International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-2, Issue-9)

Research Article **September 2013**

Using Decision Tree Model and Logistic Regression to Predict Companies Financial Bankruptcy in Tehran Stock Exchanges

Hossein Rezaie Doolatabadi * , Seyed Mohsen Hoseini, Rasoul Tahmasebi

Management Department at the University of Isfahan   Bio-Statistics Department at the Medical University of Isfahan

MA in Accountant at the Islamic Azad University UAE branch

Abstract—

Predicting financial bankruptcy is always noticed by shareholders, creditors and managers of commercial units. Therefore, it is more than a half of century that economists, educational centers, and the professional people are attempting to provide optimized models to predict financial bankruptcy. Consequently, various models have been suggested in order to predict bankruptcy. The present study attempts to use decision tree model and logistic regression in order to predict financial bankruptcy of the companies and to compare these two models with each other. For this, the prediction of financial bankruptcy, which is the main purpose of the study, has been analyzed by using decision tree chart. To conduct the research, three major hypotheses were formulated. For testing the hypotheses, a sample has been taken among the accepted companies in Tehran stock exchange during 2006 to 2011. Factor analysis, decision tree, and logistic regression were used to test the hypotheses. Two models of logistic regression and decision tree were also compared with each other.

Keywords— bankruptcy, financial ratios, factor analysis, logistic regression, decision tree

I. Introduction

The problem of bankruptcy in companies has been always complicated and intricate. For this reason, financial and accounting scholars around the world try to find some methods to predict financial bankruptcy in companies. Undesirable financial conditions of companies cause damages to various strata of the society, especially investors.

Proposing a precise definition about the involved groups in bankruptcy matter is very difficult. But it could be claimed that management, investors, creditors and legal institutes are more influenced by the phenomenon of bankruptcy than other related categories. Therefore, nowadays, bankruptcy prediction is highly underlined. By predicting the bankruptcy, investors not only prevent the danger of their investment loss, but also use it as a tool to reduce the danger of their investment basket. Managers of commercial units can also imply preventive diplomacies to prevent bankruptcy if they are informed of bankruptcy dangers in the right time. Since bankruptcy imposes expensive social and economical costs on the society, it will also be noticed from a macro-view, because waste resources of a critical economic unit could be allocated to other profitable opportunities. Accordingly, scholars tried to predict companies' bankruptcy by analyzing financial forms and financial ratios. A great deal of studies have been done in Iran and other countries some of which have proposed models for predicting bankruptcy and some of these models can predict companies' bankruptcy with a certain amount of reliability. The present study sheds more light on various bankruptcy prediction models and some research done in this area. This study deals with financial bankruptcy prediction by using decision tree and logistic regression, and then, it compares these two models with each other.

Theoretical Framework

Along with development of financial markets and accordingly, domination of rival conditions, some companies would be bankrupted and out of rival circle. This will provide worry among stockholders and other beneficiaries who seek those methods by which they can predict financial crises to prevent the loss of their primary and secondary capital. One of the methods of preventing the waste of resources and also proper operation on investment opportunities is prediction of financial distress, in that firstly, companies could be warned about entering bankruptcy with necessary warnings so that they can act upon the necessary operations. Secondly, investors can distinct the desirable opportunities of investments from the undesirable ones and invest their resources on proper opportunities and places.

Generally, the models of bankruptcy prediction could be classified in three major groups.

A) Statistical techniques

Statistical techniques are the most preliminary and most current techniques for providing models of predicting financial crises. Classical standard modeling vision has been used in these models. These models protract the signs of commercial inability which are divided into two groups of one-variable and multi-variable statistical models. Diagnosis analysis, linear probability, and legit models are among them.

B) Artificial Intelligence Techniques

Artificial intelligence technique, similar to human logic intelligence, is a system which learns and improves its own solving problem function according to past experiences. Techniques such as reverse algorithms (decision trees), relative reasoning on the subject, neural networks, and genetic algorithms are among these techniques.

C) Theoretical Models

Unlike statistical and artificial intelligence technique models, these models try to determine the reason of commercial inabilities. These models are naturally multi-variable and usually use statistical analysis to support theoretical issues. Nowadays, because of the existence of constraining assumptions in statistical techniques such as stating the problem of regressions, discriminant analysis and also the existing complications in theoretical models, usage of artificial intelligence techniques including genetic algorithm, Bayesian networks, and decision trees has tuned out to be a widespread matter in predicting various sciences especially financial and managerial affairs.

In the present article, the prediction of financial distress in the accepted companies in Tehran's stock exchange is analyzed by using decision tree. Decision tree is a method for approximation of the goal functions with broken amounts. This method is one of the most famous inductive learning algorithms which is used successfully in various applications.

Review of literature

Lots of financial research has been conducted about financial distress. We will mention some of the related ones later on in this study.

Using artificial neural networks, Shah and Morteza have introduced a model for predicting bankruptcy. The data were taken from the information of 60 bankrupted companies and 54 non-bankrupted ones from 1992 to 1994. They used eight financial ratios that 73 percent of the accuracy of prediction was achieved. Marked and Sraidom used Bayesian networks for the first time to predict bankruptcy. They used two different models in their study, i.e. simple Bayesian model and complex Bayesian model. 228 banks were their sample. Sun and Shenoy (2007) tested bankruptcy prediction by using Bayesian networks. The y used the information of 500 companies. The results of their research indicated that Bayesian model can predict bankruptcy by 89 percent of certainty. Alfaro et al. (2008) worked on two models of Adaboost algorithm and artificial neural networks to predict financial distress of companies. They used the information of 590 bankrupted and healthy companies from 2002 to 2003. The result of this research indicated that Adaboost algorithm has a better function than artificial neural networks and the power prediction of this method is 91.1 percent. Also, their research showed that ratios of profitability, debt and index of company size are the most important variables for making Adaboost model. Joseph James (2011) conducted a research about bankruptcy prediction techniques and their benefits. In this research, most of the models and techniques of prediction were analyzed. The results of this study indicated that among various methods, linear discriminant analysis method can provide a more precise prediction. Niknakht and Sharifi (1389) carried out a research entitled 'financial bankruptcy prediction of the companies in Tehran's stock exchange by using artificial neural networks. The used neural networks in this study are multi-layer perceptron which are trained after error dissemination. Two groups of bankrupted and non-bankrupted companies formed the sample of this study. The bankrupted group was selected based on code 141 of commercial law from 1999 to 2006 and the non-bankrupted one was selected randomly. The results showed that there is a significant difference between MDA and ANN. According to the results, smallness of the first type of error is prioritized on the basis of the second type of prediction error. Fadaienezhad and Eskandari (1390) carried out a research entitled 'designing and explaining of financial bankruptcy prediction model of the companies in Tehran's stock exchange'. The main research question is that which one of the models after dissemination, genetic algorithm and particle swarm optimization can predict bankruptcy of the companies with a higher accuracy. They also compared the effect of the market's data and financial ratios with each other for predicting bankruptcy. The result of this study indicates that genetic algorithm usage is effective in the increase of accuracy in predicting bankruptcy. But the comparison of genetic algorithm and particle swarm optimization reveals that statistically, it could not be proved that one of the methods is superior to the other one. Also, the results indicate that using the market's data is more effective than using financial ratios or simultaneous usage of the market's data and financial ratios. Ahangari (2011) studied about the usage of decision tree algorithm in order to predict the situation of the accepted bankrupted and non-bankrupted companies. In this study, various types of decision trees with limited variables were analyzed to predict bankruptcy. Tabarestani (2012) in a study with the title of 'using a plan based on diagnosis analysis and evaluation of the effect of efficiency variable in plan improvement to predict financial bankruptcy' applies financial ratios which are based on plan of multiple-variable diagnosis analysis. Then, in order to determine the power of the efficiency variable on predicting financial bankruptcy, this variable was used along with financial ratios. The results of the utilized plan show that efficiency variable has an important role in predicting financial bankruptcy.

II. MATERIAL

Research Hypotheses

Accordingly, the hypotheses of the research are formulated as follows:

1. It is possible to predict financial bankruptcy of the accepted companies in Tehran's stock exchange by using decision tree model.
2. It is possible to predict financial bankruptcy of the accepted companies in Tehran's stock exchange by using logistic regression model.
3. The accuracy of decision tree for predicting financial bankruptcy of the accepted companies in Tehran's stock exchange is more than logistic regression model.

Research Population and Sampling Method

The population of the present study contains the accepted companies in Tehran's stock exchange and its sample includes all the companies which have the following conditions:

1. In order to have comparable information, the fiscal year of the companies ends on 29th march.
2. In order to have homogeneous information, banks, insurances, and other intermediary agencies shouldn't be included in the investing companies.
3. The information related to the utilized variables should be available in the present study.
4. The audited financial information theses companies needs to be available to analyze and test the hypotheses.

The data collection follows a library and documentary method. The needed information for analyzing the variables, the financial forms of the accepted companies in Tehran's stock exchange was extracted. For this, the revealed information by Tehran's stock exchange and the notifying software as well as other related internet resources are among the data collection tools. Data analysis was carried out with Excel and Spss clemantine.

Variables

The variables in this research could be classified into two groups which are dependent and independent as follows:

Dependent Variables

A variable is qualitative and is subject to nominal scale. This variable is bankruptcy or lack of bankruptcy. In the present study, the companies subjected to code 141 of commercial law are considered as the bankrupted ones.

Independent Variables

To predict financial bankruptcy, it has been attempted to analyze those variables which possess theoretical support and were use in the previous studies. The list of these variables could be observed in Table 1.

Log (total assets)	X_1
Working capital / total assets	X_2
Net profi / total assets	X_3
EBIT / total assets	X_4
Current assets / current Liabilities	X_5
Operating cash flows / total Liabilities	X_6
Current assets / total assets	X_7
cash flows / total assets	X_8
total Liabilities / total assets	X_9
Non current Liabilities / total assets	X_{10}
Log (sale)	X_{11}
Net sales / total assets	X_{12}
Current assets / sale	X_{13}
Accumulated profit / total assets	X_{14}
Net profit / net sales	X_{15}
Accumulated profit / equity capital	X_{16}
Quick assets / total assets	X_{17}
Quick assets / current liabilities	X_{18}
Interest expense / Gross profit	X_{19}
EBIT / sales	X_{20}
Sales / Current assets	X_{21}
Tangible fixed assets / total assets	X_{22}
Sales / working capital	X_{23}
EBIT / interest expense	X_{24}
Gross profit / sales	X_{25}
Sales / Tangible fixed assets	X_{26}
Net profit / total Liabilities	X_{27}
working capital / Non current Liabilities	X_{28}
working capital / sales	X_{29}
Net profit – operating cash flows / total assets	X_{30}
operating cash flows /sales	X_{31}
operating cash flows / equity capital	X_{32}
equity capital / total Liabilities	X_{33}

III. METHODS

The Utilized Models

As it was mentioned in literature review, in most of the previous research, statistical models, expert systems, theoretical models for financial distress prediction, financial crises of financial functions, etc. were used. Decision tree CHID and logistic regression which are two searching data have been used in the present study. They will be explained later on in this article.

In order to prevent the occurrence of linear correlation between the data, factor analysis was used to reduce the data and we omitted those variables with high correlation degree from the data file.

A) Factor analysis

Factor analysis tries to recognize the major variables or the factors that explain the internal relationship of a set of observed variables. Factor analysis is often used in reducing the data in that it recognizes a few numbers of the factors that explain a more variance and have been observed in major variables. Factor analysis also could be used in formulating hypotheses based on regular mechanisms or observed variables for the next analyzes.

In the first degree, factor analysis is used to reduce the data or to recognize the structures.

The aim of reducing data is that it omits the variables with high correlation from the data file and substitutes them with a few non-correlated variables.

The aim of recognizing structure is to analyze the hidden relationship between variables. Factor analysis provides several strategies for making one solution.

Reduction of the data: the extraction method of the main parts begins with finding a compound line of the variables that calculates a high range of the changes of the main variables. Then, it will recognize another factor by which the remained deviations are calculated, and also this factor is not related to the previous factors. This will continue so that a lot of factors from the main variables are provided. These factors are usually replaced by the main variables. This method is usually used to reduce the number of variables in the data file.

Recognition of the structure: the second strategy provided by factor analysis goes further with the assumption that some of the changes in the data could not be explained by factors.

As a result, the whole of the variance is explained by a smaller solution. In addition to this model, factor analysis creates ideal methods for analyzing the relations between variable. However, by application of the extraction methods we need to answer these two questions:

How many of the substituted variables are required for variables' representativeness.

What do these substitute variables indicate?

B) Decision Tree Model

Decision tree is a strong and regular tool for classification and prediction. Decision tree deals with production of rules and then clarifies its prediction based on a series of rules. This algorithm begins with a test which performs the best separation of the classifications. The main purpose of the separation is obtaining a model for prediction. For this, we use a set of data called "training data" which is a set of variables and records. As for the next step, the same is happened for lower knots with less data to achieve the best rules. Finally, the tree gets bigger and bigger so that it is not possible to provide a better separation rule for the data of knot. In this stage, the effect of the created tree should be measured. To do this, a set of records with training data is used which are different from the primary data that created the tree. The criterion which is measured includes: the percentage of the data which are classified correctly and the predicted classification is identical to their real one.

The decision tree algorithms are various types and the most important ones are Quest, chaid, cS, and cart (Azar et al. 2010). The applied tests are different in each recognized algorithm of the decision tree and selection of the branches and separation of each one id carried out with a different method. In the present study, CHID decision tree has been used as a comparative algorithm.

CHID is one of the classification trees which was introduced by Breiman et al. in 1984. This model introduces a non-rotating graph similar to a tree with binary divisions based on helping variables in order to provide a classification and recognition pattern. Decision tree is constituted of three main parts including root, internal knot, and leaves and at the level of procedure, first, a helping variable is chosen as the root and according to the aims of the study, it is divided to several internal knots. Like roots, each knot is divided into other knots so that finally a class of response variable is allocated to each knot. These knots are called leaves. In order to select important variables in the pattern of tree classification, a function that is called incongruous function and an index called gene have been used. The incongruous function for a knot like t and dependent variable with k class (c_c, c_2, …, c_k) is defined as follows:

$$i(t) = \Phi[P(C = C_1|t), \dots P(c = c_k|t)]$$

Gene index often is used in tree models with binary divisions in each knot and is defined as follows:

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

Among several variables, the one is suitable that takes a higher amount for GG(T,X). This is a criterion for selection of the best one among several variables. Therefore, based in the incongruous function and gene index, first the amount of incongruous function is generally calculated for the response variable. As for the next step, according to the best binary divisions for the response variable, the amount of the incongruous function is calculated and their weight average is subtracted from the total of the incongruous function for all of the helping variables. Among the helping variables, the one that has the highest amount for this relation is selected for tree classification. For quantitative variables, binary divisions are used and a point like a (cutting point) is determined. It should be mentioned that in lots of tree classification patterns, the cutting point is determined by the applied index itself (here gene index). For qualitative variables, each level of the variable is considered as sub-branch of the classification tree (Hoseini et al. 2010).

Logistic Regression Model

In most of the research, the dependent variable has not been continuous and might have two results. For example, from two values of zero and one that the value 'one' means occurrence and the value 'zero' meant non-occurrence of the event (or vice versa), it should take just one of them. For example, we can find out the success or failure of a person in entrance exam by the help of the amount of attempt, and intelligence, or recognize the bankruptcy of a company by using several variables. In such cases, we use logistic regression. Logistic regression is similar to the regular regression except that the coefficient estimation is not identical in them. In logistic regression, instead of minimizing the square of the lines (which occurs in the regular regression), the possibility that an event will happen is maximized. The general form of logistic regression is as follows (Pirayesh, et al. 2009):

$$\mu(x) = \frac{e^{B_i x_i + \beta_0}}{1 + e^{B_i x_i + \beta_0}}$$

$\mu(x)$: dependent variable
X_i: independent variable i=1, 2, 3, 4
B_i: independent coefficient i= 1, 2, 3, 4
β_0: constant digit and e=2.71828182

When the dependent variable is one, we expect that $\mu(x)$ is getting closer to 1 and vice versa. In this article, according to the fact that the companies' bankruptcy is defined as variable zero and one, logistic regression has been used (one for the non-bankrupted companies and zero for the bankrupted ones).

Since the results of this study could be used in the decision-making process, from the viewpoint of aim it is applied and from methodology viewpoint it is descriptive-correlational of cross-sectional type. This study attempts at predicting the dependent variable (bankruptcy) by the usage of other variables (financial ratios). The aim of the study is analysis of the amount of accuracy in classification of the sample companies fro the viewpoint of bankruptcy by using financial ratios taken out of financial forms. According to the fact that the information related to financial situation of the companies in an organized and accurate form is only available in Tehran's stock exchange, the research population is restricted to the accepted companies in Tehran's stock exchange from 2006 to 2011. Decision tree and logistic regression as suitable methods of data analysis were used to investigate the accuracy of bankruptcy prediction by using financial ratios and variables and the effect of these financial ratios and variables were determined in bankruptcy prediction.
The major steps in the present study are as follows:

1. Dividing the companies from financial distress view to two groups of bankrupted and non-bankrupted ones
2. Extraction of the financial ratios from financial forms
3. Reduction of the data dimension by factor analysis
4. Prediction of the financial bankruptcy by using decision tree model and logistic regression
5. Calculation of the prediction accuracy by using logistic regression and decision tree and comparison of these two models

IV. RESULTS and DISCUSSION

Using Factor Analysis to Reduce the Dimension of the Data

The aim of using factor analysis is to achieve the weight or importance degree of each index quantitatively, and also to extract non-correlated compound indices as factors. In this regard, each factor is a liner function from several indices with different weights.
Since just some original major factors explain the most changes in the observations and the rest of the factors have trivial changes the latter ones could be given up.
There are some statistics in factor analysis process through which the researcher is able to determine and recognize the suitable data for factor analysis. KMO (Kaiser-Meyer-Olkin) is among these methods that its amount is always between 0 and 1. If this amount is less than 0.5, the data will not be suitable for factor analysis. For the data of the present study this amount is 0.629. Another test called Bartlett-Test indicates significance of the factor analysis in data and if this amount is less than 0.05, it will be another confirmation that data is suitable for factor analysis. For these data, a very small amount

was achieved that confirms the efficiency of the factor analysis for the data. The results of KMO test and Bartlett-Test are indicated in Table 2.
KMO and Bartlett's Test

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.629
Bartlett's Test of Sphericity	Approx. Chi-Square	6476.936
	Df	528
	Sig.	.000

In factor analysis method, those factors are important which have eigenvalues bigger than 1. For this, among the achieved factors, those ones which have eigenvalues more than 1 have been chosen as the main factors. The results of factor analysis with 33 indices related to bankruptcy indicated that 8 factors with eigenvalues bigger than 1 are existed that determine more than %80 of the total variance.
The related screen plot is indicated in Diagram 1 that shows 8 factors with eigenvalues more than 1.

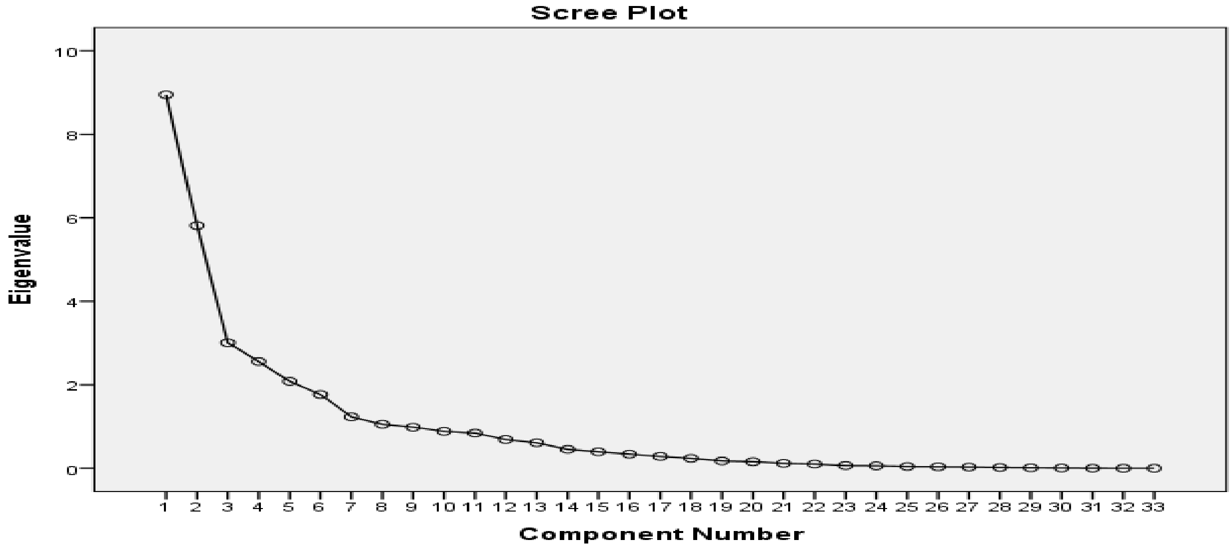

Table 3 shows the variance and variance percent of each chosen factor
Table 3.

Total varianc Explaimed

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	8.948	27.115	27.115
2	5.814	17.62	44.734
3	3.007	9.112	53.846
4	2.556	7.746	61.592
5	2.081	6.308	67.9
6	1.768	5.356	73.256
7	1.231	3.73	76.986
8	1.053	3.189	80.175

The first factor which expresses %27 of the total variance includes effective variables in bankruptcy of the companies. They are the variables x3, x4, x15, x16, x20, x 24, x25, x27, i.e. net profit/total assets, the profit before interest and tax/total assets, net profit/ net sale, accumulated loss and profit/ total assets, accumulated loss and profit/ total equity, profit before interest and tax/sale, profit before interest and tax/cost of interest, gross profit/sale, net profit/total debts.
The second factor which shows %18 percent of the variance includes comparison of the variables x2, x5, x17, x18, x28, x29, x33 with x9. This means, investment flows/ total assets, current assets/ current debts, quick assets/ current debts, investment flows/ long-term debts, investment flow/ sale, total equity/ total debts.

The third factor which shows %9 percent of the total variance includes comparison of variables x7, x26 with variable x22. That is, current assets/ total assets, sale/ tangible fixed assets, tangible fixed assets/ total assets.
The fourth factor which contains %8 percent of the total variance includes variables x6, x30, x31, x32, i.e. operating cash/total debt, operating cash-net profit/ total assets, operating cash/ sale, operating cash/ total equity.
The fifth factor shows %6 of the total variance and includes variables x12, x13, x21, i.e. sale/ current assets, net profit/ total assets, current assets/ sale.
The sixth factor expresses %5 of the total variance including x2, x11, i.e. (total assets) log, (sale) log.
The seventh factor shows %4 of the total variance which includes variable x8, i.e. cash/ total assets.
The eighth factor which shows %3 of the total variance includes variable x10, i.e. long-term debts/ total assets.

Using logistic regression to design the bankruptcy prediction model

In this part, 8 achieved factors are entered logistic regression as variables in order to predict the bankruptcy of the companies. The fitting results of the logistic regression on factors are indicated in Table 4.

Table6 . Result of Logistic regression model

	B	S.E.	Wald	Df	Sig.	Exp(B)
FAC1	-2.294	.458	25.132	1	.000	.101
FAC2	-1.076	.330	10.663	1	.001	.341
FAC3	.829	.340	5.953	1	.015	2.292
FAC4	-.449	.309	2.110	1	.146	.638
FAC5	-.772	.264	8.543	1	.003	.462
FAC6	-.658	.310	4.498	1	.034	.518
FAC7	-.057	.283	.041	1	.840	.945
FAC8	.421	.321	1.722	1	.189	1.524
Constant	.077	.284	.074	1	.786	1.080

The fitted logistic regression equation is as follows:

$$\log\left(\frac{p}{1-p}\right) = 0/077 - 2/294FAC1 - 1/076FAC2 + 0/829FAC3 - 0/449FAC4 - 0/772FAC5 - 0/658FAC6 - 0/57FAC7 + 0/421FAC8$$

Here, p shows the possibility of bankruptcy.
The first, second, third, fifth and sixth factors are the important ones in this model which have P-Value (significance value) of less than 0.05. Interpretation of the significant factor coefficient is as follows:
For each unit increase of the first factor and controlling the rest of the factors in a fixed amount, the chance of non-bankruptcy in a company is 10 (1/0/101) times more than the time that the mentioned factor is not increased for one unit.
For each unit increase of the second factor and controlling the rest of the factors in a fixed amount, the chance of non-bankruptcy in a company is 3 (1/0/341) times more than the time that the mentioned factor is not increased for one unit.
For each unit increase of the third factor and controlling the rest of the factors in a fixed amount, the chance of bankruptcy in a company is 2.292 times more than the time that the mentioned factor is not increased for one unit.
For each unit increase of the fifth factor and controlling the rest of the factors in a fixed amount, the chance of non-bankruptcy in a company is about 2.2 (1/0/462) times more than the time that the mentioned factor is not increased for one unit.
For each unit increase of the sixth factor and controlling the rest of the factors in a fixed amount, the chance of non-bankruptcy in a company is about 2 (1/0/518) times more than the time that the mentioned factor is not increased for one unit.
According to the findings shown in Table 5, based on the classification table achieved from logistic regression on the given value, 79.6 percent of the non-bankrupted companies, 88.7 percent of the bankrupted companies were classified correctly. In general, 84.5 percent of the companies were classified correctly.

Table 5. Classification based on regression

Observed		Predicted		
		Group		Percentage Correct
		Nonbankrupt	Bankrupt	
Group	Nonbankrupt	43	11	79.6
	bankrupt	7	55	88.7
Overall Percentage				84.5

Bankruptcy Prediction of the Companies in Decision Tree Model CHAID

In this part, in order to predict the financial function of the company by usage of the factors, 8 determined factors have been entered to the model by factor analysis. Diagram 2 shows the created tree.
This model has been fitted with 116 companies that %46.6 of them is non-bankrupted and %53.4 bankrupted. The important factors which are determined by decision tree are the first and the third ones. If the score of the first factor is less than 0.43 or equal to 0.43 in a company, the probability of its bankruptcy is %84.8 and the probability of its non-bankruptcy is %15.2. If the score of the first factor in a company will be higher than 0.43 and less than or equal to 0.08, the probability of bankruptcy will be %58.3 and its non-bankruptcy probability is %41.7. The companies in which the score of the first factor is higher than 0.08, if the score of the third factor is less than or equal to 0.043, the bankruptcy probability of that company will be %0 and its non-bankruptcy will be %100. And if the score of the third factor in this company is higher than 0.043, the bankruptcy probability of this company will be %59.1 and its non-bankruptcy %40.9.

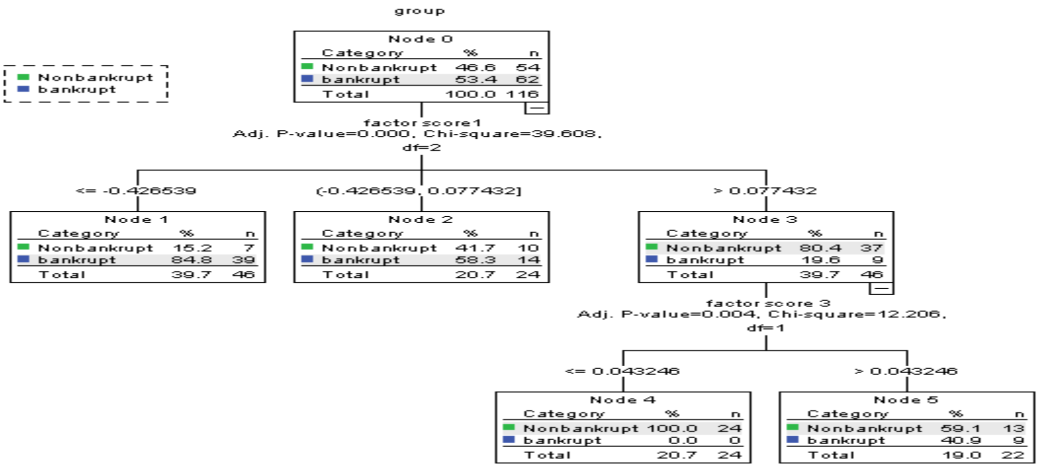

Diagram 2. The created decision tree

According to the information in Table 6, regarding the accuracy of the model, one can say that among the non-bankrupted companies 37 cases out of 46 cases and among the bankrupted companies 52 out of 70 have been recognize correctly. In general, %787.6 of the companies has been classified correctly.

Table 6. Classification table of the prediction accuracy

Classification			
Observed	Predicted		
	Nonbankrupt	Bankrupt	Percent Correct
Nonbankrupt	37	17	68.5%
Bankrupt	9	53	85.5%
Overall Percentage	39.7%	60.3%	77.6%

Growing Method: CHAID
Dependent Variable: group

Comparison of Logistic Regression and Decision Tree, Using ROC Model and Classification Table

In order to know whether the results of decision tree and logistic regression are close to each other or not, first we need to use ROC model. Diagram 3 shows ROC model. In fact, the more concave is this diagram, the prediction accuracy will be better and also, the more logistic regression diagram and decision tree are close to each other, and the results of the logistic regression and decision tree are closer to each other. Consequently, according to the area under the curve it could be said that prediction of logistic regression is better than decision tree.

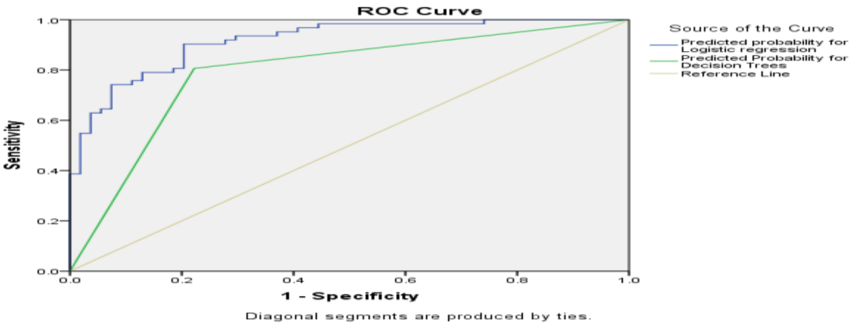

Diagram 3. ROC

According to Table 7 that shows classification of logistic regression and decision tree, logistic regression classifies %84.5 of the total of the companies correctly while decision tree model classifies %77.6 of the companies correctly. Therefore, we observe that the prediction ability of logistic regression for this set of data is more than decision tree.

Observed		Predicted for logistic Regression		
		Group		Percentage Correct
		Nonbankrupt	Bankrupt	
Group	Nonbankrupt	43	11	79.6
	bankrupt	7	55	88.7
Overall Percentage				84.5
Observed		Predicted for decision trees		
		Nonbankrupt	Bankrupt	
Group	Nonbankrupt	37	17	68.5%
	bankrupt	9	53	85.5%
Overall Percentage				77.6%

IV. CONCLUSION

Testing the Research Hypotheses

The Results of Testing the First Hypothesis

The first hypothesis of this research is formulated as follows:

It is possible to predict financial bankruptcy of the accepted companies in Tehran's stock exchange by using decision tree model.

To test this hypothesis, we need to compare the accuracy of bankruptcy classification of the companies in different models in the form of bankrupted and non-bankrupted companies. For testing this hypothesis, we compared the classification accuracy of bankruptcy prediction of the companies in different models in the form of bankrupted and non-bankrupted companies.

Decision tree Chid model also was used to test the hypothesis. It shows that the percentage of accuracy for bankruptcy prediction for non-bankrupted companies is %65.5 and for bankrupted ones it is %85.5. The total accuracy of the model is %77.6.

The Results of Testing the Second Hypothesis

The second hypothesis is formulated as follows:

It is possible to predict financial bankruptcy of the accepted companies in Tehran's stock exchange by using logistic regression model.

Logistic regression analysis is used to test this hypothesis. According to the achieved results we can interpret that it is possible to predict bankruptcy by using logistic regression model. In fact the accuracy percentage for predicting the bankrupted companies is %88.7 and for non-bankrupted ones it is %79.6. The total accuracy of this model is %84.5.

The Results of Testing the Third Hypothesis

The third hypothesis of the present article is formulated as follows

The accuracy of decision tree for predicting financial bankruptcy of the accepted companies in Tehran's stock exchange is more than logistic regression model.

To test this hypothesis, the financial bankruptcy of the companies was compared in the form of bankrupted and non-bankrupted companies. Decision tree CHID and logistic regression were also used to test this hypothesis. Based on the results, we one can conclude that bankruptcy prediction accuracy by using logistic regression is better than decision tree model, though the results of logistic regression model and decision tree model are close to each other.

Limitations of the Study

To accomplish the aims is always along with some limitations that make it slower to get to the aims. As a process of achieving the aims of solving the research problem, research is not an exception of this fact. Therefore, in this part there is an attempt to declare the limitations of the study so that the reader will be able to operate with more knowledge upon the generalization of the results and have a fair judgment on the procedure of the study. In this regard, the limitations of the present study are explained as follows:

a) The time period of this study is from 2006 to 2011. Thus, application of the results of this study to the before and next periods should be done with caution.
b) The extracted data from the financial forms of the companies have not been adapted regarding inflation. Under the condition of adapting the mentioned data, different results might be achieved.
c) Investment, leasing, and insurance companies were omitted from the population of the present study because of their specific nature of activities and the recorded digits in their financial forms. Therefore, application of these results to these companies should be carried out with caution.

Suggestions

Based on the results of the study, in this part, two kinds of suggestions are introduced. First, applied suggestions that we hope will help the users of the financial and accounting information especially investors in decision-making. And second, suggestions for the future studies which could be considered as guides for the next studies regarding the present topic.

Suggestions Based on the Results of the Study

According to the results of the first and second study, the following suggestions could be proposed:

A) Tehran Stock exchange can gain a relative confidence from the suitable financial conditions of the future companies by using the models applied in this study to accept new companies.
B) According to the standards of auditing, the auditor is charged of evaluating the maintenance of the unit activity that is under investigation, and he report on it obscurely. Therefore, the related financial ratios can help the auditor in auditing the financial forms of the company.
C) Banks and credit-financial institutes can use the models of this study to pay loans with huge sums to the industry applicants.
D) Stock exchange agents, analysts and financial counselors whose duty is to analyze the financial situation of the companies in stock exchange and to illustrate the future financial conditions of the companies for the applicants of stock and also those who are the financial providers of the companies can use the results of this study to have a more appropriate analysis.

Suggestions for Future Studies

Some of the subjects and cases which could be taken into account for future studies include:

A) The hypotheses of the present study were tested without considering the industry in which the companies work. It is recommended that the researchers also formulate the hypotheses based on the level of each industry.
B) Two data-based models were used in this study to predict bankruptcy. The future studies can revolve around the models such as: neural networks, genetic algorithm, the nearest neighbor algorithm, etc.
C) The studies about the models could be conducted for the companies out of Tehran stock exchange.
D) In order to increase the level of confidence of the results, the study could be done in longer periods.

REFERENCES

[1] m. Ahangari, “using decision trees to predict bankrupt and non – bankrupt companies listed on stock Exchany,” 2011.
[2] E. Alfaro, N. Garcia, “Bankruptcy Forecasting: An Empirical Comparison of Adaboost and Neural Networks,” Decision Support Systems., Vol. 45, pp. 110-122. 2008.
[3] Azar et all, “ Model design Man power Data Mining Approach,” Journal of information Te chnology and Manag Nent ., No. u. pp. 22 -3. 2010.
[4] S. M. Bryant, “A Case-based Reasoning Approach to Bankruptcy Prediction Modeling,” Intelligent System Accounting, Financial and Management., Vol. 6, pp.195-214. 1997.
[5] Hossenini etall , “ classification tree used to predict diabctic and type,” diabetic patients at diag nosis compared with the model of Isfahan medical school., no lou-pp.27-17. 2010.
[6] J. joseph, “Bankruptcy prediction and its Advantages in the frech Hospitality Industry,”The Msc programmer in Economics and Business Admi stration, 2011.
[7] M. Nikbakt, “Tehran Stock Exchanye company bankruptcy prediction using neural networks,”Journal of Tech nology Manag Ment, Vol 2 (u), pp.180 – 163. 2010.
[8] R .Pirayesh , “to design a Mathematical Model based on cash flow forecasting bankruptcy and listed company-es in Tehran Slock Exchange ,”Journal of investment No. u. pp. 9u – 73. 2009.
[9] J.R. Shah, M.B. Murtaza, “A Neural Network Based Clustering Procedure for Bankruptcy Prediction,” American Business Review, pp. 80-86. 2000.
[10] L Sun, P. Shenoy, “Using Bayesian Networks for Bankruptcy Prediction,”European Journal of Operational Research, Vo1.180, Issue. 2, pp. 738-753. 2007.