International Journal of Applied Operational Research
Vol. 5, No. 1, pp. 63-72, Winter 2015

Journal homepage: ijorlu.liau.ac.ir

The Comparison of Credit Risk between Artificial Neural Network and Logistic Regression Models in Tose-Taavon Bank in Guilan

S. M. Taghavi Takyar, R. Aghajan Nashtaei*, E. Chirani

Received: 16 July 2014 ; **Accepted:** 7 December 2014

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. Different statistic methods have been applied for this purpose, such as Discriminant Analysis, Probit Regression, Logistic Regression, Neural Network and so on. Among these methods, Neural Network has been considered mostly because of its high flexibility in recent years. In this research, many efforts have been made to examine the efficiency of Logistic Regression and Neural Network models for credit decision of natural applicants receiving installment loans for selling in Tose-Taavon Bank, Guilan. For this reason, customers who had applied for loans from the beginning of 1388 (2009) to the end of 1392 (2013) and also had complete information files were 376 cases and reviewed based on the independent variables of this research such as applicant's income, facility profit, repayment period, the amount of guarantor's loan, and the type of assurance taken. The result of this survey shows that Logistic Regression and Neural Network models are both highly efficient for predicting applicants' credit risk, but comparing these two models shows that Neural Network is more efficient and more accurate.

Keywords: Credit Risk, Neural Network, Logistic Regression, Tose- Taavon Bank.

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

* **Corresponding Author.** (✉)
E-mail: taghavi.takyar@gmail.com (S. M. Taghavi Takyar)

S. M. Taghavi Takyar, R. Aghajan Nashtaei, E.Chirani
Department of Business management, Rasht Branch, Islamic Azad University, Rasht, Iran

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. [1]

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. [2]

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. [3]

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]

In this paper, the background of research in the field of measuring credit risk is given and after that the conceptual model and theoretical framework of this article are proposed. In the next section, we briefly discus about the methodology used to measure credit risk in the banking systems. Topics in this section include hypotheses, methodology, population and statistical sample, methods of statistical analysis and hypothesis testing, and finally some findings and recommendations for establishing a system to measure and manage credit risk in the banking system of our country are presented.

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. [6]

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 than other parametric statistical techniques. [13]

Bekhet & Eletter (2014) 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 Theoretical models to measure credit risk

Various approaches in the fields of mathematics, statistics, econometrics and researches in operations such as mathematical programming, probabilistic and deterministic simulation, artificial neural networks, survival analysis, game theory, discriminant analysis, logit analysis and probit analysis are involved in developing a model for the accurate measurement of credit risk. Similarly, the development of financial market theories like arbitrage theory, option pricing theory and pricing model of capital assets have contributed to the development of accurate models for measuring credit. [14]

In this article by checking the efficiency of logistic models and artificial neural networks (RBF), we tried to design and explain the most appropriate model for measuring credit risk of customers of Tose-Taavon Bank in Guilan province.

3.1 Logistic regression model

One of the most important classic models often used for prediction is regression model. If the dependent variable is 0 or 1, logistic regression model shall be used [4].

Logistic regression is a model in which predictor variables (independent) can be both in quantitative scale and categorical scale where the dependent variable is a two-level category. These two categories in a way refer to membership or non-membership in a group (companies that are not able to repay their loans). In logistic regression, the concept of odds is used for the amount of dependent variable. In statistics, odds means the ratio between probability of an event $\left(\mathrm{P_i}\right)$ and probability of the non-occurrence of $\left(1-\mathrm{P_i}\right)$. Probability between 0 and 1 may vary while odds can be more than 1. The key word in logistic regression analysis is an element named Logit which is the natural logarithm of odds. Logistic regression is defined as follows: [14]

$$Z_i = Ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \sum_{i=1}^{n} \beta_i X_i \tag{1}$$

In logistic regression model, the probability of required event (non-repayment of the loan facility by the customer) is calculated by the following equation:

$$P_i = \pi_i\left(x_1, x_2, \ldots x_k\right) = \frac{e^{\beta_0 + \sum_{i=1}^{k} \beta_i x_i}}{1+e^{\beta_0 + \sum_{i=1}^{k} \beta_i x_i}} \tag{2}$$

3.2 Radial basis function neural networks (RBF)

Radial basis function neural networks are the most widely used basic functions. 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. [16]

RBF network architecture:
RBF 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.

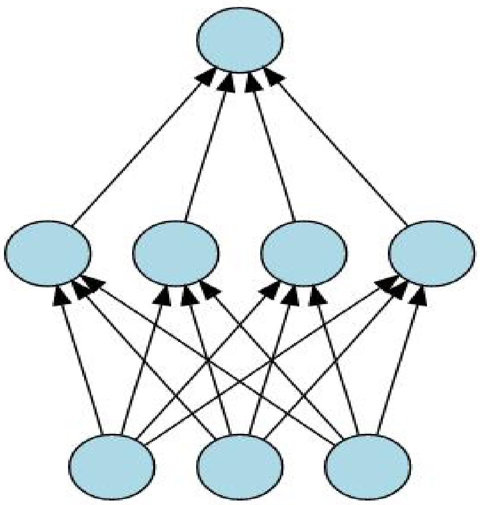


Fig. 1 schematic form of a neural RBF network

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:

$$\varphi(x) = \sum_{i=1}^{N} a_i p(\|x - c_i\|) \tag{3}$$

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].

Teaching RBF networks:
RBF networks are usually taught by a two-step algorithm. In the first stage, vectors of Central c_i for RBF functions in the hidden layer are selected. This can be done in many ways: centers

can be randomly sampled from some of the examples, or K-means clustering can be used. It should be noted that this stage is unsupervised.
The second stage is simply fitted by a linear model with w_i coefficients for the output of the hidden layer according to objective function.

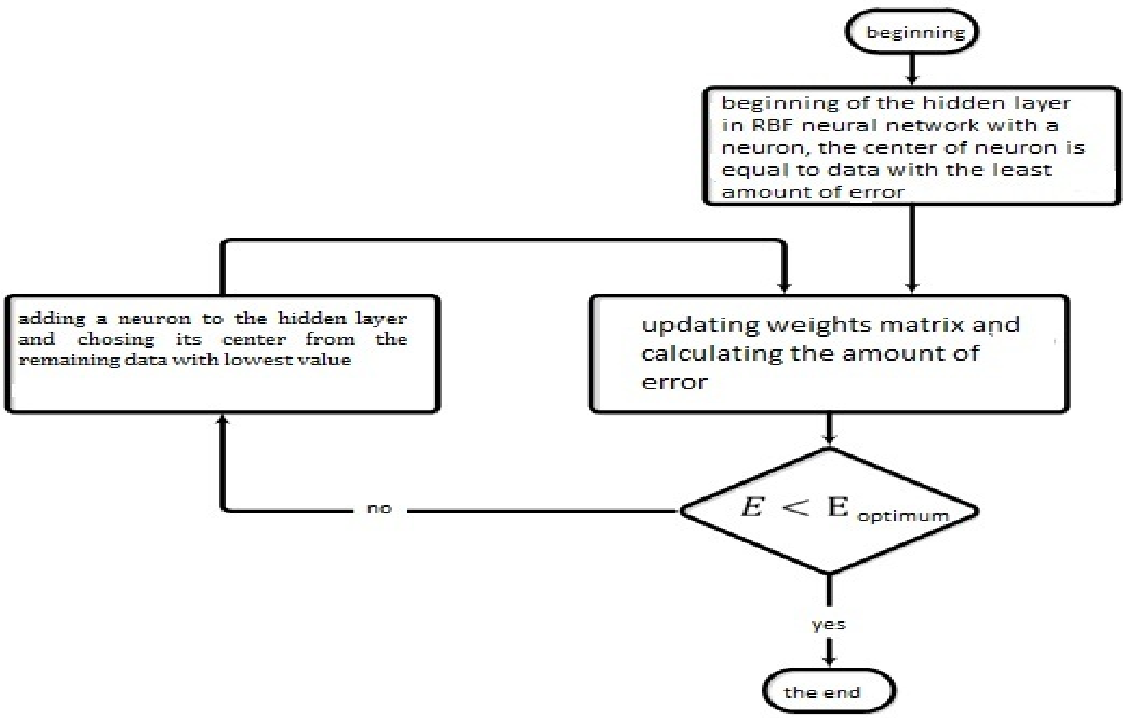
Algorithm for implementing RBF neural network
1. We first consider the hidden layer of the neural network architecture with a neuron. The center of neuron is equal to the data with the least amount of error that is typically considered to be an optimal value for error. This value determines the desired frequency in algorithm for getting the ultimate answer of RBF network.
The error is usually considered as follows:

$$E = \sum_p \sum_k (y_k(x^p) - t_k^p)^2 = \sum_p \sum_k (\sum_{j=0}^{M} w_{kj}\phi_j(x^p, c_j, \sigma_j) - t_k^p)^2 \quad (4)$$

In which x^p is the input vector obtained by p-th sampling , y_k is the output of network and t_k^p is as follows:

$$y_k(x^p) = \sum_{j=0}^{M} w_{kj}\phi_j(x^p) = t_k^p \quad (5)$$

In which w_{kj} is the weights matrix (coefficients matrix) and $\phi_j(x^p)$ is the matrix of stimulation function.
2. Based on the input and output, weights are updated and the amount of error is calculated.
3. If the error was larger than the desired error
* A neuron is added to the hidden layer
* Its center is chosen between the remaining data with lowest value
4. Continuing this trend until the amount of error is lesser than optimum amount of error.
The algorithm is shown as a flowchart in the following figure



4 Research hypotheses

Checking hypotheses in applied researches is of particular importance. In the following hypotheses, an appropriate model for predicting credit risk is identified:
First hypothesis: the logistic regression is an efficient model for credit decision of bank customers,

The second hypothesis: artificial neural network model is an efficient model for credit decision of bank customers.

5 Research variables

In this study, the credit decision of customers is the dependent variable which is considered as a binary variable with a value of 1 for the accepted application (credit risk is low) and 0 for the rejected application (credit risk is high); independent variables of this study include: the profit of facilities granted by the bank, the amount of mortgage, the total income of the applicant, the period of repayment (number of installments) and guarantor and the type of assurance taken by the Bank

6 Statistical population and sample

The studied population consisted of nearly 7652 actual credit cases belonging to natural customers of Tose-taavon bank in Guilan which used the facilities provided by this bank from the beginning of 1388 (march of 2009) till the end of 1392 (march of 2014). 376 cases involved the natural customers of loan installments for buying basic goods which had a complete data file according to the parameters evaluated in this study. These cases are divided in two groups: good customers (lower credit risk) and bad customers (higher credit risk). Out of the 376 cases, 302 cases belonged to good customers and 74 cases involved bad customers.

7 Methods of data analysis and hypothesis testing

In this study, Pearson correlation analysis was used to identify a significant relationship between each independent and dependent variables and coefficients significance test (Wald Test) and likelihood test were used to evaluate the efficiency of designed logic model.

8 Research findings

8.1 The first hypothesis test result:

Correlation analysis between studied variables

The results showed that there was a significant negative correlation between the amount of the loan (Rials) and credit decision of bank customers at the level of 99% (r = -0/277). The results also showed that there was a significant positive between the number of installments and credit decision of bank customers at the level of 99% (r = 0/155) (Table 1).

According to Table 1, there was no significant relationship between the variables of loan interest rates (%) and the guarantor and the type of assurance received (%) and monthly income of the applicant (Rials) and credit decision of bank customers (Table 1).

Table 1 Pearson correlation between credit decision of bank customers and studied variables

Variables	Correlation coefficient	significance level
Loan interest rate (percent)	0/059ns	0/251
The amount of loan (Rials)	-0/277**	0/000
The number of installments	0/155**	0/003
Guarantor and the type of assurance received	-0/049ns	0/344
Monthly income of the applicant (Rials)	0/046ns	0/376

ns is non-significant and * and ** are significant correlations at 1 and 5 percent levels

Logistic regression analysis

The results showed that the credit decision of the bank customers is determined by variables such as loan rate (Rials) and the number of installments. Also, Wald values in Table 2 show that variables of loan rate (Rials) (42/575) and the number of installments (25/560) had a better contribution to predict the credit decision of bank customers (Table 2).

Another finding of logistic regression model is the significant positive impact of loan rates (Rials) and the number of installments on credit decision of bank customers (Sig <0.01). The likelihood amount for this model is 313/801 that indicates the ability of model to predict credit risk.

Table 2 Logistic regression analysis to estimate the factors influencing the credit decision of bank customers

variables	B	S.E.	Wald	Sig.
Loan rate (Rials)	0/000	0/000	42/575	**000/0
Number of installments	0/124	0/024	25/560	**000/0
Constant	0/058	0/599	0/009	**923/0

Likelihood Ratio Statistic (L.R. Statistic) = 313.801
Probability (L.R Statistic) = 0.000
Mcfadden R^2 = 82.7

* and ** are significant at 95 and 99 percent of certainty respectively.

Based on this outcome, logit model can be presented as follows:

$$Ln(\frac{P}{1-P}) = 0/058 + 0/0001X_1 + 0/124X_2$$

In which X1 = loan rate and X2 = number of installments.

Also, the percentage of correct prediction in this model is 82/7 percent. This means that 82/7 of the observations are properly segregated.

8.2 The second hypothesis test result

The number of data samples is 376 in which 326 data are intended network training and 50 data are intended for network testing. As previously stated, there are tunable parameters in the network that can be adjusted using trial and error and expert knowledge. According to this, network was coded in MATLAB with a hidden layer and taking 23 units for every network with 5 inputs and 1 output. It should also be noted that the Green's function centers are selected randomly and in a fixed way where a batch training method is used. It should be noted that MATLAB tools were not used for encoding because these pre-written tools have limitations in implementing neural network.

RBF neural network with the settings mentioned was taught by 85% of the available data and its training time with optimized written code takes about 12 seconds which is acceptable for the network.

In the following, it was necessary for the network to be evaluated. The desired output and the output obtained from the training network are shown below.

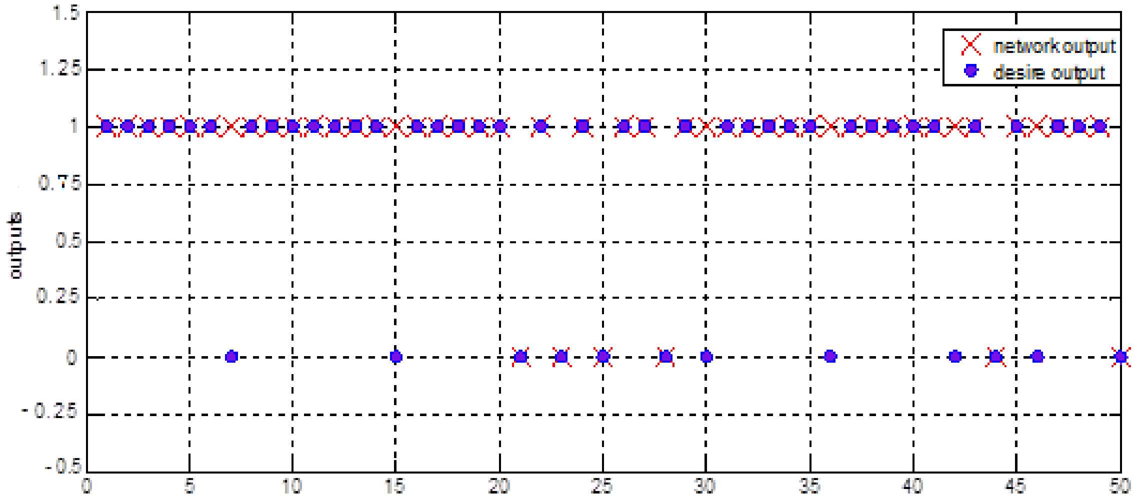


Fig. 2 The desired output and the output obtained from the training network

The network has been able to predict the output with an 88 percent accuracy and the desired output and network are shown below at the same and can be seen that the network only precidited 6 data incorrectly.

9 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:

1. 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.
2. Designing and establishing a software system for logistics and neural network models to predict the credit risk of customers.

3. 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.
4. Reviewing and revising the credit risk prediction model based on the continuous feedback of results.
5. Determining the credit capacity of each customer based on the determined amount of credit risk based on this model.
6. 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.
2. Yurdakul, F., (2014). Macroeconomic Modeling Of Credit Risk For Banks. Procedia Social and Behavioral Sciences, 109,784-793.
3. Bekhet, H. A., Fathi Kamel Eletter, S., (2014). Credit risk assessment model for Jordanian commercial banks : Neural scoring approach. Review of Development Finance 4, 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. 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 micro-economic participation in improving the bank's credit risk. Journal of knowledge and research investment. Third year. Number Nine. 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. No. 17. Page S133-158.
7. Mirzaei, H., Nazarian, R., Baegheri, R., (1390). Surveying factors affecting credit risk of banks legal entities. Quarterly Journal of Economic Research. Nineteenth year. Number fifty eight. Pages 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. The second issue. Pages 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. No. 32. Pages 85-110.
10. Boguslauskas, V., Mileris, R., (2009). Estimation of credit risk by Artificial Neural Networks models. Eng. Econ. 4, 7-14.
11. Khashman, A., (2010). Neural network for credit risk evaluation : investigation of different neural models and learning schemes. Exp. Syst. Appl. 37 (9), 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. 61 (4), 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. 40(1), 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. Number forty-four. Pages 213-234.
15. Sajadi, S., (2012). Design of radial basis function neural networks based on Fuzzy K nearest neighbors. Acta Compuare, 1, 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.