BANKRUPTCY PREDICTION BY USING SUPPORT VECTOR MACHINES AND GENETIC ALGORITHMS

SALEHI Mahdi
Ferdowsi University of Mashhad, Iran

ROSTAMI Neda
Islamic Azad University Science and Research Khorasan-e-Razavi Branch

Abstract:
The original purpose of this study is comparing of Support Vector Machine and Genetic Algorithm and impact of financial ratios on accuracy of bankruptcy prediction. In according to some limitations in traditional statistical models, we used two models of Support Vector Machine and Genetic Algorithm. One of findings in this research is impact of financial ratios on accuracy of bankruptcy predicting and it shows that improper selection of financial ratios do not have high resolutions. Besides, they can decreases accuracy of prediction and may wrong introduce results of the research. Moreover, Support Vector Machine was more powerful than Genetic Algorithm in year's t. However, it cannot be introduced which of them is better. Identifying of the most effective financial ratios as predictor variables and create a more powerful models, which can improve accuracy of prediction and reduce bankruptcy risk and its heavy cost will be decreased. This research focuses on identifying the most effective financial ratios and the most powerful model for predicting of bankruptcy.

Key words: *Bankruptcy predicting, Support Vector Machine, Genetic Algorithm, financial ratios*

1. Introduction

Bankruptcy and failure of company as unfavorable financial phenomena always be important. Since mid 20th century, in according to growth of technology, wide environmental changes and increase competitions have raised chance of failure and in other words, with the advent of joint stock companies and increase demand of companies to funding from outside sources need to assess the financial situation of companies by investors and lenders in order to provide more confidence to invest in the companies. Reasonable concerns of investors and income of their capital and advantages and disadvantages of bankruptcy predicting in companies, micro economics and others have led to wide research in identifying bankruptcy and methods of its predicting which have been done in cross the globe. If we can predict bankruptcy and identifying chance of bankruptcy and then planning and solve the problem, we will troubleshoot and problem-solving methods to prevent from waste of physical and humanity sources. Moreover, this model can be good leader for investors. In fact, predicting bankruptcy is one of methods, tools for predicting future situation of company and chance of bankruptcy can estimate by financial ratios, and these models due to rapid growth of technology and rise in economical competitions are very important for researches and financial decision makers. By using of this model, creditors, investors and managers can be able to predict bankruptcy of companies in several prior and after year's occurrence. Following, in according to findings of these models, we can do necessary actions for preventing losses. In fact, predicting models are separated in two or several levels by at least errors. Thus, in process of predicting, we are following a model which can be able predict of bankrupt or non-bankrupt of a company based on a series of financial data by classified statistics and machine learning which is called supervised learning (Premachandra *et* al, 2009). However, result of researchers indicates that in between predicting bankruptcy models, average accuracy of artificial intelligence in predicting of bankruptcy is higher than statistical models.

2. Literature review

There have been several models for Predicting bankruptcy. Beaver (1966) with a sample of 158 companies, 30 financial ratios to be selected as the best indicator of a company's financial health and it shows that ratios are as one of the best indicators for financial health of companies and he believed ratios are difference in healthy and failure companies. Altman (1968) by using five financial ratios via MDA model, he could propose that MDA model can evaluate bank loans, internal control process and investment options. Frydman *et* al, (1985) were the first persons who used PRA model. Min and Lee (2005) by Support Vector Machines could design model for predicting companies, their results show that Support Vector Machine had better performance compare with other traditional statistical models. Kim and Kang (2012) in their research show superiority of Support Vector Machine compare with genetic algorithm. They selected a sample include of 1200 Korean auditing companies and with select of seven financial ratios (income to total assets ratio, earnings before interest and taxes to interest expense, retained earnings to total assets, cash return, total debt to total assets, Inventory to sales and total assets) showed that Support Vector Machine with accuracy of 72.45% is more powerful compare with the other models. Moreover, optimized models with Genetic Algorithm have higher performance than the classified individual models.

2.1. Financial ratio

Bankruptcy prediction models are one of techniques and tools to predict the future state of companies that evaluate bankruptcy by combining group of risk by financial ratios. in the most models of bankruptcy have used of accounting information that is used often in the form of financial ratios. The role of accounting information derived from the difference between the company's financial was one of the controversial issues in the healthy and unhealthy companies in recent decades. Altman (1968) argued that financial ratios of bankrupt companies to non-bankrupt financial ratios are significantly different. Chen (2011) in their empirical study that included 42 financial, 33 financial ratios, 8 non-financial ratios and used one of the combination of macroeconomic indicators to get the result, financial ratios are more effectively in anticipating of bankrupt. Hossary (2006) believed that 79% of research in the field of bankruptcy prediction used financial ratios as predicting variable. Gahlon (1988) showed that related information about cash flows, including variations in cash flows from operations and cash coverage ratio index, are suitable for predicting bankruptcy. Blum (1974), Deaken (1972), Mensah (1983) argued that the current cash to total debt as a crucial predictor.

2.2. Support vector machines

Support Vector Machine is statistical learning algorithms and it is looking for optimal screen separator by the vectors supporting and this way they can solve classified problems (Yoon, 2010), Support Vector Machine is so simple that it can be analyzed by mathematics. Recently, Support Vector Machine has frequent applications such as credit ratings and predicting time series. Support Vector Machine using a linear model to classify of nonlinear boundary through a nonlinear mapping into a high dimensional feature space to perform linear model created in the new space. Non-linear decision boundaries in the new space will create an enormous page. Thus, Support Vector Machine is an algorithm that is a certain kind of linear model and find that the maximum margin of the enormous page. Maximum of the enormous page can lead to maximum separation in among the classes. Moreover, the nearest educational data to the maximum margin of the enormous page refers to the support vectors. Support vectors can determine the boundary amongst classes and the rest of the training data is not necessary to specify the binary class boundaries (Min & Lee, 2005). Studies conducted by this method have been shown that, Support Vector Machine In terms of performance is better than other methods such as bankruptcy prediction ANN-CBR-MDA (Ravi Kumar& Ravi, 2007).

Vapnik proposed the support vector machines (Vapnik, 1995). The support vector machines (Support Vector Machines) [1] have been highly concerned in recent years. Based on the structured risk minimization (SRM) principle, Support Vector Machines seek to minimize an upper bound of the generalization error instead of the empirical error as in other neural networks. Additionally, the Support Vector Machines models generate the regress function by applying a set of high dimensional linear functions. The Support Vector Machine regression function is formulated as follows (Pai and Lin, 2005):

$$y = w(x) + b,$$

The training algorithms of Support Vector Machines try to find the optimal separating hyperplane by maximizing the margin between the hyperplane and the data and thus minimizing the upper bound of the generalization error. Delivering promising results makes, the Support Vector Machines extensively applicable in many information-processing tasks, including data classification, pattern recognition and function estimation. Support Vector Machines are ordinarily used as binary classifiers that separate the data space into two areas. The separating hyperplane is not explicitly given. A small number of data points, called support vectors, represent it. However, the real data are often linearly inseparable in the input space. To overcome this, data are mapped into a high dimensional feature space, in which the data are sparse and possibly more separable. In practice, the mapping is also not explicitly given. Instead, a kernel function is incorporated to simplify the computation of the inner product value of the transformed data in the feature space. That is, choosing a kernel function implies defining the mapping from the input space to the feature space (Wu and Wang, 2012).

2.3. Genetic algorithm

Genetic algorithm, which is known as evolutionary approach, introduced for the first time by Holland (1975). Genetic algorithm is one of search algorithms and it is based on the genetic living creatures. Following, coding should be done and it is the most common format the chromosomes in genetic algorithms. Each decision variable in Format of binary form and then get together and make this variable chromosome and also a fitness function must be devised solutions to each encrypted value of the run to select parents for reproduction. Combination and mutation operators are used to produce new offspring. This process is repeated several times and produce the next generation population and crowd converged standards will be reviewed and if the process is terminated.

Fundamentally genetic algorithms are a class of search techniques that use simple forms of the biological processes of selection inheritance variation Strictly speaking they are not optimization methods per se, but can be used to form the core of a class of robust and _exible methods known as genetic algorithm based optimizers.

Genetic Algorithms are a family of computational models inspired by evolution. These algorithms encode a potential solution to a specie problem on a simple chromosome-like data structure and apply recombination operators to these structures as to preserve critical information.

Genetic algorithms are often viewed as function optimizer, although the ranges of problems to which genetic algorithms have been applied are quite broad. An implementation of genetic algorithm begins with a population of (typically random) chromosomes. One then evaluates these structures and allocated reproductive opportunities in such a way that these chromosomes, which represent a better solution to the target problem, are given more chances to `reproduce' than those chromosomes that are poorer solutions. The 'goodness' of a solution is typically with respect to the current population (Terela, 2003).

3. Research methodology

The population of the study is listed companies in the Tehran Stock Exchange and the sample included 158 companies from the community who were selected based on article 141 of the Commercial Code Tehran bankrupt each year and each industry will also choose a non-bankrupt. Note that the model Support Vector Machine and genetic algorithm per year the T-1 and T-2 were also financial information relating to the year the 1383-1387 financial statements of the Company the samples were extracted.

H_1: Support Vector Machine is stronger in accuracy of bankruptcy prediction than Genetic Algorithm.

One of the best ways for cross-validation is partitioning a sample of data into training set, and testing set. In this research, k-fold cross validation method was applied; which the sample is randomly partitioned into k subsamples. Of these k subsamples, one subsample is used as the validation data for testing the model, and k − 1 sub-sample are used as training data. In training phase, the training data set applies to the Support Vector Machine algorithm, and then Support Vector Machine calculates Lagrange multipliers and afterwards the weights and bias. Now the data that has not been applied to Support Vector Machine are applied and the error rate is computed.

$$Maximize \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j y_i y_j K\left(x_i^T, x_j\right)$$

$$Subject\ \ to \sum_{i=1}^{n} \lambda_i y_i = \circ$$

$$\circ \le \lambda_i \le C \qquad , i = 1, \ldots, n$$

$$K\left(x_i^T, x_j\right) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2 \times \sigma^2}\right)$$

; That is called Gaussian kernel function.

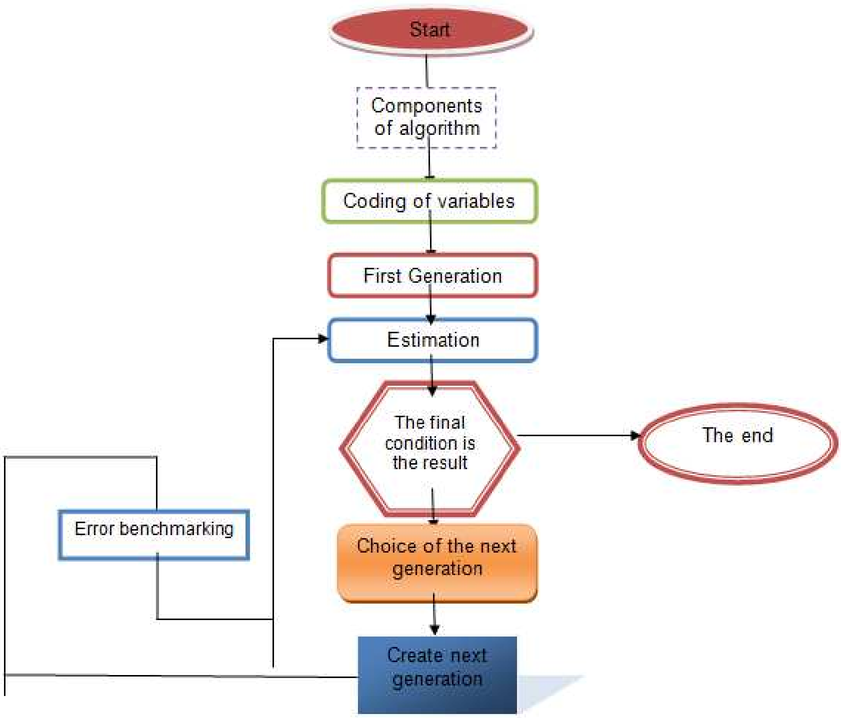
3.1. Potential variables

The predictive variables of the study are divided into two parts. The first part of the first phase of the independent variables is a set of 56 financial ratios of liquidity, profitability, operating and financial leverage, which is through the study of research conducted in the field of bankruptcy. The second stage includes 9 variables out of 56 primary variables which have features the ability to separate and remove variables which cannot separate healthy and unhealthy companies. Therefore, three variables like returns of assets, fixed assets turnover ratio and ratio of equity to capital. Table 1 shows the results of variables. Further, Figure 1 shows the method of variable selection.

Table 1. Independent Variables

Return on investment	Working Capital to Sales	Operating profit to sales
Asset returns	Working Capital to Total Assets	Fixed assets to equity
Percent return on investment	Working capital to total liabilities	Current Assets to Sale
Capital Working ratio	Profit to gross profit	Net profit on sale
Useful measure of loan	Operating profit to equity	Collection period
Sale to inventory	Accumulated earnings to total assets	Debt to equity ratio
Total liabilities to equity	Sale to receivable accounting	Cash adequacy ratio
Inventory to sale	Operation profit to interest expense	Liability ratio
Ratio of current assets	Gross profits to total debt	Cash ratio
Ratio of inventory to capital working	Capital Working to equity	Debt coverage ratio
The total financial cost of debt	Gross profit to total income	Working Capital
Debt cost to Gross Profit	Size	Debt to Capital*
Cash to total debt	Net Income to Total Debt	Quick ratio
Cash to total assets	Operating profit to total assets	Cash turnover ratio
Operating cash flow to debt	Total debt to equity	The financial burden of loans
Operating cash to equity	Equity to total assets	Inventory turnover
Operating cash flow to assets	Equity to capital*	Fixed assets turnover*
Operating cash to sale	Working capital to long-term debt	Accounts receivable to total debt
	Sales to Total Assets	Current liabilities to total assets

Figure 1. Algorithm Genetic



By comparing two models of Support Vector Machine and Genetic Algorithm and calculate error I and error II by following equitation of 2, 3 and 4:

$$Correct\ \ Rate = \frac{Number\ \ of\ \ True\ \ Prediction}{Total\ \ Number\ \ of\ \ Samples} \tag{2}$$

$$Error_1 = \frac{***}{***} \tag{3}$$

$$Error_2 = \frac{***}{***}$$

If correct rate closer to 100 and error I and error II closer to zero, prediction algorithm become closer to the reality. Error I is number of failure predictions of bankrupt companies and error II is number of number of failure prediction of non-bankrupt companies to total non-bankrupt companies and cost of these errors are different significantly. Cost of false classified as healthy company when company goes toward bankrupt (error I) is highly more than when a healthy company identified as failure company (error II). Thus, in bankruptcy models should be a balance between the types of forecast error. Results of the algorithms implemented in the first year, Year's t and t-1 show in Table 2.

Table 2. Results of the algorithms before deleting the outliers

year	algorithm	Accuracy of the prediction	Type I error	Type II error
T	SVM	73.1818	3.4567	52.7517
	GA	84.1502	12.5523	18.1617
t-1	SVM	74.18972	8.069597	42.73388
	GA	80.25692	17.44597	21.36908
t-2	SVM	67.0949	11.3225	54.4344
	GA	74.5059	27.8386	23.0155

Although in the first stage, the accuracy of the prediction of both models is not satisfying, but genetic algorithm compared to Support Vector Machine has higher accuracy of prediction and smaller type II error in three years t, t-1 and t-2. However, type I error is not consistent with these two results, so we compare the two models by using T-test (because the output of the algorithms are normal). The results are presented in Table 3.

Table 3. Results of the T-test in year t

	t	df	p-value
Error	-2.459	18	0.024
E_0	-2.622	18	0.017
E_1	4.703	18	0.000

Results of T-test in year's t and t-2 show that p-value of accuracy of prediction is less than .05 but type I error is not consistent with accuracy of prediction and type II error. In year's t-1 significance level of accuracy of prediction is more than 0.05 so the hypothesis is not significant. The results are presented in Tables 4 and 5.

Table 4. Results of T-test in year t-1

	t	df	p-value
Error	-1.179	18	0.254
E_0	2.653	18	0.016
E_1	-1.675	18	0.111

Table 5. Results of the T-test in year t-2

	t	df	p-value
Error	-2.196	18	0.041
E_0	5.999	18	0.000
E_1	-2.868	18	0.010

In second stage, Genetic Algorithm and Support Vector Machine models are designed and compared based on the 9 variables selected among 56 initial independent variables from the first stage. Increasing accuracy of prediction, decreasing errors of Support Vector Machine and Genetic Algorithm and aligning these criterions are observed in this stage compared to the first stage, which Table 6 shows the results of this stage.

Table 6. Results of the algorithms after deleting the outliers

year	algorithm	Accuracy of prediction	Type I error	Type II error
T	SVM	94.3083	4.21	7.3285
	GA	87.2727	14.1111	10.63869
T-1	SVM	91.2253	12.1	5.6153
	GA	84.20949	22.42208	9.4641
T-2	SVM	81.18577	24.6564	13.7
	GA	81.60079	28.88717	9.345599

In year’s t and t-1, accuracy of prediction of Support Vector Machine is more than genetic algorithm, and its type I and II errors are less. Considering the alignment of the results, we can conclude that in year’s t and t-1, Support Vector Machine model is more powerful than genetic algorithm. However, in year t-2 accuracy of prediction and type I error of genetic algorithm is higher. Tables 7, 8 and 9 demonstrate the results of this stage.

Table7. Results of T-test in year’s t

	T	df	p-value
Error	2.516	18	0.220
E_0	-3.211	18	0.005
E_1	-0.760	18	0.45

Table 8. Results of T-test in year t-1

	T	df	p-value
Error	1.709	18	**0.081**
E_0	-1.870	18	**0.078**
E_1	-0.906	18	**0.377**

Table 9. Results of T-test I year t-2

	T	df	p-value
Error	-0.127	18	0.900
E_0	-0.628	18	0.538
E_1	1.034	18	0.315

In year t absolute value of t statistic for accuracy of prediction and type I error is more than corresponding value in t-student (t=1.73) and according to aligning of type II error with accuracy of prediction and type I error and also this fact that significance level for all three criterions is less than .05, with the confidence of 95 % we can say that Support Vector Machine is more powerful than Genetic Algorithm. However, in years t-1 and t-2 significance level of accuracy of prediction is more than .05.

4. Conclusions

In this research, we try to compare the two models the Support Vector Machine And genetic algorithm. The compare did based on importance of financial ratios in bankruptcy prediction in two pre-election the Effective ratios and after the election was conducted and the results indicate that the in the first stage in year's t and t-2 Algorithm Genetic has the highest accuracy and lowest type II error. However, type I error is not aligned with the other criteria cannot make a decision which of them is better and it can be said as result of available ratios which cannot separate healthy and failure company. Although, in year's t-1 the genetic algorithm with the accuracy of 84.15%, higher than expected Support Vector Machine and significance level is greater than 0.5 the research hypothesis rejected. In the second stage of the selection - effective ratios and remove of financial ratios that were unable to separate the two groups leads to improve accuracy and decline errors significantly , moreover results of models become aligned and making decision about bankruptcy prediction models become easier. In fact, remove of ineffective observations in the second stage can improve accuracy of prediction with 94.30% and it is obvious that Support Vector Machine has better performance compare with genetic algorithm and the hypothesis is confirmed. Furthermore, our results are consisting of research by Kim and Kang (2012). Although, the hypothesis is rejected in year's t-1 and t-2 but evaluating criteria become aligned together and making it easier to choose a robust model.

Due to the high accuracy, prediction with Support Vector Machine is proposed that genetic algorithm or Support Vector Machine combine with other models like honeybee colonies and it is better to use of market price and compare it with the results of this research. Moreover, it is proposed that investigate, is there any difference between variable methods lead to the selection of the different independent variables.

5. References

Altman, E. I, (1968).Financial ratios, discriminate analysis and the prediction of corporate bankruptcy, the Journal of Finance, 23(4), 589-609.

Beaver,W, H.(1966), Financial ratios as predictors of failure . Journal of Accounting Research, 4, Empirical Research in Accounting: selected studies, (supplement), 71-111.

Blum, M. (1974). Failing Company Discriminant Analysis. Journal of Accounting Research, 12(1), 1-25.

Chen,M.Y.,(2011). Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. Computers and Mathematics with Applications, 62, 4514-4524.

Deakin, E. (1972). A Discriminant Analysis of predictors of Business Failure. Journal of Accounting Research, 10(1), 167-179.

Frydman, H., Altman, E., Kao, D. (1985). Introducing Recursive Partitioning for Financial Classification: the case of Financial Distress. Journal of Finance, 40(1), 269-291.

Gahlon,J.M.,Vigehand.R.L.,(1988).Early Warning of Bankruptcy Using flow Analysis. The Journal of commercial Bank Lending.pp:4-15.

Holand,J.A.,(1975),"Adaptation in natural and artificial systems", Ann Arbor, MI: university of Michigan press.

Hossari,G.,(2006). A Ratio-Based Multi-Level Modeling Approach for Signaling corporate collapse: A study of Australian corporations. Swinburne university of Technology. Australia, A thesis for the degree of Doctor of philosophy.

Kim, M.J., Kang, D. K., (2012).classifiers selection in ensembles using genetic algorithms for bankruptcy prediction. Expert systems with Applications,39,9308-9314.

Mensah, Y.M. (1983). The Differential Bankruptcy predictive Ability of specific prediction Models: A Methodological study. Journal of Accounting Research, 22(1), 380-395.

Min,Y, H., Lee, Y.C.(2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert systems with applications , 28, 603-614.

Pai, P, F., Lin, C, S (2005), A hybrid ARIMA and support vector machines model in stock price forecasting, the international journal of management science, vol. 33:497-505

Premachandra,I.M.,Bhabra,G.S.,Sueyoshi,T.,(2009), "DEA as A Tool For Bankruptcy Assessment: A comparative study with logistic Regression Technique" ,European Journal of operational Research 193,pp:412-424.

Ravi Kumar, P. Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques. European Journal of operational Research, 180, 1-28.

Trelea, I, C) 2003) the particle swarm optimization algorithm: convergence analysis and parameter selection, Information Processing Letters 85" 317–325

Vapnik V (1995), the nature of statistic learning theory. New York: Springer.

Wu, K, P., Wang, S, D (2012), Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space, Pattern Recognition, Vol. 42, pp. 710 – 717.

Yoon, J.S., Kwon, Y.S. (2010). A practical approach to bankruptcy prediction For small businesses: substituting the unavailable financial data for credit card sales information. Expert systems with Applications, 37, 3624-3629.