Iranian Journal of Operations Research
Vol. 12, No. 1, 2021, pp. 184-190

Assessing Default Probability of EN Bank Legal Customers using Support Vector Machine Method

M. T. Taghavifard [1,*], R. Habibi[2]

According to current development in credit allocation and recent economic crises, planning for identification of credit risk has found special importance for investors, banks, shareholders and financial analysts, so that they are able to make proper decisions. Although credit loss is a common cost in banking industry, however, increase in this loss might affect the bank performance. Therefore, there is a strong need to reassess current approaches in risk evaluation of each loan and default rate of loan portfolios. Banks usually have their own internal validation models for loan risk measurement but these approaches are inappropriate and utilize simple mathematical approaches based on incomplete premises. In this paper, we have tried to estimate the possibility of default for legal customers using 20 financial ratios for 200 healthy and 200 unhealthy companies receiving civil participation facilities from Eghtesad Novin (EN) Bank in 2009 and 2010 and 4 approaches for choosing financial ratios including remarks from credit experts of Raah Eghtesad Novin Co., Altman, comparison between averages and choosing correlation attribute. Results show that Support Vector Machine approach can differentiate between healthy and unhealthy companies with average accuracy of 84.63% using all chosen ratios.

***Kewords**: Default risk, Banking, Support Vector Machine, Legal customer*

Manuscript was received on 01/31/2021, revised on 10/14/2021 and accepted for publication on 01/23/2022.

1. Introduction

According to the nature of its activities, mainly equipment and resource allocation, banking industry is highly faced with credit risk. Although operational and liquidity risks are among banking industry risks, however, credit risk has special importance because this attribute itself can be considered as one of the influent factors in liquidity risk outbreak. In the other hand, all customers risks including currency rate risk, interest rate risk, operational risk, etc. are transferred to banks through credit risk and affects these relationships. Although credit loss is a common cost in commercial activities for banks, but increase in losses can threaten the existence of financial institutions. Accurate prediction and initial alerts for the possibility of failure in payment causes risk control, inhibiting wrong decision making, reduction in cost of supervision on debt payments and reduction in credit evaluation time.

Blocking banking resources in previous, delayed and suspicious to be remunerated maturity times, not only reduces facility allocation ability in banks, but also causes problems in achieving banks' goals through negative impact on productivity and reduces economic development. Therefore, planning for prediction of failure in facility repayment has special importance for shareholders, creditors, auditors and bank managers.

One way for choosing right decision while receiving facilities is attention to financial, technical and personal capabilities of customers, which is done in 2 ways, namely inter-organizational approach and rating institutes. International rating agencies such as Moodays and Standard & Poor's are criticized due to inefficiency of their models in risk prediction in debt repayment for banks and companies. Moreover, local and international rating agencies are likely to analyze big companies while financial system and banks need models for risk analysis of SME's as well. Banks also use methods based on experts' judgments including 5P, 5C, and LAPP in order to assess customers but these approaches are

* Corresponding Author.
[1] Associate Professor, Allameh Tabatabai University, Tehran-Iran. dr.taghavifard@gmail.com,
[2] Faculty member, Iran Banking Institute, Tehran-Iran

elementary and based on simple mathematical methods with incomplete premises. Besides, due to high cost and being subjective, each variable has serious problems. Therefore, the need for prediction system for possibility of fault is considered.

Ability to distinguish between good and bad account customers and prediction and controlling unfavorable impacts of customer's fault is one basic component in credit risk management, which is manageable by utilizing proper quantitative methods. The main objective of this paper is to examine efficiency of Support Vector Machine approach to present an accurate prediction for the probability of fault in facility allocation for companies and help banks allocate resources properly. The article is structured as follows. In next section research background is presented. Then, methodology and empirical results are discussed in sections 3 and 4 and results are presented in the 5th section.

2. Literature Review

In this section, domestic and foreign studies regarding this topic are discussed. These studies are about evaluating the probability of failure and reviewing the applications of Support Vector Machine in other fields are ignored.

Discriminant analysis (DA) introduced by Beaver (1966) was the first applied method for bankruptcy prediction. There are many models such as logit, probit, hazard models, recursive partitioning and neural networks for bankruptcy prediction. Beaver (1966) described that some financial ratios such as cash flow to total debt ratio detects business failure well before it actual occurs.

Altman (1968) studied a sample of unhealthy companies between 1969 and 1973. He named this model "Zeta". Their sample consists of 58 unhealthy including 32 manufacturer and 26 retailer. The results show that Zeta model has better prediction ability rather than Z model until 5 years before failure. Aziz et al. (1988) in their research used 6 financial ratios based on committed accounting and 3 ratios for cash flows of operations, and reach to the conclusion that prediction accuracy is not affected by adding variables of operational cash flows. The results show that this algorithm has better prediction accuracy in comparison with Lucitanalysis and neuro- networks.

Martin (1977) used "Early warning of bank failure", wherein a logit regression approach is implemented. Later, Moro et al. (1980)'s work of "Financial ratios and the probabilistic prediction of bankruptcy" based on logit model was published, where he used a large sample,105 bankrupt firms and 2;058 nonbankrupt firms, being different from the previous studies. The drawback of this study is that he disregarded the statements after bankruptcy. Some of the other researchers applying logistic and probit model to bankruptcy analysis can be counted as Wiginton (1980), Zavgren (1983) and Zmijewski (1984). As to the other statistical methods used to predict company failures, there are namely neural networks (Tam and Kiang, 1992). Wilson et al. (1999) asserted that the contribution of the ability and willingness of a firm to pay its creditors to credit analysis is highly important.

Glennon and Nigro (2005) designed a model for predicting companies' failure using Support VectorMahcine. Their study shows that Support Vector Machine model shas better performance than traditional statistical models. Results show that Adabost algorithm has better performance than artificial neuro-networks and its predicting potential is 91.1%. Moro et al. (2010) planned a pattern using neuro-network approaches. In his pattern key financial ratio values are used. Rebeiro'sspatternhas total accuracy of 94% and examines 65 financial ratios in previous studies.

3. Methodology

Support Vector Machine (SVM) is one of the tools for accurate predictions. We have examined SVM approach for evaluating and predicting the probability of failure. A prediction is an estimation of future's events and it is aimed at error reduction in decision making. Predictions are not usually accurate, and the magnitude of error in prediction depends highly on prediction system.

The idea behind SVM stems from Statistical Learning Theory and has a plenty of applications in regression, classification, clustering and generally in estimation. The approach of SVM in its early

stages of development was confined to two-class classifications. This approach was then generalized to multi-class classifications using a variety of techniques (Yaari & Khaanlou, 2008). For example, for a 2-class problem, SVM is applied to nonlinear and seprable problems.

Here, some types of SVM are identified.

a) Hard margin SVM. This type of SVM assumes that data are classified with no error and all data to be classified in proper classes. For 2 classes problems, decision function $D(x)$ is as follows:

$$D(x) = \sum_{i \in S} a_i y_i x_i^T x + b$$

where S is a set of indices for support Vectors. Bias value is

$$b = \frac{1}{|S|} \sum_{i \in S} (y_i - w^T x_i)$$

Finally, unknown data x is classified as follows:

$$\begin{cases} D(x) < 0 \ \ if \ \ x \in class2 \\ D(x) > 0 \ if \ x \in class1 \end{cases}$$

If $D(x) = 0$, x is in boundary condition and is unclassifiable. If learning data are separable, $\{x|-1<D(x)<1\}$ is a generalization district.

b) Soft margin SVM. is a condition in which we accept wrong classification for some data. It means we avoid over-accuracy. In SVM with constant margin, it is assumed that learning data are separable and linear. In order to make our model separable, non-negative and weak variables are entered respectively in relation. This will always lead to acceptable answer. Decision function is obtained like support vector machine with constant margin:

$$D(x) = \sum_{i \in S} a_i y_i x_i^T x + b$$

where S is set of indices for support vectors. Since α_I's are not zero, in above relationship, just support vectors are appeared. For an infinite α_I the following expression holds:

$$b = \frac{1}{|S|} \sum_{i \in S} (y_i - w^T x_i)$$

For increasing accuracy in calculations, we calculate averages. Machine's core: in a support vector system, if learning data are not linear, obtained classification are not so generalizable. So, in order to enhance linear distinguishability, input space is mapped into a space obtained from an internal product with upper dimension named as attribute space.

Decision function is:

$$D(x) = \sum_{i \epsilon S} \alpha_i y_i H(x_i, x_j) + b$$

Like previous method, we obtain average:

$$b = \frac{1}{|U|} \sum_{i \epsilon S} \left(y_i - \sum_{i \epsilon S} \alpha_i y_i H(x_i, x_j) \right)$$

Principally, 2 types of cores are used in support vector machine which are:

- Polynomial core: a polynomial of the order of d is based on the following formula:

$$H(x, x') = (x^T x' + 1)^d$$

- Radial Basis Functions Core: is defined by the following relationship:

$$H(x, x') = \exp(-\gamma \|x - x'\|^2)$$

where, γ is a positive parameter for radius control.

- **Model Selection**: in learning support vector machines, choosing a core and proper value for margin parameter is necessary. In order to find an optimum classification, we must determine marginal parameter. Determining optimum classifier is named as Model Selection. Usually model selection is done using try and error method based on different parameters of core and marginal parameter. Also, in order to determine the accuracy of the model, a K-tuple validation is used, in which learning data are divided into k subset with almost same size which are chosen randomly. Then, selected classifier is learned through k-1 subset and tested using remaining subset. Learning of k subset whose related learning data are omitted are calculated and averaged.

- **Sample size**. In order to determine sample volume, during negotiation with experts of credit bureau of EN Bank, files of allocated facilities for about 5000 companies are announced and based on Coqueranformula and with higher estimation, 400 companies were chosen. Regarding the fact that most researches about failure prediction is based on this premise that the number of chosen companies in each categories are equal, therefore these 400 companies are divided equally between healthy and unhealthy companies, but disrespect to this fact will not reduce the validity of the process. The method of calculating the sample volume based on Cochran formula is as follows:

$$n = \frac{Nt^2pq}{Nd^2 + t^2pq}$$

In the above formula the maximum error permitted (d) is 0.05, confidence coefficient is 0.95, t=1.96 and p and q are 0.5 and N is the population size. Based on the definition from Central Bank of IRI, all companies whose facilities are delayed and suspicious to be remunerated maturity time status are considered as unhealthy companies.

Previous dues, dues for 2 month before or more and less than 6 month of delay, delayed dues, dues for more than 6 and less than 18 month of delay and cut for share repayment, suspicious to be remunerated dues, dues whose time for repayment is passed more than 18 months, burnt dues, a part of dues for credit institutions that, regardless of maturity time are not repayable to due to reasons such as death our failure or other reasons. Determining new maturity times for due repayment, dividing debt for those whose dues are classified in previous maturity times.

Therefore, in this study, based on definition of Central Bank, companies whose condition is in "current" status are classified in healthy companies, and otherwise they are considered as unhealthy.

- **Database**. The following strategy is used about the main financial database for proper regulation of prediction model:
- A set of 200 unhealthy companies is chosen so that there are at most 10 lost data.
- 200 tuple unhealthy companies are sampled randomly so that a balanced set is obtained.
- Lost values are replaced with the nearest year.

In order to run this study, 43 financial ratio is hosen. In choosing these ratios the following subjects are considered:

1. There must be access to information. According to the fact that used information are obtained from credit bureau of EN Bank, this office just presented data about balancesheet for analysis. Therefore, determined ratios are changed on this basis
2. There must be the least overlap between ratios.
3. It must be used in domestic studies.

According to above conditions, finally current ratios, instant, current asset, defensive gap, cash flow, net capital flow, capital flow, maturity period, operation flow period, product to capital inflow, current capital flow, times of receiving, constant capital flow, sell change percent, constant capital to

special value, asset outcome, inflow capital outcome, loan efficiency index, Doupanratio, constant asset flow, interest change percent are added to the model.

- **Variable selection**. Variable selection are done in 5 ways, and best variables are chosen in theses ways:

1. Financial ratios which are used in Rah EghtesadNovin Co.
2. Altman Method
3. Average comparison method
4. Correlation attribute selection method
5. Variable ommittance method

Then model parameters and core are selected. On this basis, the best wayfor variable selection are utilized. The best approach for 2008 is discussed.

- **Changeability reduction**. In order to reduce changeability, main sample is divided into 2 parts. So that with using a part, modelling is done and this is tested on another part. Also, in order to examine the performance,data related to main part are divided to some auxiliaryparts. Moreover, each auxiliarypart are replaced with test part. All operations are done using Matlab Software.

The procedure of evaluation is based on occurrence probability classification matrix:

Condition	unhealthy	healthy	total
Unhealthy	*tp*	fn	Pos.
Healthy	*fp*	tn	Neg.

tp denotes unhealthy condition of a company which is predicted as unhealthy as well and *fn* denotes unhealthy condition of a company and is predicted to be healthy. Also, *fp* shows a company to be healthy and is predicted as unhealthy and *tn* shows that a company is healthy and is predicted as healthy. Pos. and Neg. are the total number of companies which are unhealthy and healthy in practice, respectively.

- Error classification. Type I error (economic risk): a healthy company is considered unhealthy. This error is calculated using $\frac{fp}{fp+tn}$. Type II error (credit risk): an unhealthy company is identified as healthy. This is so much important because predictor should not have a mistake that leads to wrong decision making. In order to calculate this error we use $\frac{fn}{fn+tp}$.

- Accuracy. All conditions that reach to right answer in model. The formula for accuracy is $\frac{tp+tn}{tp+fp+fn+tn}$. To apply data analysis in this paper, we have used EXCEL, SPSS 16, WEKA 3.7.9 and MATLAB softwares.

4. Data and empirical results

In this section, we present obtained results from various approaches of financial ratios, model selection and parameter selection.

4.1. Financial ratios used

In this section, ratios in professional reports from experts are utilized. The reason for choosing this method was that all experts of credit bureau use results from these reports as well and choosing it is mandatory in all files for facilities more than 5 billion Rls. Ratios, limits, prediction matrix, accuracy, 1st type and 2nd type errors are as follows:

a) Ratios: current, instant, due receiving period, product flow period, constant assets flow, total assets flow, sales outcome, assets outcome, shareholders right outcome, debts, privilege, total debts to shareholder rights
b) Limits: interest cost coverage ratio is ignored due to lack of presenting interest cost by bank in most companies.
c) Hypotheses: according to the fact that parameters in support vector machine are selected with try and error approach, therefore, for congruency in all selected methods, we use below constants. K=1- shows that all data are divided into 10 parts and 9 parts of which are used for learning and the remaining one is for testing model. Selected core is Gaussian which has parameters of c andδ.

δ	c	Core	K
1.8	3.9	Gaussian	10

d) Prediction Matrix:

		Reality	
		unhealthy	healthy
Prediction	Unhealthy	153	36
	Healthy	47	164

e) Accuracy, type I and II error:

Type II error	Type I error	Model Accuracy
23.5%	18%	79.25%

4.2. Financial ratios used by Altman method

Based on average ratios, characteristic variables are selected which includes some financial ratios with totally different behavior in healthy and unhealthy companies. Based on Altman theory (1968), these variables have significant role in predicting the probability of failure.

In order to examine this issue, average for all financial ratios for both types of companies are calculated. In order to see the procedure of these ratios I healthy and unhealthy, we use linear plot. Obtained results are shown in the following Figure:

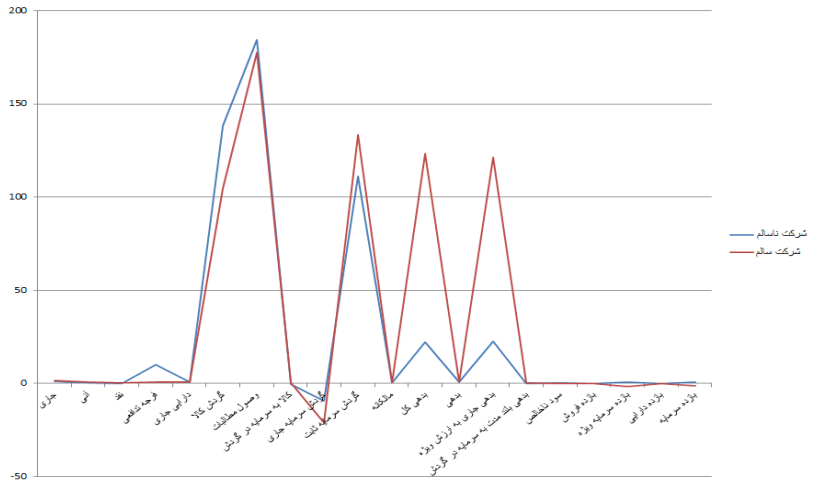

Fig.2 :Ratio examination based on Altman method

5. Conclusion

The probability of default for legal customers is estimated using 20 financial ratios for 200 healthy and 200 unhealthy companies receiving civil participation facilities from Eghtesad Novin (EN) Bank in 2009 and 2010 and 4 approaches for choosing financial ratios including remarks from credit experts of Raah Eghtesad Novin Co., Altman, comparison between averages and choosing correlation attribute. Results show that Support Vector Machine approach can differentiate between healthy and unhealthy companies with average accuracy of 84.63% using all chosen ratios.

References

[1] Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance, pages 589–609.
[2] Aziz, A., Emanuel, D., and Lawson, G. H. (1988). Bankruptcy prediction - an investigation of cash flow based models. Journal of Management Studies, 25(5):419–437.
[3] Beaver, W. (1966). Financial ratios as predictors of failures. Journal of Accounting Research, 5:71–111.
[4] Fan, A. and M., P. (2000). Selecting bankruptcy predictors using a support vector machine approach. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, pages 354–359.
[5] Glennon, D. and Nigro, P. (2005). Measuring the default risk of small business loans: A survival analysis approach. Journal of Money, Credit and Banking, Blackwell Publishing, 37(5):923– 47.
[6] Martin, D. (1977). Early warning of bank failure: A logit regression approach. Journal of Banking and Finance, pages 249–276.
[7] Moro, R., Hoffmann, L., and Härdle, W. (2010). Learning machines supporting bankruptcy prediction. SFB 649 "Economic Risk".
[8] Tam, K. and Kiang, M. (1992). Managerial application of neural networks: the case of bank failure prediction. Management Science, 38(7):926–947.
[9] Wiginton, J. (1980). A note on the comparison of logit and discriminant models of consumer credit behaviour. Journal of Financial and Quantitative Analysis, 15(3):757–770.
[10] Wilson, N., Hope, R., and Summers, B. (1999). Predicting corporate failure and payment behaviour: Addressing some pertinent issues for practitioners. Credit Management Research Centre, Leeds University Business School.
[11] Zavgren, C. (1983). The prediction of corporate failure: The state of the art. Journal of Accounting Literature, 2:1–38.
[12] Zmijewski, M. (1984). Methodological issues related to the estimation of financial distress prediction models. Journal of Accounting Research, 20(0):59–82.