Presenting a Model for Validation Companies and Institutions for Ranking Bank Customers with a Neural Network and Logistic Regression Approach

Mahboubeh Mahmoudzadeh

Master of Financial Managemen Facualty of management,university of tehran ,Iran

Abstract: Today, one of the main concerns of banks and credit institutions in allocating bank loans and credits is reviewing the risk of credit customers. One of the ways of reducing credit risk and delayed payments can be funding and launching validation companies and institutions. A validation institution is an independent organization which gathers public and legal information, information related to identity, credit transactions and payment records of customers and organizations according to current laws and those that must be implemented. One of the important and main tasks of validation institutions is reviewing the credit information of customers who have visited credit institutions and banks in order to receive loans. This review is in the respect of measuring the rate of customers' credit and also ranking credit customers. The present research is in the respect of presenting a model to these validation institutions in the respect of ranking regal customers of banks with an approach to multilayer perceptron neural network and logistic regression. In this research, firstly, the effective variables on credit risk have been extracted from credit reports and financial statements and ultimately, the ranking model has been presented through two methods: logistic regression and neural network. And among the two methods of ranking credit customers, the multilayered perceptron artificial neural network method is a more efficient method by considering that it has been more accurate in ranking.

Keywords: Neural Network, Multilayered Perceptron, Logistic Regression, Ranking

Introduction

One of the important issues which is crucially important in allocating bank credits is reviewing and evaluating the risk of credit customers. Achieving or not achieving the predicted rate of return in evaluating capital projects is called risk. In other words, not being confident about receiving the future earnings of investment is said to be the risk. Business cycles, inflation, political conditions and many other factors affect lack of confidence in future. One of the best principles of investment in all fields is that the risk obtained from investment shall be suitable for the return of that investment.

The thing that is especially important in giving credit loans is reviewing and evaluating the probability of the original capital with the profit obtained from giving credit. In fact, reviewing and evaluating the risk of banking industry has an approach equivalent with the investment industry. In terms of investing in a specific company, the investor studies and evaluates the conditions of the company and after financial analyses and making sure that investing in that company has more benefits than other companies, the investor attempts to invest in that company and whenever he/she needed the price of investment, he/she can easily sell his stock with a slight change and receive his/her capital. However this issue is looked at from another angle in the banking industry which means that in this section, the reference company visits the bank to receive a loan and attempts to get a loan. The required bank also acts after reviewing the documents of the company. It is obvious that here bank is the investor which mean that the bank also tries to give bank loans to a company that has the ability of paying the loan itself and its interest in due time. But, here, the

only difference between a real investor and a bank is that, on one hand banks consider a fixed interest for each kind of Islamic contracts and receive them in the due time along with the loan itself.

Here, the thing that becomes significant is that banks try to give their loans to companies that are able to have a return in proportion with the interest of the given loans in addition to having a low risk; thus if a company's interest is much more than the interest obtained from giving loans, in receiving credit loans, it will be the same as the company which has return in proportion of the interest of the given loans. Therefore, the ban tries to choose a company with the minimum of risk among these companies. On the other hand, at the time of getting credit loans which is considered to be the same as bank's capital, they cannot act like a real investor.

This means that the bank determines a time for the repayment while giving loans and asks its customers to pay back the loan itself along with its interest in the due time. In other words, repaying the loan before the due time is not possible. In any case, here, evaluation of the risk of the companies seeking loans for the banks is also crucially important. This means that if a country is not able to act successfully in the economic environment, it would not be able to repay the given loans. Thus the bank giving loans will also face several problems.

Review of literature
Artificial neural network

Artificial neural networks are a model for processing information which has been created by imitating the biological neural networks such as human brain. The key element of this model is the new structure of its data processing system and it has been composed of many elements (neuron) with strong internal connections which work in coordination with each other for solving special issues. Artificial neural networks, by processing the empirical data, transfer the knowledge or the law hidden beyond the data to the network structure which is called learning. Basically, the ability of learning is one of the most important features of an intelligent system. A system which can learn is more flexible and is programmed easier; therefore it can be more responsive in terms of new issues and equations.

The used neural network is the multilayered perceptron neural network. Multilayered perceptron has been arranged in layers. Naturally, they are referred to as multilayered perceptron. The overall structure of this model has been presented in figure (1).
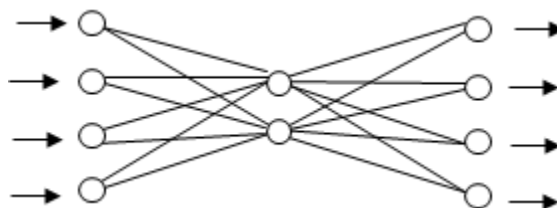

Figure 1.Multilayered perceptron

The new model has three layers: an input layer, an output layers and a layer between them which is not directly connected to input data and output outcomes. In fact, this layer is called a hidden layer. Each unit acts like a perceptron in the hidden layer and output layer. The only difference is that the sigmoid function has been used instead of step function. Units of the input layer is only responsible for distributing the input values to the next layer and therefore, no calculation is done. By changing the nonlinear function from a step function to sigmoid function and adding a hidden layer, we are forced to change the rule of learning model. Our new model shall be able to recognize more complex models. The learning rule of multilayered perceptron is called the delta overall rule or "post-spreading rule".

Logistic regression

Logistic regression is regression statistical model for binary dependent variables. This model can be considered as an extended linear model which uses the logit function as a link function and its error follows polynomial distribution. This model is like:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i},$$

Where: i = 1,...,n,

$$(2\text{-}2) \qquad p = \Pr(Y_i = 1).$$

$$(3\text{-}2) \qquad p = \Pr(Y_i = 1 | X) = \frac{e^{\alpha + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i}}}{1 + e^{\alpha + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i}}}.$$

The logistic regression method is used when the dependent variable (response variable) and has two modes. This method has also presented various statistical tests and has a wide recognition ability in itself.

Credit Ranking

Credit ranking is a statistical tool which is used in order to determine the degree of customers' risk. Credit ranking can determine the error probability in the determination of parameters based on which the customers are given credit. Credit ranking helps the bank improve the process of efficiency and specifies proper methods for credit decisions. Credit ranking is a statistical method which is used in order to predict the probability of the error of future loan borrowers or how many of the current borrowers experience failure. Statistical techniques use the historical data in this method. Credit ranking attempts to specify the impact of various demanders' features and personality; but it can be said that their most important operation is giving loans to the demanders. It is necessary for these institutions to establish an efficient system for doing this important activity of theirs so that the operation of giving loans in the current competitive markets would have the required speed and efficiency and also the probability of the original loans not returning, which is most important base of credit and financial institutions, would be minimized. From this perspective, one of the basic applications of the credit ranking systemsbecomes apparent which is basically a way for recognizing the differences between the groups in a population.

Thomas mentions two basic reasons for developing the current credit ranking systems:

Due to the economic conditions, the institution needs to identify the techniques of predicting the consumer risk for automatically complying with the new conditions.

Companies hope to identify customers who are very profitable instead of trying to reduce customers who don't repay the loan in due time.

Credit ranking systems

Credit ranking systems can be divided into three categories:

Judgment systems

Ranking based on statistical techniques

Intelligent systems

Judgment systems are very slow and expensive. Generally, when the number of demands is high and the number of experts is low, these systems do not have the necessary efficiency. In terms of the statistical methods, each of these techniques also requires specific assumptions. It is obvious that with the absence or fading of the presumptions, the accuracy of the outputs is doubted. When the laws of decision making are clear and information is valid, expert systems are of great help to solving the problems. But, the laws of the institutions that give loans are often not clear and information does not even exist or a part of information is not accurate; in this case the neural networks are a very proper option.

Statistical methods and population
Statistical population
      Given the fact that the ultimate purpose of this article is to present a model for validation companies and institutions for ranking the legal customers of the bank; therefore the statistical population includes legal customers whose credit report has been prepared by the validation company.
      Given the fact that the required statistical population for this research is the legal customers of Mellat Bank and by considering the fact that necessary information has been extracted from the reports of the company Tadbirgaran-e Behsaz-e Mellat; therefore the statistical population is the legal customers of Mellat Bank whose report has been prepared in the interval of July, 2011 to December, 2011 by this company and there is a number of 1850 reports.

The variables used in ranking
      Variable is a concept to which more than two values or numbers are allocated. In other words, features which can be observed or measured are referred to as variables and two or more values or numbers can replace them. The number or the value allocated to the variable is indicative of the change from a person to another or from a mode to another.
Variable is divided into two categories based on the role it has in the research:
Independent variable
Dependent variable
      The independent variable is the driving variable within data which is measured, manipulated or selected by the researcher so that its impact or relationship with another variable is determined. The dependent variable is the response variable, the output or the criterion and it is an aspect of the behavior of an organization which has been stimulated. The dependent variable is observed and measured so that the impact of the independent variable on it would be clarified and specified. The dependent variable is the credit risk of the legal customers of the bank.
      The independent variable is the probable causes of the dependent variable which includes the following items:
Liquidity ratios, activity ratios, debt or leverage ratios, profitability ratios

Liquidity ratios
Current ratio $= \frac{\text{current assets}}{\text{current debts}}$
Immediate ratios $= \frac{\text{current assets} - \text{savings}}{\text{current debts}}$

Activity ratios
Ratio of assets' turnover $= \frac{\text{net sales}}{\text{total assets}}$

Debt or leverage ratios
Debt ratio $= \frac{\text{total assets}}{\text{rights of stockholders}}$
Ratio of total debts to total assets $= \frac{\text{total debt}}{\text{total assets}}$
Ownership ratio $= \frac{\text{rights of stockholders}}{\text{total assets}}$

Profitability ratios
Ratio of profit margin $= \frac{\text{net profit}}{\text{net sales}}$
Ratio of return of investment $= \frac{\text{net profit}}{\text{total assets}}$
Ratio of return of the rights of stockholders $= \frac{\text{net profit}}{\text{rights of stockholders}}$

Models' fitting
      In order to fit the model of credit ranking, two methods have been used: neural network and logistic regression model. In order to select the best model, we compare their accuracy in the classification with each

other in such way that the method which has the minimum error in the classification is selected as the reference model.

Logistic regression method

In order to have a better understanding of the effect of the dependent variables, the Logit analysis predicts the probability of the occurrence of an event instead of predicting whether or not that event occurs; through this the dependent variable can include a range of rates between zero and one. In order for the relationships of the independent variable and dependent variables to be a number between zero and one, the logit analysis uses an assumed relationship between these. At the low levels of the dependent variable, the probability becomes closer to zero. And therefore the variables will be like a curve and the more the dependent variables increase, the slope begins to be reduced at a level that the dependent variables of probability get close to one. However, the value is never more than one. Regression linear models cannot comply with such relationship and therefore the logit analysis has been created and expanded; but this unique feature of the assumed relationship between independent and dependent variables requires a different approach in estimating the model and interpreting the coefficients.

Model fitting

By using the output of the SPSS19 software, it is concluded that only 317 cases have been used out of the reviewed 318 cases in the logistic regression analysis. Also the output obtained from the statistic of Omnibus test for the model show that since the P-Value obtained from this test is smaller than 0.05. Therefore, the model fitting will be satisfying. Also the R^2 obtained from the model fitting is approximately 74%. In other words, the percentage of the changes of the independent variable, which is expressed by the independent variable, is 74%. The accuracy of the ranking of the model is accurately approximately 99.1%. also out of 9 variables involved in the method of logistic regression model, 9 variables affect the independent variable; therefore the defined model is as follows:

$Y = -493.65 + 55.53v1 + 37.495v2 - 16.281v3 + 77.07v4 + 41.72\ v5 + 55.74\ v6 + 74.31\ v7 + 59\ v8 + 23.66\ v9$

Where the variables are as follows:

V1= current ratio
V2= immediate ratio
V3= ownership ratio
V4= the ratio of total debt to total assets
V5=debt ratio
V6= ratio of profit margin
V7= rate of the return of investment
V8= ratio of assets' turnover
V9= rate of the return of the rights of stockholders

By considering the model, we understand that the variable of ownership ratios has a diverse relationship with the dependent variable (credit risk); therefore the productive statistic of the ranking of the importance of the dependent variables is as mentioned in table (1):

Table 1.ranking of variables' important by using logistic regression

Degree of significance	Order or variables' ranking
0.025	Debt ratio
0.024	Ratio of assets' turnover
0.021	Ratio of profit margin
0.009	Ratio of total debt to total asset
0.009	Current ratio
0.007	Immediate ratio
0.001	Rate of the return of the rights of stockholders
0.001	Ownership ratio
0.001	Rate of the return of investment

The most important effective factor on the dependent variable in this method is the variable of debt current and the least effective variable on the dependent variable is the variable of the rate of the return of investment.

Neural network method

For the credit ranking of the legal persons, the neural networks method, which has been made of intelligent networks, has been used with is of the perceptron (multilayered kind). At this stage, the purpose is also specifying the number of the middle layers with the minimum error and the maximum percentage of accurate ranking. In other words, the most important neural network analysis is the network structure with the minimum middle layers and minimum error and prioritization of independent variables.

In this method, which was analyzed with the SPSS19 software, out of 318 cases, 234 cases were selected as training sample and 83 cases were classified as the hold out sample.

Structure of multilayered neural network

The architecture of the obtained neural network includes 9 input layers, 2 middle layers and 1 output layer with two levels. The performance function used in the middle layer is the tangent hyperbolic and the error function used in the input layer is an entropy error function and the error function used in the output layer is the entropy error function and the performance function is the softmaxfunction. And also the percentage of inaccurate prediction in the training sample is equal to 0.4% and in the hold out sample, the percentage of inaccurate prediction is equal to 1.2%.

A multilayered perceptron can have one or two hidden layers. The hyperbolic tangent function used in this layer gets the real values and turns them into values in the (-1, 1) interval. When the selective structure is automatic, this activating function is used for all of the units of the hidden layers.

The function used in the output layer is Softmax. This function receives the true values and returns them without a change. When the automatic selective structure is selected, in case of presence of the variable dependent to scale, this activating function is used for the output layers.

The used error function is the entropy error function and the reason for selecting such error function is minimize the error of the model.

Generally, the accuracy of the ranking of the model in the neural network model is equal to 99.6% in the training sample. The figure of the obtained neural network is as shown in figure (2).
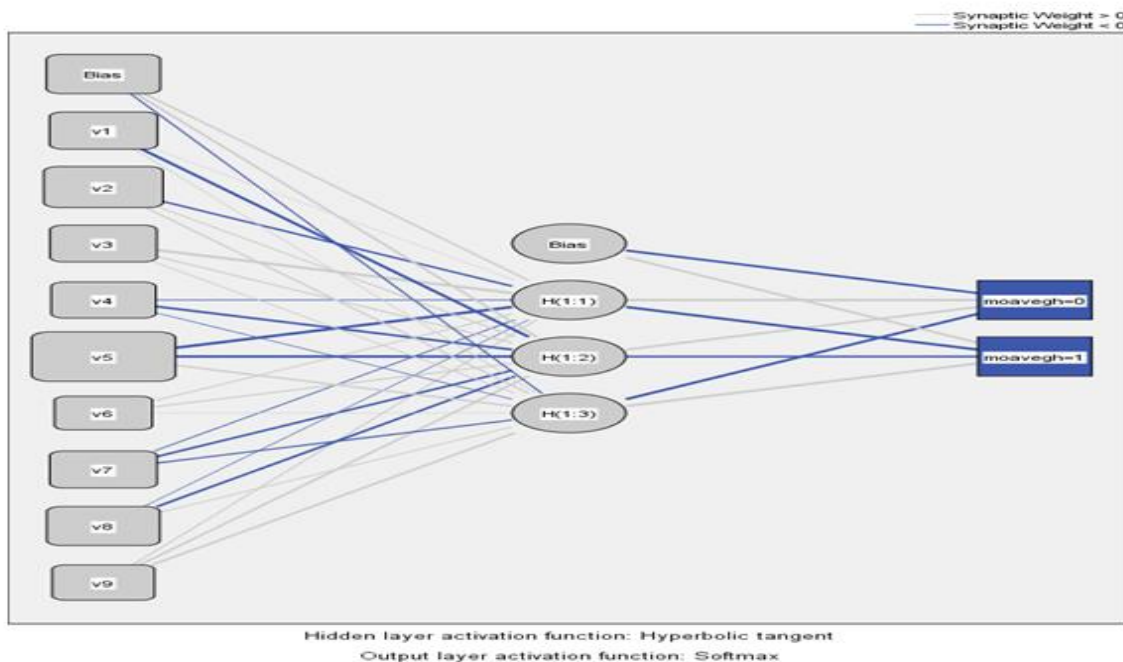

Figure 2. obtained neural network model

The order of the importance of independent variables in the model of neural network structure has been shown in both table and figure (2).

Table 2. order of variables' prioritization by using neural network

Order of prioritization of the variables	Degree of importance
Debt ratio	0.242
Immediate ratio	0.143
Ratio of assets' turnover	0.118
Current ratio	0.097
Rate of the return of investment	0.095
Ownership ratio	0.092
Ratio of total debt to total asset	0.84
Rate of the return of the rights of stockholders	0.074
Rate of profit margin	0.056

Also by considering the above issues, it is concluded that the model of neural network structure uses all of the independent variables of the issue such as logistic regression model for predicting the dependent variable.

Given the items mentioned above and by considering more accuracy in the classification, this result is obtained that among these two methods, ranking with the method of artificial neural network method is a more efficient method.

Conclusion

Among the two methods of ranking credit customers, artificial neural network method is a more efficient method of classification by considering its higher rate of accuracy (according to table (3).

Table 3. comparison of the accuracy of classification of the two neural network and logistic regression methods

Model	Neural network	Logistic regression
Accuracy of classification (minimum error)	%99/6	%99/1
Number of variables involved in the model	9	9

References

Alborzi, Mahmood (1998). "Research method from the perspective of statistics".ShahidBeheshti University, Iran's 4th international conference on statistics.

Altman,Edward.I(2000)"predicting Financial Distress of companies ,Revisiting The Z_score and Zeta Models ,business credit",July

Anthony Saunders,(2000) "Understanding your credit score "

AsghariOskooyi, Mohammad Reza (2002). Application of neural networks in predicting time series, quarterly of Iran's economic researches, 12th issue, Fall.

Erbil and T. Jackson (2004). "Being familiar with neural networks", tanslated by MahmoodAlborzi (Tehran: publication of Sharif University, 2nd edition).

Fensterstock,Albert(2003) "Credit scoring Basics,businesscredit"March

Ghiyasvand, Ahmad (2008). Application of statistics and SPSS software in data analysis.

Haykin,s (1994) ,s .Neural Network :A ComrehensiveFoondation ,Macmillan

Jahankhani, Ali; Parsian, Ali (2009).Financial management.

James Matlews,(www.Generations.org/content/2000/Anin troduction to Nevral Network.ASP)

Jensen,H.L (1996) ,Using Neural Networks For Credit Scoring ,EfrairnTurhanchicage :Inwin

Kiss ,Ferenc.Credit Scoring Processes From Knowledge Management Perspective , Periodical Polytechnics .Soc .Vol11,No.1

Maghsood, Hossein (2006). Prediction of the market of stock exchange in Tehran's stock exchange by using the neural network, Master's thesis of Islamic Azad University.

Menhaj, Mohammad Bagher (2010). Principles of neural network, Amir Kabir Industrial University.

Moosavi, et al. (2009).Ranking the validation criteria of bank customers with Delphi approach, Tabriz University.

Morsman ,E.(1997) "Risk Management and Cerdit Culture ".Journal of lending Cerdit Risk Management ,special Ed. June

Norusis, Marija (2009). Neural networks in SPSS, translated by Fathipoor, Jalilian Amir Reza; Najba, Maziar.

P.Mandic ,A . chambers (2001) ,RecurruntNeuer Networks for Prediction ,2d.ed.

Salimi, et al. (2009).Presentation of a scientific model for the validation of Iranian companies and review of the effect of this credit on the stock price.

Series of approaches for the effective management of credit risk, Bank of Islamic Republic of Iran.

Sinkey.Jr , Joseph .F (1992) ,Commerical Rant Financial Management , 4th Edition macmillan

Thomas ,Lync (2000), A Survey of credit and Bahavioral Scoring :Forecasting Financial Risk of lending to Customer ,International Journal of forcasting

Zirilli ,Josehp .S (1997) ,Financial Prediction Using Neural Networks ,International Thomson Computer press

Zurada,J.M(1992),Introduction to Artificial System,BostonPws . Publishing Company , p.xv.