Volume 3, Issue 6, June 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Bankruptcy Prediction of Listed Corporations in Tehran Stock Exchange Using Data Mining Techniques

Zeinab rashidi [1]
[1]*Department of Public management, khorasgan (Isfahan) Branch, Islamic Azad University, Isfahan, Iran.*

Sayed mohsen hosseini[2]
[2]*Department of Biostatistics at Medical University of Isfahan, Isfahan, Iran*

Abstract— Aims: This study aimed at predicting bankruptcy based on two data mining techniques, i.e. logistic regression and classification and regression tree (CART). Study design: This was an applied, descriptive- analytical, cross-sectional study. Place and Duration of Study: This research was carried out in Iran. Annual financial statements of companies in Tehran stock market (Iran) during 1999-2010 were evaluated. Methodology: No sampling was performed. In total, 98 successful companies and 71 bankrupt companies were included. In designing the models, financial ratios were considered as independent variables and successful and bankrupt companies were regarded as dependent variables. Results: According to the CART, the most important variables in predicting bankruptcy were return on assets, debt ratio, operating profit to total assets ratio, working capital to total assets ratio, and operating profit margin. The logistic regression model suggested return on assets, debt ratio, and operating profit to total assets ratio as predicting variables. The overall accuracy of prediction using the CART was 100% on training data and 92.8% on test data. The accuracy of regression model was 95.9%. Conclusion: Both models revealed return on assets, debt ratio, and operating profit to total assets ratio to be the most important variables in predicting bankruptcy. However, considering the area under the receiver operating characteristic (ROC) curve, the logistic regression model had better performance in predicting bankruptcy.

Keywords— Prediction of bankruptcy, Data mining, Decision tree, Classification and regression tree, Logistic regression.

I. INTRODUCTION

Bankruptcy prediction seems to be most popular topic of the application of DM techniques on financial data. Corporate bankruptcy causes economic damages for management, investors, creditors and employees together along with social cost. For these reasons bankruptcy prediction is an important issue in finance. Increasing competitiveness among enterprises reduces the availability of resources and augments the risk of bankruptcy. Research has shown that companies tend to postpone the official announcement of their bankruptcy until it is too late [48]. Therefore, a crucial problem is whether it is possible to predict the time of a company's bankruptcy before it actually occurs. Bankruptcy of companies generally affects the liquidity of the capital market and the economic development. Banks incline to stop giving credit to bankrupt companies or to ask for higher loan interest to compensate for the increased risk. Likewise, investment companies, e.g. insurance companies and pension funds, shun buying shares in the stock market and prefer to buy bank bonds or invest in similar markets. All these factors will result in reduced liquidity in the capital market, higher cost of capital, and lower economic development [44]. Investors and creditors have a great propensity for prediction of corporate bankruptcy since it imposes enormous costs on them. Accordingly, bankruptcy prediction has turned into a major issue in financial literature during the past for decades. Numerous scientific studies have sought to identify best bankruptcy prediction models based on available data and statistical methods. Researchers in most developing countries have also made huge efforts to design and introduce new prediction models based on different economic and financial circumstances [25].

Currently, the Data Mining or Knowledge Discovery(KDD) technologies ara rapidly developed to discover useful paterns in great amount of data, especially in financial domain problems. In some early related works, Beaver [7], Altman [1] and Ohlson [45] are the pioneers of the financial distress empirical approach. Beaver, in particular, was one of the first researchers to study the prediction of bankruptcy using financial statement data. However, his analysis is very simple in that it is based on studying one financial ratio at a time and on developing a cut-off threshold for each ratio. The approaches by Altman and Ohlson are essentially linear models that classify between distress and non-distress firms using financial ratios as inputs. Altman used the classical multivariate discriminate analysis technique (MDA). Both the MDA model and the linear regression model (LR) have been widely used in practice and in many academic studies. They have been standard benchmarks for the financial distress prediction problem. The main criticism of the MDA is that the restrictive statistical requirement posed by the model [15]. Binary models such as Probit [59],Logit [45, 11], were able to overcome the main problems of MDA. In traditional statistical analysis like Logit model, the modeler was required to specify the precise relationship between inputs and outputs and any restrictions that might be implied by theory. Whereas artificial neural network (ANN) is not required to specify the nature of the relationships involved. Research studies on ANN for bankruptcy prediction started in 1990 [8, 46, 52, 13, 53, 18, 2, 56, 9, 10, 16, 57, 60, 49, 3, 4, 14, 21, 44] used to forecast financial distress for bank and other business. The true power and advantage of ANN lies in their ability to represent both linear and non-linear relationships and in their ability to learn these relationships directly from the data being modeled. Often remarked upon as a major drawback of ANN is the fact that their internal functional structure remains unknown once they have been trained. In effect, an ANN remains a "black box" that may produce useful results, but cannot be precisely understood. That is to say, it is difficult or impossible to explain how decisions were made [31]. Other methodologies applied to this problem include genetic algorithms [54, 50] and support vector machines [19, 20], decision tree [30, 51, 22, 31, 35, 17, 23], hybrid intelligent systems [29, 32, 5, 33, 34, 12, 41, 36, 58, 37, 38, 39, 40, 41, 42, 43]. the decision tree has become a very popular data mining technique and commonly used for classification. The advantages of decision tree are simple to understand and interpret, require little data preparation, able to handle nominal and categorical data, perform well with large data in a short time and the explanation for the condition is easily explained by boolean logic. However, it ignores the relationship between attributes, output attribute must be categorical. What is more, the decision tree algorithms are unstable, trees created from numeric datasets can be complex [31].

Among the Iranian studies on bankruptcy prediction models, none has focused on classification and regression trees (CART). A CART is nonparametric, stable, precise, and easy to implement. It also has simple results. The present study predicted bankruptcy with two data mining techniques, i.e. CART and logistic regression. It tried to provide a desirable model to warn company managers, government officials, shareholders, investors, creditors, financial analysts, and employees before bankruptcy. The significance of this research was the use of the mentioned methods to efficiently predict bankruptcy in new samples.

The rest of the paper is organized as follows: the next section discusses the methodology adopted for the study followed by the results and discussion section. The conclusion section ends the paper.

II. MATERIAL

II-A. Statistical Population and Sample Selection

The study population comprised companies which had provided the Tehran stock market with their annual financial statements during 1999-2010. No sampling was performed and all eligible companies were studied. Companies with changes in financial year during the period of the study and investment companies, financial intermediaries, holding or leasing companies, and banks were not included. Other companies were only included if their financial statements were available. Based on data from one year before bankruptcy, the companies were categorized as either bankrupt or successful. The last year a company is subject to article 141 of Iran's commercial code is considered as the year before bankruptcy. In order to confirm the success of a company during a particular year, its profitability during that same year is calculated. This criterion has been selected due to the relation between earnings and retrained earnings as the profit of a company at the end of a financial year is transferred to its accumulated profit and loss account. Article 141 uses the ratio of accumulated profit and loss to a company's capital (accumulated loss > 50% of the capital) to classify companies. A census was taken and all eligible companies (98 successful and 71 bankrupt companies) were included.

II-B. Employed Hypotheses

According to theoretical principles of research, we considered the following hypotheses:

i. The designed CART is capable of predicting the bankruptcy of companies in Tehran stock market.
ii. The designed logistic regression model is capable of predicting bankruptcy of companies in Tehran stock market.

II-c. Variables:

In this research, successful and bankrupt companies were assumed as dependent variables and weighted as 1 and 0, respectively. Independent variables were the best financial ratios suggested by previous studies (Table 1).

TABLE 1. independent variable description

variables	**Description**
CA/CL	Ratio Current (Current Assets / Current Liabilities)
WC/TA	Working Capital / Total Assets
TD/TA	Debt Ratio (Total Liabilities / Total Assets)
NS/TA	Assets Turnover Ratio (Net Sales / Total Assets)
CA/TA	Current Assets / Total Assets
EBIT/TA	Operating Profit / Total Assets
EBIT/NS	Margin of Operating Profit (Operating Profit / Net Sales)
ROA	Return on Total Assets (Net Profit/ Total Assets)
ROE	Return on Equity (Net Profit/Equity)
FS	Size Corporation (Logarithm Net Sales)

III. METHODS

This study used two data mining methods, i.e. logistic regression and CART, which are explained in the following sections:

III-A. Decision Trees

Decision trees use very simple techniques to provide a non-parametric classification model for available observations. While no complicated technique is employed and the resulting model is easy to understand, decision trees can predict as complex problems as can neural networks.[26] A decision tree is a powerful yet uncomplicated method to categorize data to distinct, homogeneous classes in a tree-shaped diagram. The tree is formed based on a set of questions. Each question usually involves one variable. A decision tree comprises three parts: a root, internal nodes, and external nodes (leaves). In order to construct a decision tree, a covariate is considered as the root and split into a number of internal nodes based on a set of questions and features. Several algorithms of decision trees (e.g. Quest, Chaid, C5.0, and CART) have been developed.[6] In the current study, we used CART as a comparative algorithm.

Breiman et al. suggested CART as a type of classification tree in 1984. A CART is an acyclic graph which provides a classification model through binary splits based on different covariates. Similar to any other decision tree, a CART comprises a root and internal and external nodes. A covariate is initially considered as the root and then split to a number of internal nodes according to study objectives. Every internal node will be further split and this process continues until each node represents a class of the response variable. These final nodes are called external nodes or leaves. In order to identify important variables, CART uses the impurity function and Gini index. In a node t, impurity function for a covariate with k classes ($c_1, c_2, \ldots, c_k$) is defined as:

$$i(t) = \Phi\left[P(C = C_1 \mid t), \ldots p(C = C_k \mid t)\right]$$

In most decision models, Gini index is used with binary splits at each node. The index is defined as:[6]

$$i(t) = \text{gini}\ (t) = 1 - \sum_{j=1}^{k} p^2\left[c = c_j \mid t\right]$$
$$= \sum_{k \neq 1} P(c = c_k \mid t) P(c = c_1 \mid t)$$

The above equation will equal zero when all observations belong to the same class. On the other hand, it will maximize when all classes have equivalent possibility. Considering x as the covariate based on which the node t is split to n branches (T_j, j= 1, 2, …, n), a reduction will be observed in the impurity function. This value is calculated based on the Gini index as follows:

$$\text{gini gain} = GG(T, X)$$
$$= \text{gini}\ (t) \quad \sum_{j=1}^{n} p(t_j \mid t)\ .\ \text{gini}\ (t_j)$$

Among a number of variables, a variable with maximum GG (T, X) is the best one. Therefore, the impurity function is first calculated for the response variable. Afterward, for all covariates, the impurity function is calculated for both splits (the best splits for the response variable). Their mean weight is then subtracted from the total impurity function. The best covariate for the first-level splitting is the one resulting in the highest value. In handling of quantitative variables, binary splits are made and a cut-off point (like A) is determined. In many classification trees, the cut-off point is actually determined by an index (in this case the Gini index). While working with qualitative variables, each level will form a branch of the classification tree. [24]

In order to clarify the appropriate size of a CART, complexity cost is calculated. A favorable tree will fit not only the learning data set (available observations) but also the test data set (new observations). Although a larger tree will provide more accurate classification on the learning data set, it will reduce precision in test data sets bigger than a specific size. The mentioned method will hence balance the size and accuracy of the classification tree, i.e. it decides which nodes are supposed to be pruned based on the number of nodes and Length of branches.

III-B. Logistic Regression Model

In many studies, the dependent variable is not continuous and may have to results. For instance, it may take either one or zero to represent the presence or absence of an event (or vice versa). We might be willing to determine the success or failure of a person in the university entrance exam based on his/her level of efforts and intelligence. We may also be interested in evaluating financial distress based on a number of variables. While logistic regression is similar to ordinary regression, the two methods differ in estimation of coefficients, i.e. instead of minimizing the squared errors (as performed in ordinary regression), logistic regression maximized the possibility of an event.[47]

IV. RESULTS and DISCUSSION

IV-A. Corporate Bankruptcy Prediction Using the CART In this method, a decision tree is constructed to classify the data. Each observation is placed in the most closely related class, i.e. the goal is to make observations in each class have the highest similarity. On the other hand, the variables are considered as splitters instead of connectors. In other words, a node in the decision tree is split based on the significance of each variable in formation of the split nodes. The following section discusses the implementation of the CART to predict corporate bankruptcy. We used SPSS Clementine (SPSS Inc., Chicago, IL, USA) to test the assumed hypotheses. In order to predict bankruptcy, the available data set was divided into two equal subsets of learning and test data. Then, 10 variables were entered in the model. Table 2 summarizes the importance of variables in splitting of the companies in the tow data sets. The most important variables were return on assets ratio, debt ratio, operating profit to total assets ratio, working capital to total assets ratio, and operating profit margin.

TABLE 2. independent variable importance

Independent Variable	Importance	Normalized Importance
ROA	.449	100.0%
TD/TA	.407	90.6%
EBIT/TA	.332	73.9%
WC/TA	.309	68.9%
EBIT/NS	.294	65.5%
CA/CL	.277	61.7%
ROE	.100	22.2%
FS	.078	17.4%
NS/TA	.020	4.5%
CA/TA	.009	2.1%

Growing Method: CART

The CART constructed based on the learning data set is demonstrated in Diagram 1. As it is seen, return on assets ratio had the highest significance in predicting corporate status. For instance, in the class with return on assets ratio ≤ 0.015, 2.7% of corporates in the learning data set were successful and 97.3% were bankrupt. In the next stage of classification, corporate size plays the most important role. The accuracy of prediction in this class is 100%. On the other hand, in the classes with return on assets ratio > 0.015, debt ratio had the highest importance in the next level of classification. In the class with debt ratio ≤ 0.960, return on assets was the most significant variable in the next level of classification and the accuracy of prediction in this class was again 100%. In the lass with debt ratio > 0.960, the accuracy of prediction is 100% and the classification terminates.

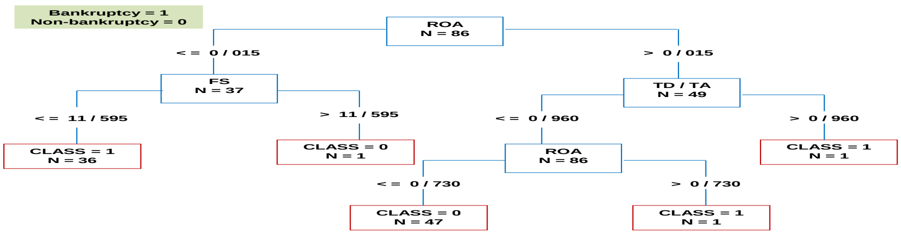


Diagram 1. The decision tree constructed from the learning data set based on the study variables

Table 3 provides full information about the construction of the CART from both learning and test data sets. The total number of data items was 86 corporates (48 successful and 38 bankrupt companies) in the training phase and 83 corporates (50 successful and 33 bankrupt corporates) in the test phase. The CART contains eight nodes. The table also presents the accuracy of prediction at each node, the value of each branch, and the significance of variables in classification.

TABLE 3. decision tree

Sample	Node	0		1		Total		Predicted Category	Parent Node	Primary Independent Variable		
		N	Percent	N	Percent	N	Percent			Variable	Improvement	Split Values
Training	0	48	55.8	38	44.2	86	100	0				
	1	1	2.7	36	97.3	37	43	1	0	ROA	.426	<= .015
	2	47	95.9	2	4.1	49	57	0	0	ROA	.426	> .015
	3	0	.0	36	100	36	41.9	1	1	FS	.023	<= 11.595
	4	1	100	0	.0	1	1.2	0	1	FS	.023	> 11.595
	5	47	97.9	1	2.1	48	55.8	0	2	TD/TA	.022	<= .960
	6	0	.0	1	100	1	1.2	1	2	TD/TA	.022	> .960

	7	47	100	0	.0	47	54.7	0	5	ROA	.023	<= .730
	8	0	.0	1	100	1	1.2	1	5	ROA	.023	> .730
Test	0	50	60.2	33	39.8	83	100	0				
	1	4	11.4	31	88.6	35	42.2	1	0	ROA	.426	<= .015
	2	46	95.8	2	4.2	48	57.8	0	0	ROA	.426	> .015
	3	4	11.8	30	88.2	34	41	1	1	FS	.023	<= 11.595
	4	0	.0	1	100	1	1.2	0	1	FS	.023	> 11.595
	5	46	97.9	1	2.1	47	56.6	0	2	TD/TA	.022	<= .960
	6	0	.0	1	100	1	1.2	1	2	TD/TA	.022	> .960
	7	46	97.9	1	2.1	47	56.6	0	5	ROA	.023	<= .730
	8	0	.0	0	.0	0	.0	1	5	ROA	.023	> .730

Growing Method: CART

Table 4 includes prediction results and the final CART model.

TABLE 4. the accuracy of prediction using the classification and regression tree (cart) based on the findings of the study

		Prediction		
	Actual	**Non-Bankrupt**	**Bankrupt**	**Accuracy of Prediction (%)**
Training Sample	**Non-Bankrupt**	48	0	100
	Bankrupt	0	38	100
	Total Accuracy(%)			100
Test Sample	**Non-Bankrupt**	46	4	92
	Bankrupt	2	31	93/9
	Total Accuracy(%)			92/8

IV-B. Corporate Bankruptcy Prediction Using Logistic Regression

SPSS for Windows 20.0 (SPSS Inc., Chicago, IL, USA) was used to build the logistic regression model. Ten ratios were simultaneously entered in the model. Chi-square test was performed to evaluate the overall significance of the model. The appropriate step for the model was calculated as twice the log-likelihood (-2LL). As the significance level of the chi-square test (0.001) was lower than the error rate (0.05), the model has been statistically significant (Table 5). On the other hand, the optimal step is the one minimizing -2LL.

TABLE 5. the results of chi-square test regarding the overall statistical fitting and significance of the logistic regression model

	chi-square	**df**	**sig**	**-2 log likelihood**
Logistic Regression Model	207/818	3	0/001	22/133

After the significance of the whole model had been confirmed, the Wald test was applied to assess the significance of independent variables, i.e. all variables were entered in the model but non-significant ones were excluded and only significant variables finally remained in the model. The remaining independent variables were statistically significant at 0.05 level and significantly related with dependent variables. The most influential factors were operating profit to total assets ratio, debt ratio, and return on assets ratio. The operating profit to total assets ratio and return on assets ratio were negatively related with bankruptcy. In contrast, debt ratio had a positive relation with bankruptcy. Therefore, higher debts and lower operating profit to total assets ratio and return on assets ratio can all increase the probability of bankruptcy (Table 6).

TABLE 6. the wald test results and the obtained coefficients

Independent Variables	**B**	**S.E**	**Test Wald**	**df**	**Sig**
Ebit/Ta	-74/962	24/219	9/580	1	0/002
Debt Ratio	26/609	9/285	8/213	1	0/004
Roa	-9/549	4/505	4/493	1	0/034
Constant	-18/539	6/555	7/998	1	0/005

The next stage in this study was to test the assumed hypotheses through evaluating the designed corporate bankruptcy prediction model. Table 7 shows the logistic regression model's accuracy of prediction. As the table shows, among the 98 successful companies, the model predicted 94 to be successful and 4 to be bankrupt (accuracy of prediction = 95.9%). On the other hand, among the 71 bankrupt corporates, the prediction was correct about 68 (accuracy of prediction = 95.8%). Therefore, the model's overall accuracy of prediction was 95.9%.

TABLE 7. accuracy of prediction in the logistic regression model

Actual	Prediction		
	Non- Bankrupt	Bankrupt	Accuracy of Prediction (%)
Non- Bankrupt	94	4	95/9
Bankrupt	3	68	95/8
Total Accuracy(%)			95/9

IV-C. Testing the hypotheses :

This study employed two data mining methods (CART and logistic regression) to predict corporate bankruptcy. Accuracy of classification is crucial in data mining models. Table 8 presents the accuracy of the two models based on the study findings. According to our findings, the hypothesis implying the inability of the CART and logistic regression models to predict is rejected. As a result, the first and second hypotheses suggesting the capability of the two models to appropriately predict. Accuracy of prediction in the CART model was 100% for both successful and bankrupt companies using the learning data set and 92.0% for successful companies and 93.9% for bankrupt companies using the test data set. The accuracy of prediction in the logistic regression model was 95.9% and 95.8% for successful and bankrupt companies, respectively.

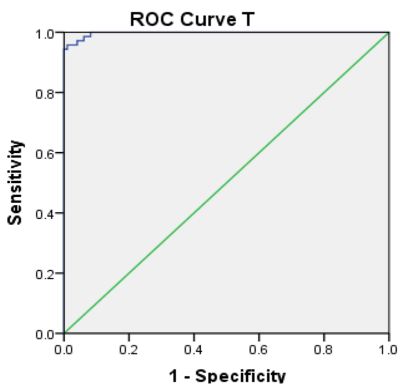
TABLE 8. comparison between the accuracy of prediction in the classification and regression tree (cart) and logistic regression models

Model		Accuracy Of Prediction (%)	
		Non-Bankrupt	**Bankrupt**
CART	Training Sample	100	100
	Test Sample	92	93/9
Logistic Regression		95/9	95/8

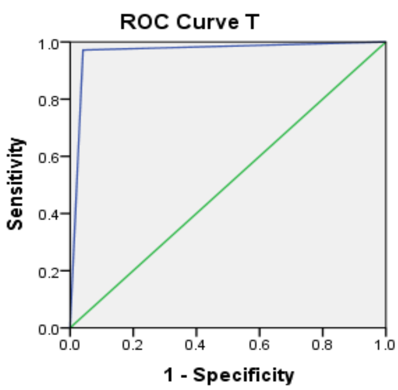
Modelling studies use the area under receiver operator characteristic (ROC) curve to assess the accuracy of the prediction made by a model. The area under the ROC curve quantifies a model's ability to discriminate between two results (discriminatory power). The closer the area under the ROC curve is, the model is more precise. However, values close to 0.5 indicate low accuracy and unreliable prediction. The two models in the current study were compared through determining the area under the ROC curve based on data mining methodologies. According to the calculated values, the logistic regression model had better performance in predicting bankruptcy (Table 9).

TABLE 9. area under receiver operator characteristic (roc) curve obtained from the study results

Logistic Regression	CART
0/997	0/986



Logistic Regression



CART

V. Conclusion

Based on our findings, data from financial statements has a high predictive power. Both logistic regression and CART models suggested return on assets ratio, debt ratio, and operating profit to total assets as the most important variables in prediction of bankruptcy. We found the logistic regression model to have a more favorable area under the ROC curve which indicated its better performance in predicting bankruptcy. In contrast, a previous research [31, 35] reported the

CART to have higher area under the ROC curve compared to the logistic regression model. Economic fluctuations and political variables are the most significant external, uncontrollable factors (on the part of companies) leading to corporate bankruptcy in Iran. On the other hand, high production costs, payable interest expense, and production bureaucracy are the key the internal, controllable factors causing bankruptcy in the country [28].

V-A. Study Limitations

Reaching a goal is always decelerated due to various limitations. As a process to solve the research problem, research is no exception. We hereby inform the readers to generalize the results obtained by this study with more caution and to have a fair judgment about the process of research. The limitations of the present study are as follows:

a) As mentioned before, this research used data from companies currently registered in Tehran stock market. Hence, generalizing its findings to other time periods or non-stock companies has to be made more cautiously.

b) Several factors can undoubtedly help in determining a company's financial distress. However, difficulties in quantifying some variables and lack of adequate data about certain variables forced us to include the effects of a limited number of variables. Therefore, the effects of other factors cannot be underestimated.

c) Data from some bankrupt companies was not accessible.

d) The annual number of bankrupt companies is limited.

V-B. Practical Suggestions Based on Study Results

According to our findings and the significance of the studied subject, the following practical suggestions are presented:

a) The Securities and Exchange Organization of Tehran can also use the models designed in this study to gain insight into the future financial state of companies before they are accepted in the exchange organization.

b) Banks and financial and credit institutions can use the obtained models before granting high amount loans to industries.

c) The designed models can also be used by stock brokers, analysts, and financial consultants who are responsible for analyzing the companies' financial state in the stock market and describing their future status for potential stock buyers.

d) Company managers are constantly searching for ways to evaluate their weak points and to predict future threats. Using models based on financial data is a method to evaluate future financial weaknesses and the resulting financial distress. Therefore, the models presented in the current study can be appropriate options for the managers to detect financial distress.

e) Investment companies hunt for apt opportunities for investment. They make various evaluations to find the most profitable industry. As they require to assess corporate bankruptcy, the models designed in this study can be beneficial to them.

f) Industries may decide to merge with or buy other companies in order for product development and raw material control. They have to appraise the financial state of available companies to prevent future financial problems. Such industries can use the models constructed in the current study to predict other firms' future financial status.

References

[1] E. I. Altman, "Financial ratios, discriminate analysis and the prediction of corporate bankruptcy," Journal of Finance., vol. 23(4), pp. 578-609. 1968.

[2] E. I. Altman, G. Marco, F. Varetto, "Corporate distress diagnosis: comparisons using linear discriminate analysis and neural networks," Journal of Banking and Finance., vol. 18, pp. 505-529. 1994.

[3] A. F. Atiya, "Prediction for credit risk using neural networks: a survey and new results," IEEE Tractions on Neural Networks., vol. 12(4), pp. 929-935. 2001.

[4] E. Alforo, N. Garcia, "Bankruptcy forecassting: an empirical camparison of ad boost and neural networks,"International Advances in Economic Research., pp.110-122 .2008.

[5] H. Ahn, K. J. Kim, "Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach,"Applied Soft Computing., vol. 9(2), pp. 599-607. 2009.

[6] A. Azar, P. Ahmadi, M. V Baste, "Design human resource selection model using data mining approach,"Journal of Information and Technology Management., vol. 4, pp. 3-22. 2010. Iran.

[7] W. H Beaver, "Financial ratios as predictors of failure,"Journal of Accounting Research., vol. 4, pp. 71-111.1966.

[8] T. Bell, G. Ribar, J. Verchio, "Neural nets versus logistic regression: a comparison of each model's ability to predict commercial bank failures, in proceeding of the 1990 deloitte touche/univ,"Kansas Symp. On Auditing Problems, pp. 29-53. 1990.

[9] J. E. Boritz, D. B. Kennedy, "Effectiveness of neural network types for prediction of business failure,"Expert Systems with Applications., vol. 9(4), pp. 95-112. 1995.

[10] B. Back, T. Laitinen, K. Sere, "Neural networks and genetic algorithms for bankruptcy predictions,"Expert System Application., vol. 11(4), pp. 407-413. 1996.

[11] B. Back, T. Laitinen, K. Sere, M. van Wezel, "Choosing Bankruptcy Predictors using Discriminant Analysis, Logit Analysis and Genetic Algorithms,"Technical Report no. 40, Turku Centre for Computer Science, Turku School of Economics and Business Administration. 1996.

[12] A. Bahrammirzaee, "A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems,"Neural Computing & Applications., vol. 19(8), pp. 1165-1195. 2010.

[13] P. Coats, L. Fant, "A neural network approach to forecasting financial distress,"Journal of Business Forecasting., vol. 10(4), pp. 9-12. 1992.

[14] W. S. Chen, Y. K. Du, "Using neural networks and data mining techniques for the financial distress prediction model,"Expert Systems with Applications., vol. 36(2), pp. 4075-4086. 2009. doi: http://dx.doi.org/10.1016/j.eswa.2008.03.020

[15] E. B. Deakin, "A discriminant analysis of predictors of Business failure,"Journal of Accounting Research., vol. 10, pp. 167-179. 1972.

[16] H. L. Etheridge, R. S. Sriram, "A comparison of the relative costs of financial distress models: artificial neural networks, logit and multivariate discriminate analysis,"Intelligent Systems in Accounting, Finance, and Management., vol. 6, pp. 235-248. 1997.

[17] İ. H. Ekşi, "Classification of firm failure with classification and regression trees,"International Research Journal of Finance and Economics., vol. 76, pp. 113-120. 2011.

[18] D. Fletcher, E. Goss, "Forecasting with neural networks an application using bankruptcy data,"Information Management., vol. 24, pp. 59-167. 1993.

[19] V. Fan, v. Palaniswami, "Selecting bankruptcy bredictors using a support vector machine approach,"Proceedings of the Internal Joint Conference on Neural Networks, 2000, p. 354-359.

[20] S. Falahpour, R. Raie, "Application of support vector machine to predict financial distress using financial ratios,"Journal of Accounting and Auditing Studies., vol. 53, pp. 17-34. 2005.Iran.

[21] S. Falahpour, R. Raie, "Predicting financial distress of corporations by using artificial neural networks,"Financial Research., vol. 17, pp. 39-69. 2009.Iran.

[22] A. Gepp, K. Kumar, S. Bhattacharya, "Business failure prediction using decision trees,"Journal of Forecasting., vol. 29(6), pp. 536-555. 2009.

[23] A. Gepp, J. H. Wilson, K. Kumar, S. Bhattacharya, "A comparative analysis of decision trees vis-a-vis other computational data mining techniques in automotive insurance fraud detection,"Journal of Data Science., vol. 10(3), pp. 537-561. 2012.

[24] M. Hossini, M. Tazhibi, M. Amini, A. zare, H. Jahanihashemi, "Application of classification trees to predict diabetic rtyvpany comparison with diagnostic function model in patients with type 2 diabetes,"Journal of Isfahan Medical School., vol. 104, pp. 17-27. 2010. Iran.

[25] F. L Jones, "Current techniques in bankruptcy prediction,"Journal of Accounting Literature., vol. 6, pp. 131-164. 1978.

[26] H. A. Kiers, J. P. Rasson, P. J. Groenen, "Data analysis, Classification, and Related Methods,"IFCS, University of Namur. Belgium : Springer. 2000.

[27] P. R. Kumar, V. Ravi, "Bankruptcy prediction in banks and firms via statistical and intelligent techniques – a review, Eur,"J. Oper. Res., vol. 180, pp. 1-28. 2007.

[28] A. Kamijani, J. Saadatfar, "Conditional probability of the optimal model for forecasting economic bankruptcy of companies in iran," Monthly Benefit., vol. 57 .2007. Available:www.magiran.com. Iran.

[29] K. C. Lee, I. Han, Y. Kwon, "Hybrid neural network models for bankruptcy predictions," Decision Support Systems., vol. 18(1), pp. 63-72. 1996.

[30] W. C. Lee, "Genetic programming decision tree for bankruptcy prediction," In Proceedings of the 2006 Joint Conference on Information Sciences(JCIS), 2006, p. 8-11.

[31] w. c. Lee. An empirical comparison of bankruptcy models -evidence from Taiwan. 149-168. Available: http: //ir.lib.au.edu.tw/bitstream/987654321/812/1/CM05-pa_lee_01-0001-u.

[32] P. C. Lin, J. S. Chen, "A genetic-based hybrid approach to corporate failure prediction," International Journal of Electronic Finance., vol. 2(2), pp. 241-255. 2008.

[33] H. Li, J. Sun, "Forecasting business failure in China using case-based reasoning with hybrid case respresentation," Journal of Forecasting., vol. 29(5), pp. 486-501. 2009.

[34] H. Li, J. Sun, "Predicting business failure using multiple case-based reasoning combined with support vector machine," Expert Systems with Applications., vol. 36(6), pp. 10085-10096. 2009.

[35] H. Li, J. Sun, J. Wu, "Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods," Expert Systems with Applications., vol. 37(8), pp. 5895-5904. 2010.

[36] H. Li, Y. C. Lee, Y. C. Zhou, J. Sun, "The random subspace binary logit (RSBL) model for bankruptcy prediction," Knowledge-Based Systems., vol. 24(8), pp. 1380-1388. 2011.

[37] H. Li, J. Sun, "Empirical research of hybridizing principal component analysis with multivariate discriminant analysis and logistic regression for business failure prediction," Expert Systems with Applications., vol. 38(5), pp. 6244-6253. 2011.

[38] H. Li, J. Sun, "Predicting business failure using an RSF-based case-based reasoning ensemble forecasting method," Journal of Forecasting. 2011.

[39] H. Li, J. Sun, "Predicting business failure using forward ranking-order case-based reasoning," Expert Systems with Applications., vol. 38(4), pp. 3075-3084. 2011.

[40] H. Li, J. Sun, "Predicting business failure using support vector machines with straightforward wrapper: A re-sampling study," Expert Systems with Applications., vol. 38(10), pp. 12747-12756. 2011.

[41] H. Li, J. Sun, "Mining business failure predictive knowledge using two-step clustering," African Journal of Business Management., vol. 5(11), pp. 4107-4120. 2011. DOI: 10.5897/AJBM10.158. Available: http://www.academicjournals.org/AJBM.

[42] F. Lin, C. C. Yeh, M. Y. Lee, "The use of hybrid manifold learning and support vector machines in the prediction of business failure," Knowledge-Based Systems., vol. 24(1), pp. 95-101. 2011.

[43] H. Li, J. Sun, "Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples–Evidence from the Chinese hotel industry," Tourism Management., vol. 33(3), pp. 622-634. 2012.

[44] M. R. Nikbakht, M. Sharifi, "Financial bankruptcy prediction of corporations in stock exchanges by using artificial neural network," Journal of Industrial Management., vol. 2(4), pp. 163-180. 2011.

[45] J. A. Ohlson, "Financial ratios and the probabilistic prediction of bankruptcy," Journal of Accounting Research., vol. 18(1), pp. 109-131. 1980.

[46] M. D.Odom, R. Sharda, "A neural network model for bankruptcy prediction," IJC NN International Joint Conference on Nearal Networks, 1990, 2, p. 163-168. Available from: http://ieeexplore.ieee.org.[Accessed 28 May 2005].

[47] R. Pirayesh, A. Mansori, S. Amjadiyan, "Design cash flow-based mathematical models to predict bankruptcy companies accepted in Tehran Stock," Journal of Capital Development., vol. 4, pp. 73-94. 2009. Iran.

[48] P. Ravi Kumar, V. Ravi, "Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review," European Journal of Operational Research., vol. 180(1), pp. 1-28. 2007.

[49] J. R. Shah, M. B. Murtza, "A network based clustering procedure for bankuptcy prediction," American Business Review., vol. 18(20), pp. 80-86. 2000.

[50] K. Shin, Y. Lee, "A genetic algorithm application in bankruptcy prediction modeling," Expert Systems with Application., vol. 23(3), pp. 321-328. 2002.

[51] J. Sun, H. Li, "Data mining method for listed companies' financial distress prediction," Knowledge-Based Systems., vol. 21(1), pp. 1-5. 2008.

[52] K. Y. Tam, M. Kiang, "Managerial applications of the neural networks: the case of bank failure predictions," Management Science., vol. 38(7), pp. 926-947. 1992.

[53] G. Udo, "Neural network performance on the bankruptcy classification problem," Computers and Industrial Engineering., vol. 25, pp. 377-380. 1993.

[54] F. Varetto, "Genetic algorithms in the analysis of insolvency risk," Journal of Banking and Finance., vol. 22(10), pp. 1421-1439. 1998.

[55] A. Verikas, Z. Kalsyte, M. Bacauskiene, A. Gelzinis, "Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey," Soft Computing-A Fusion of Foundations, Methodologies and Applications., vol. 14(9), pp. 995-1010. 2010.

[56] R. L. Wilson, R. Sharda, "Bankruptcy prediction using neural networks," Decision Support Systems., vol. 11(5), pp. 545-557. 1994.

[57] Z. Yang, M. Platt, H. Platt, "Probabilistic neural networks in bankruptcy prediction," Journal of Business Research., vol. 44, pp. 67-74. 1999.

[58] C. C.Yeh, D. J. Chi, M. F. Hsu, "A hybrid approach of DEA, rough set and support vector machines for business failure prediction," Expert Systems with Applications., vol. 37(2), pp. 1535-1541. 2010.

[59] M. E. Zmijewski, "Methodological issues related to the estimated of financial distress prediction models," Journal of Accounting Research., vol. 22(1), pp. 59-82. 1984.

[60] G. Y. Zhang, M. Hu, B. C. Eddy Patuwo, D. Indro, "Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis," European Journal of Operational Research., vol. 116(1), pp. 16-32. 1999.

