate the probability that the customers are good or bad. Table 1 shows network information.

Table 1. Network information

| Input Layer | Factors | 1 | gender |
|---|---|---|---|
| | | 2 | Age |
| | | 3 | Education |
| | | 4 | Job |
| | | 5 | Work experience |
| | | 6 | Type of loan |
| | | 7 | Amount of loan |
| | | 8 | Individual loan frequency |
| | | 9 | Individual account turnover average |
| | | 10 | Time period of loan |
| | | 11 | Type of collateral |
| | | 12 | Interest rates |
| | | 13 | Penalty rates |
| | | 14 | History of customer relationship with bank |
| | | 15 | Received services |
| | | 16 | Status of customer's bank account |
| | | 17 | Bank's branch ranking |
| | | 18 | Value of collateral |
| | Number of Units | | 600 |
| Hidden Layer(s) | Number of Hidden Layers | | 1 |
| | Number of Units (Excluding the bias unit) in Hidden Layer 1 | | 12 |
| | Activation Function | | Hyperbolic tangent |
| Output Layer | Dependent Variables | 1 | Bad Customer |
| | Number of Units | | 2 |
| | Activation Function | | Softmax |
| | Error Function | | Cross-entropy |

## 4. Data Analysis and Results

The data description is shown in Table 1. The dependent variable (Y) is an ordinal variable. Some of the independent variables are also ordinal and some of them are scale. SPSS (version 19) software is used for modeling. 70 percent of data is considered for training of neural network model, 20 percent of data is considered for testing model during training and also 10 percent of data is considered for testing model after completion. For building the provided model, only one hidden layer with hyper tangent activation function is used. The numbers of nodes in the inner layer will be selected automatically. The output of software analysis is shown in Table 2.

Table 2. Classification

| Sample | Observed | Predicted | | |
|---|---|---|---|---|
| | | .00 | 1.00 | Percent Correct |
| Training | .00 | 209 | 31 | 87.1% |
| | 1.00 | 14 | 119 | 89.5% |
| | Overall Percent | 59.8% | 40.2% | 87.9% |
| Testing | .00 | 3 | 1 | 75.0% |
| | 1.00 | 0 | 5 | 100.0% |
| | Overall Percent | 33.3% | 66.7% | 88.9% |
| Holdout | .00 | 4 | 1 | 80.0% |
| | 1.00 | 1 | 3 | 75.0% |
| | Overall Percent | 55.6% | 44.4% | 77.8% |

Table 2 is divided to three parts: Training, Testing and Holdout. In the first part of table (Training), 87.1 percent of customers who classified to good customers and 89.5 percent of customers who classified to bad customers were estimated correctly. In second part of table (Testing), the correct predicted percents are 75% and 100%. In third part of table (Holdout), the correct predicted percents are 80% and 75%. Also according to Table 3, overall percent error for holdout data is 22.2%. It is clear that in provided model, all the variables don't have the same effect on estimation and some are more effective in this model. In Table 2 and Table 3, dependent variable is bad customer. In Figure 2, importance of variables in the model is presented as normalized. Based on this figure, individual loan frequency (X8) and amount of loan (X7) have greatest effect on customer's good or bad estimation.

Table 3. Model summary

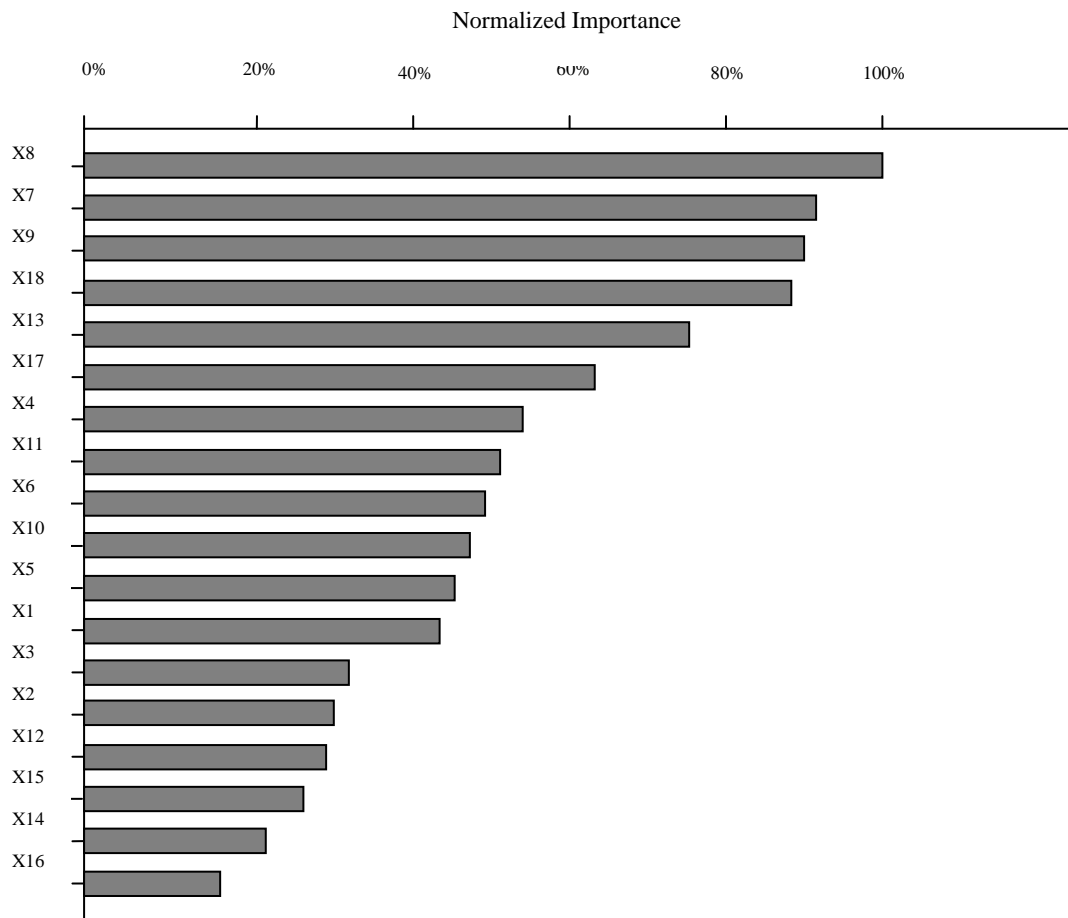| | | |
|---|---|---|
| **Training** | Cross Entropy Error | 111.195 |
| | Percent Incorrect Predictions | 12.1% |
| | Stopping Rule Used | 1 consecutive step(s) with no decrease in error |
| | Training Time | 00:00:10.047 |
| **Testing** | Cross Entropy Error | 1.950 |
| | Percent Incorrect Predictions | 11.1% |
| **Holdout** | Percent Incorrect Predictions | 22.2% |

Normalized Importance



Figure 2. Importance of variables

## 5. Conclusion

In this paper, we have presented an application of artificial neural network to measure bank customers' credit risk. We discussed the importance and necessity of customers' credit risk measurement and explained the architecture of ANN models. After determining of required variables, the collected data were entered into the model. Results of this study show that individual loan frequency and amount of loan have most important effect in identifying classification criteria of good customers and bad customers and also status of customer's bank account, history of customer relationship with bank and received services have least important effect. It means that bank managers and policy makers should focus on number of times that customers have received loans from bank and each time, how much was the amount of loan. This strategy reduces risk of non-repayments and increases the bank's profit.

## References

Angelini, E., di Tollo, G., & Roli, A. (2008). A neural network approach for credit risk evaluation. *The Quarterly Review of Economics and Finance*, *48*(4), 733-755. http://dx.doi.org/10.1016/j.qref.2007.04.001

Atiya, A. F. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on Neural Networks*, *12*(4), 929-935. http://dx.doi.org/10.1109/72.935101

Curry, B., Morgan, P., & Silver, M. (2002). Neural networks and non-linear statistical methods: an application to the modelling of price–quality relationships. *Computers &amp; Operations Research*, *29*(8), 951-969. http://dx.doi.org/10.1016/S0305-0548(00)00096-4

Eletter, S. F., & Yaseen, S. G. (2010). Applying Neural Networks for Loan Decisions in the Jordanian Commercial Banking System. *International Journal of Computer Science and Network Security*, *10*(1), 209-214.

Ghodselahi, A., & Amirmadhi, A. (2011). Application of Artificial Intelligence Techniques for Credit Risk Evaluation. *International Journal of Modeling and Optimization*, *1*(3), 243-249. http://dx.doi.org/10.7763/IJMO.2011.V1.43

Gouvêa, M. A., & Gonçalves, E. B. (2007). *Credit Risk Analysis Applying Logistic Regression, Neural Networks and Genetic Algorithms Models.* Paper presented at the Production and Operations Management Society (POMS), Dallas, Texas, U.S.A.

Hall, M. J. B., Muljawan, D., Suprayogi, & Moorena, L. (2009). Using the artificial neural network to assess bank credit risk: a case study of Indonesia. *Applied Financial Economics*, *19*(22), 1825-1846. http://dx.doi.org/10.1080/09603100903018760

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359-366. http://dx.doi.org/10.1016/0893-6080(89)90020-8

Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Syst. Appl.*, *37*(9), 6233-6239. http://dx.doi.org/10.1016/j.eswa.2010.02.101

Matoussi, H., & Abdelmoula, A. (2009). Using a Neural Network-Based Methodology for Credit–Risk Evaluation of a Tunisian Bank. *Middle Eastern Finance and Economics*, *4*, 117-140.

Pacelli, V., & Azzollini, M. (2011). An Artificial Neural Network Approach for Credit Risk Management. *Journal of Intelligent Learning Systems and Applications*, *3*(2), 103-112. http://dx.doi.org/10.4236/jilsa.2011.32012

Salehi, M., & Mansoury, A. (2011). An evaluation of Iranian banking system credit risk: Neural network and logistic regression approach. *International Journal of the Physical Sciences*, *6*(25), 6082-6090. http://dx.doi.org/10.5897/IJPS10.640

Steiner, M. T. A., Neto, P. J. S., Soma, N. Y., Shimizu, T., & Nievola, J. C. (2006). Using Neural Network Rule Extraction for Credit-Risk Evaluation. *International Journal of Computer Science and Network Security*, *6*(5A), 6-16.

Vasconcelos, G. C., Adeodato, P. J. L., & Monteiro, D. S. M. P. (1999, July 20-22). *A Neural*

*Network Based Solution for the Credit Risk Assessment Problem.* Paper presented at the IV Brazilian Conference on Neural Networks, São José dos Campos.

Yu, L., Wang, S., & Lai, K. K. (2008). Credit risk assessment with a multistage neural network ensemble learning approach. *Expert Systems with Applications*, *34*(2), 1434-1444. http://dx.doi.org/10.1016/j.eswa.2007.01.009

**Copyright Disclaimer**