BUSINESS FAILURE PREDICTIONS IN ISTANBUL STOCK EXCHANGE

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ONUR TEKEL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF
MASTER OF BUSINESS ADMINISTRATION
IN
THE DEPARTMENT OF BUSINESS ADMINISTRATION

JUNE 2009

Approval of the Graduate School of Social Sciences

Prof. Dr. Sencer Ayata
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Business Administration.

Prof. Dr. Cengiz Erol
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Business Administration.

Assoc. Prof. Dr. Ramazan Sarı
Supervisor

Examining Committee Members

Assistant Prof. Dr. Derya Erel	(AIBU, BA)	____________
Associate Prof. Dr. Ramazan Sarı	(METU, BA)	____________
Associate Prof. Dr. Uğur Soytaş	(METU, BA)	____________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name:

Signature :

ABSTRACT

BUSINESS FAILURE PREDICTIONS IN ISTANBUL STOCK EXCHANGE

Tekel, Onur

MBA, Department of Business Administration

Supervisor: Assoc. Prof. Dr. Ramazan Sarı

June 2009, 46 pages

This study aims to develop business failure prediction models using the data of selected firms from ISE markets. The sample data comprise ten selected financial ratios for 27 non-going concerns (failed businesses) and paired 27 going concerns. Two non-parametric classification methods are used in the study: Artificial Neural Networks (ANN) and Decision Trees. The classification results show that there is equilibrium in the classification of the training samples by the models, but ANN model outperform the decision tree model in the classification of the testing samples. Further, the potential usefulness of ANN and Decision Tree type data mining techniques in the analysis of complex and non-linear relationships are observed.

Keywords: Artificial Neural Networks, Decision Trees, Business Failure Prediction, Data Mining

ÖZ

İSTANBUL MENKUL KIYMETLER BORSASI'NDA İŞLETME BAŞARISIZLIĞI TAHMİNLERİ

Tekel, Onur

Yüksek Lisans, İşletme Bölümü

Tez Yöneticisi: Doç. Dr. Ramazan Sarı

Haziran 2009, 46 sayfa

Bu çalışmada İMKB pazarlarından seçilmiş şirketlerin verileri kullanılarak işletme başarısızlığı tahmin modelleri geliştirilmeye çalışılmıştır. Örneklem verisi 27 adet başarısızlığa uğramış şirket ile onlarla eşlenmiş 27 adet faaliyetlerine devam eden şirketin seçilmiş onar adet mali oranından oluşmaktadır. Çalışmada iki adet parametrik olmayan sınıflandırma yöntemi kullanılmıştır: Yapay Sinir Ağları (YSA) ve Karar Ağaçları. Sınıflandırma sonuçları alıştırma örnekleminin sınıflandırmasında modellerin dengede olduğunu göstermektedir, ancak YSA modeli test örnekleminin sınıflandırmasında karar ağacı modeline üstün gelmiştir. Ayrıca, YSA ve Karar Ağacı gibi veri madenciliği tekniklerinin kompleks ve lineer olmayan ilişkilerin analizindeki potansiyel faydaları gözlemlenmiştir.

Anahtar Kelimeler: Yapay Sinir Ağları, Karar Ağaçları, İşletme Başarısızlığı Tahmini, Veri Madenciliği

To My Family

ACKNOWLEDGEMENTS

I wish to express my deepest appreciations to my supervisor Assoc. Prof. Dr. Ramazan Sarı for his invaluable guidance, advice, criticism, encouragements and insight throughout the preparation of this thesis. Likewise, I want to present my special thanks to Asst. Prof. Dr. Seza Danışoğlu and Assoc. Prof. Dr. Uğur Soytaş for their suggestions and guidance. Many thanks also to my friends Ekin Uzun and Ö. Mert Giray for reading my thesis and giving me valuable criticism during the revising phase. I am also grateful to my family for their everlasting support and encouragements.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

Making good decisions is directly related to reducing the unstability and uncertain conditions that each and every decision-maker encounters. In a frequently changing business atmosphere, educated guesses about the future are really valuable, because these guesses give the organizations the chance of being ready against the situations that are unusual. Thus, every professional uses forecasting methods to be able to see at least one step ahead from their current position.

> Forecasting methods can be examined in two broad categories: qualitative methods (depend on subjective assessments of people, known also as "judgmental methods"), and quantitative methods (which are mechanical procedures and depend on manipulation of the historical data, known also as "objective methods"). Judgmental techniques are generally useful when the raw data that is needed to run a mathematical method is not present. Thus, decision-makers are forced to make rough estimates by using judgmental methods. Although these methods are very useful, they are biased toward the user group who develops them. Quantitative methods category can be examined in two segments: time series and causal. Time series techniques heavily rely on historical data which is measured in successive and equal time periods. The causal methods use the links between variables and the effects of variables on the change of another variable. (Chase and Charles 1997)

Whatever the method is, usually an accurate forecast means significant increases in revenue, decreases in expenses, or both. Moreover, costly mismatches of capacity to demand can often be avoided through consistent forecasting, and thus professionals who are responsible for monetary decisions become effective in cost controlling issues (Freeman 2003).

Statistical classification, which is a type of causal forecasting, is a procedure of placing individual items into groups, based on quantitative information about the characteristics of these items. Some well known statistical classification methods are: logistic regression (LR), decision trees, neural networks and discriminant analysis. These methods are used heavily in prediction of financial performance studies. Some of these studies are credit risk measurement (Altman and Saunders 1998), business failure prediction (Beaver 1966), and financial distress prediction (Chen and Du 2009). The results of these studies are important for a group of people and financial institutions who care about the financial position and the lifetime of the firms. Investors, financial market specialists, creditors, stockholders and managers are some of the main stakeholders within this group.

The expectations of the stakeholders in Turkey are not different from the rest of the world. Knowing that fact, this study examines the potential usefulness of financial ratios in the prediction of business failures in Turkey. In this study, "listing" status of the firms in the Istanbul Stock Exchange (ISE) markets is used as an indicator of financial performance, and the firms which were permanently delisted from the ISE markets are thought as "Financially unsuccessful". Within this frame, ANN and decision tree methods were used to classify "listed" and "delisted" firms which were selected for the study, using the key determiner characteristics of financial ratios. The results of the models that were developed using the methods above were compared to show the similarities and differences in the prediction capabilities of two methods.

Looking to the business failure prediction studies that were done for the Turkish case, it can be said that most of these studies were done for the banking sector. Moreover, studies that compare non-parametric techniques are relatively rare. Therefore with this study, it is tried to compensate the need for a study which compares two non-parametric techniques and examines the sectors other than the banking sector.

The structure of the rest of the study is as follows. In the next chapter, relevant literature about the prediction of financial performance issues is reviewed. Chapter three includes the generation of the data set, and the explanation of the variables and methods used in the study. ANN and decision tree models are developed in chapter four. Chapter five presents the results that are derived from the study and finally chapter six concludes.

CHAPTER 2

LITERATURE REVIEW

As a main title, "financial performance prediction" has been studied by scholars more than a half century. Under this main title, "prediction of bankruptcy", "prediction of business failures", "prediction of stock performances", "corporate financial distress prediction" are the subtitles that comes to minds first. It will probably not be wrong to say that the study pioneered all these titles, belongs to Beaver (1966). In this study, key accounting ratios of some failed and non-failed firms are compared at a point in time within a univariate analysis frame.

Beaver's univariate study was improved by Altman (1968) by the application of multivariate approach. In this study (also known as Z-score model), Altman used linear discriminant analysis for the first time to develop a business failure prediction model. This model used a set of financial ratios (actually five ratios) to obtain a way of measure which makes a distinction between failed and non-failed firms. Moreover, this model proved to be successful in predicting business failures in 94% of the initial sample.

After Altman (1968)'s study, several studies that uses discriminant analysis to predict business failures were made. Deakin (1972) used the same 14 ratios that Beaver (1966) used, and tried to find a linear combination of these ratios which best predicts a potential business failure. Edmister (1972) tested the usefulness of financial ratios in predicting failures of small businesses. Blum (1974) constructed a model called "The Failing Company Model" by examining three general denominators (liquidity, profitability and variability) underlying the cash-flow framework. Altman et al. (1974) developed a multiple discriminant model for assessing the credit worthiness of commercial loan applicants in textile sector of France. Altman et al. (1977) constructed a second generation model with several improvements to the Z-score approach.

There are models that are constructed with methods besides discriminant analysis. Ohlson's (1980) Logit model and Zmijewski's (1984) Probit model tried to obtain the probabilities of firm failures and measure the effects of some specific variables on these probabilities. In addition to that, Zmijewski (1984) examined sampling biases on the selection of the firms in his study.

Coming to the non-parametric techniques, decision trees and artificial neural networks are the well known ones. Decision trees are trees that classify data, based on the attribute values. Attributes are represented by nodes and values are represented by branches in a decision tree. Friedman's (1977) RPA (Recursive Partitioning Algorithm), Kass' (1980) CHAID (Chi squared Automatic Interaction Detector) algorithm, Breiman's (1984) CART (Classification and Regression Trees) algorithm and Quinlan's (1986, 1993) ID3 (Iterative Dichotomiser 3) and C4.5 (an extension of ID3) algorithms are the most known decision tree algorithms. Frydman et al. (1985) applied RPA to their financial distress model and compared this technique to discriminant analysis. ID3 method was applied in credit scoring by Carter and Catlett (1987) and corporate failure prediction by Messier and Hansen (1988). Chung and Tam (1993) compared ID3 method to two other inductive learning algorithms by applying these methods to a construction project assessment task and a bankruptcy prediction task. Shirata (1998) studied the bankruptcy in Japan in her empirical study with a CART model. Koh and Low (2004) used CHAID algorithm in their going concern prediction study which they compared the usefulness of neural networks, logistic regression and decision trees. In this study they observed that decision tree going concern prediction model outperformed neural network and logistic regression models. Chen et al. (2006) made a comparative study for the prediction of business failures in China. In this study, they used logistic regression, neural network, decision tree (CHAID), and linear discriminant analysis models. They found logistic regression and neural network models as optimal prediction models for this study. Doumpos and Zopounidis (2006) explored the combination of different classification methods in developing efficient models for credit risk assessment. Within this context they used CART algorithm. Koyuncugil and Ozgulbas (2007) examined small and medium-sized enterprises of Turkey in their study. In this study, they created financial performance profiles of these enterprises by using CHAID algorithm, and also they tried to develop an early financial performance warning mechanism for these enterprises.

Artificial neural networks can be defined as computational (or mathematical) models inspired by the nature of biological neural networks. A neural net is made-up of simple processing elements called neurons, units, cells, or nodes. Each neuron is connected to other neurons by directed communication links with associated weights, and these weights represent information that is used to solve a problem (Fausett 1994). Odom and Sharda (1990)'s study was one of the first studies that applied ANN to bankruptcy prediction. In this study, they used Altman (1968)'s financial ratios as the inputs of the ANN model that they created for a number of bankrupt and solvent US firms. They also created a multiple discriminant analysis (MDA) model as a comparison. Tam and Kiang (1992) studied bank failure prediction and made a comparison between ANN and some other methods (MDA, LR, k-nearest neighbor (kNN), ID3). Empirical results of this study showed that ANN offered better predictive accuracy than the other methods used in this study. Salchenberger et al. (1992) studied prediction of thrift failures. In their study they compared neural networks and logit models in accuracy of predicting the failures, and observed that the ANN model outperformed the LR model. Fletcher and Goss (1993) built an ANN and a LR model using the company failure data of the study of Gentry et al. (1985), and compared these models according to their accuracy in predicting the bankruptcies. The result of this study showed that the ANN model predicted the bankruptcy data better than the LR model. Wilson and Sharda (1994) made an explanatory study that compared the predictive capabilities of ANN and MDA. Both of the models used Altman (1968)'s five financial ratios as inputs and the dataset which includes some bankrupt and non-bankrupt firms, was obtained from Moody's Industrial Manual. In this study, authors suggested that the ANN model outperformed the MDA model in predicting the corporate bankruptcy. Boritz and Kennedy (1995) studied bankruptcy prediction by comparing a number of techniques including different neural network procedures, MDA and LR. In this study, they also used the bankruptcy prediction models that were developed by Altman (1968) and Ohlson (1980) as benchmarks. However, the results of the comparisons were inconclusive, or with their own words, neural network solutions didn't achieve the "magical" results. Fernandez and Olmeda (1997) compared ANN with MDA, LR, C4.5 and MARS (Multivariate Adaptive Regression Splines) in their study which examines Spanish banking sector. According to this study the ANN model had a better accuracy level. On the other hand, the authors showed that when two or more methods were combined the predictions generally were more accurate than any single method in the study. Kiviluoto (1998) used self-organizing maps (SOM), which is an ANN approach, in his study of classifying Finnish firms according to financial situation of being healthy or bankrupt-prone. As a result, he showed that an unknown firm could be easily and quite reliably characterized on the basis of its position on the map. Lin and McClean (2001) made a bankruptcy prediction study by using the financial ratios of bankrupt and non-bankrupt firms from UK. They applied MDA, LR, C5.0 (a decision tree algorithm), ANN and hybrid classifiers to the data, but before applying the methods they used human judgment and ANOVA (analysis of variance) techniques for selecting the appropriate ratios from a group of financial ratios. At the end of the study, they found the ANOVA feature selection was better than the human judgment feature selection except for MDA, and also ANN and decision tree methods showed better performance than the statistical methods. Moreover, hybrid classifiers that used more than one method were better than all methods in predicting the failure in the first year. Charitou et al. (2004) developed failure prediction models for UK public industrial firms using LR and ANN. They observed that ANN models achieved the highest overall classification rate of 78%. Brockett et al. (2006) examined the ability of two statistical (MDA and LR) and two neural network (back-propagation and learning vector quantization (LVQ)) models to identify financially troubled life insurers. This study showed that the neural network models outperformed the traditional statistical approaches. Boyacioglu et al. (2009) aimed to apply and evaluate four different neural network models, support vector machines and three multivariate statistical methods to the problem of predicting bank failures in Turkey. Results of this study showed that two neural network models (multi-layer perceptron (MLP) and learning vector quantization (LVQ)) could be considered as the most successful models in predicting the financial failure of banks.

In this particular study, ANN and decision tree techniques were used to develop models that classify the selected firms according to their financial performances and using financial ratios. With this study it was aimed to compensate the need for a study which compares non-parametric techniques and examines the positions of sectors other than the banking sector for the Turkish case.

CHAPTER 3

DATA AND METHODS

3.1 Data and Variables

The data set used in this study was obtained from the Istanbul Stock Exchange (ISE)'s website and includes the data of two groups of firms. Firms from the first group are those whose stocks were de-listed (it can be explained as a financial performance failure) from the ISE markets permanently as from 2000. Firms from the second group are those whose stocks are presently listed in the ISE markets. Selection procedure for these two groups is as follows:

Group 1 (de-listed group) consists of 27 firms, which were directly selected from the list that was published in the ISE website[1] according to the following criteria. The first criterion was the availability of the financial statements of the firms in the ISE web site. It was a problematic issue to reach the financial statements of these firms since they were de-listed. It was only possible to reach the statements for the years prior to the year that the firms were de-listed. This situation raised the question of which year prior to failure should the statements belong to?

Looking to the previous studies, it can be seen that there is not a unique approach to this situation. For instance Beaver (1966) used the data of the first year before failure in his study; on the other hand Chen and Marshall (2006) used data that

[1] See Appendix A for the full list and the source.

belonged two years before failure in their study. There are several different examples about the usage of the statements in the literature.

In this study, firstly data from the last year prior to failure (or de-listing date) were tried to be used, but the statements were not available for all firms. Therefore, the availability of data was checked for two years prior to failure for all of the firms. This time, most of the firms' statements were available except for a few of them. A further inspection was done by going two years deeper in the past to see whether the statements of those years for these exceptional firms were available, but result was still negative for some of the firms. Due to this data availability issue and the issue of keeping the predictive accuracy of the data, it was thought to use the data from two years before the failure for the de-listed group.

The second criterion of the firm selection stage for Group 1 was the sectors of the firms. In the original list which was downloaded from the ISE web site, there used to be banks and insurance companies. These firms were excluded from the study since this kind of financial institutions have different financial statement structure. Therefore, it was thought that for having a standard form of financial statements in the study, excluding these firms would be beneficial.

After applying these two criteria, there were 29 firms on hand. In the second part of this selection process, firms which are presently "listed" (these are Group 2 firms which are going concerns) were selected based upon a paired-sample design. In this part, previously selected 29 firms were paired with going concerns[2] according to their sectors and asset sizes. The core issue in this stage was the asset sizes. The asset sizes of those 29 firms were obtained from the financial statements and they were tried to be paired with the going concerns (from the same sectors) so that the firms in the pairs could have close asset sizes. However, sometimes finding such pairs was not possible. Therefore, in such situations, pairing the firms whose asset sizes had the same number of digits (in terms of TL)

[2] Financial statements of these going concerns belonged to the year that their failured pairs' financial staments belonged to.

or pairing the firms that had at most one digit difference was applied as a heuristic to match failed firms and going concerns. For instance, a Group 1 firm with TL 300 million asset size, could be paired with a Group 2 firm with TL 400 or TL 40 million asset size, but not with a firm with TL 4 million asset size. By using this heuristic, it was ensured that the firms in the same pairs had the same scale in terms of asset sizes.

Two of the Group 1 firms could not be paired with any going concerns in this stage, because one of them was a very large company and the other one was a small company. Thus, after the pairing process there were 27 pairs on hand which meant 54 firms[3]. In this study, the paired-sample design was applied in order to provide a control over the factors (sectoral factors and factors caused by size) that might have affected the predictive accuracy of the study (Beaver 1966).

Financial ratios are generally good predictors of performance, since the interpretation and the calculation of them are easier than trend and static (vertical) analyses (Akman and Mugan 2007). Moreover, financial ratios can reveal a company's strengths and weaknesses better than the individual items on the financial statements (Brigham and Ehrhardt 2008).

In this study, 11 variables were used to create the models. 10 of them were financial ratios and were used as explanatory (independent) variables, and the one remaining variable was a dummy variable that showed whether a firm was listed or de-listed. Three criteria were used to select these 10 ratios from the set of ratios available in the literature. These criteria were 1) popularity in the literature, 2) performance in the previous studies, and 3) potential relevancy to the study and computability with the information gained from the financial statements (Beaver 1966 and Altman 1968). The ratios that were used in the study are as follows:

[3] See Appendix B.

- **Current Ratio (Current Assets / Current Liabilities):** "This ratio is a test of short-term liquidity or in other words it tests a firm's ability to meet the short-term obligations" (Akman and Mugan 2007).
- **Quick Ratio (or Acid Test):** "This ratio is calculated by deducting inventories from current assets and then dividing the result by current liabilities" (Brigham and Ehrhardt 2008).
- **Working Capital / Total Assets:** "Working capital is defined as the difference between current assets and current liabilities. Therefore, this ratio is a measure of the net liquid assets of the firm relative to the total capitalization" (Altman 1968).
- **Earnings Before Interest and Taxes / Total Assets (Basic Earning Power):** "This ratio shows the raw earning power of a firm's assets, before the influence of taxes and leverage and a good measure for comparing firms with different tax situations and different degrees of financial leverage" (Brigham and Ehrhardt 2008).
- **Retained Earnings / Total Assets:** "This ratio is a measure of cumulative profitability over time, and the age of a firm is implicitly considered in this ratio" (Altman 1968).
- **Sales / Total Assets (Total Assets Turnover Ratio):** "This ratio measures the turnover of all the firm's assets or in other words indicates the sales generating ability of the firm's assets" (Brigham and Ehrhardt 2008).
- **Return on Total Assets:** "This ratio measures the profitability of total resources available to the business. It indicates how efficiently the management has used the resources to create income" (Akman and Mugan 2007).
- **Return on Common Equity:** "Stockholders invest to get a return on their money. Therefore, this ratio tells how well stockholders are doing in an accounting sense" (Brigham and Ehrhardt 2008).

- **Total Liabilities / Total Assets (Debt Ratio):** "This ratio measures the percentage of funds provided by creditors, thus it is an indication of financial leverage used by the company" (Akman and Mugan 2007).
- **Paid-in Capital / Total Assets:** "This ratio measures how much of the firm's funds are coming from the firm's capital in terms of percentage. It is very similar to debt ratio" (Turetken 1999).

In the following sections, predictive models were developed and used to test the predictive accuracy of these financial ratios. Therefore, in the "Methods" section a general description of the methods was given and in the "Models" chapter the predictive models were built by using these methods.

3.2 Methods

"Artificial neural networks (or neural networks) are useful for recognizing patterns in a data set when the type of the relationships between the dependent and independent variables is unknown or complex" (Koh and Low 2004). "Neural networks consist of basic units (nodes) that imitate the behavior of biological neurons found in the nature" (Berry and Linoff 2004). These nodes are inside the layers, and a neural network model consists of two or more layers. "Each node receives inputs from at least one node in a previous layer and combines these inputs, then generates output to at least one node in the next layer" (Koh and Low 2004) (see Figure 1). In a neural network model, each input has its own weight and these weights are used to calculate the weighted combination of the input signals (data) that a neuron receives from the other neurons. These signals activates the neurons, the neurons process these input signals and transfer the signals as outputs for the neurons which they are connected to. This is the basic description of how a neural network processes the data.

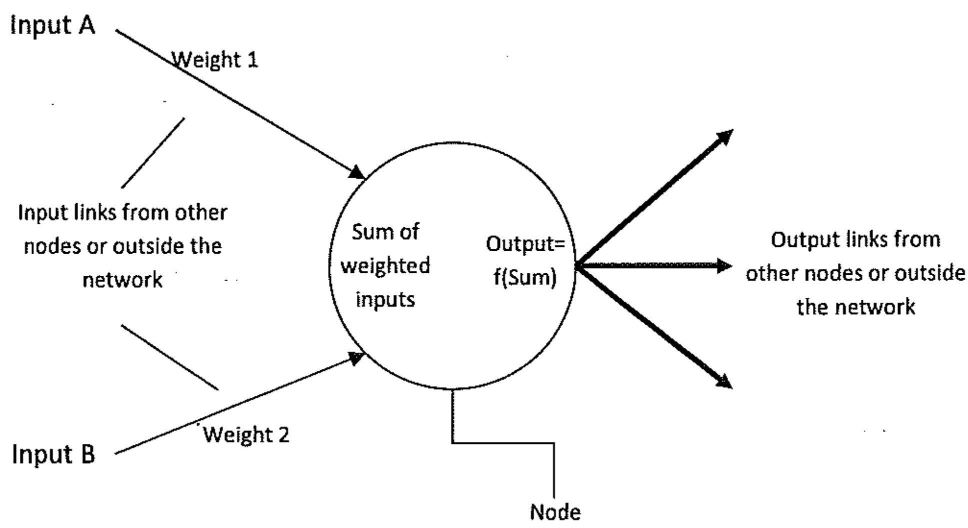

Figure 1 A Sample Node and Its Data Stream

In a feed-forward neural network, the units are organized into three layers: input layer, hidden layer and output layer (see Figure 2). Signals flow from input layer to the output layer through hidden layer. Each unit in the hidden layer is fully connected to all units in the input layer. "The units in the hidden layer calculate their output by multiplying the value of each input by its corresponding weight, adding these up and applying a transfer function" (Berry and Linoff 2004). In the output layer, a transfer function is applied to the information coming from the hidden layer again and the final output is obtained.

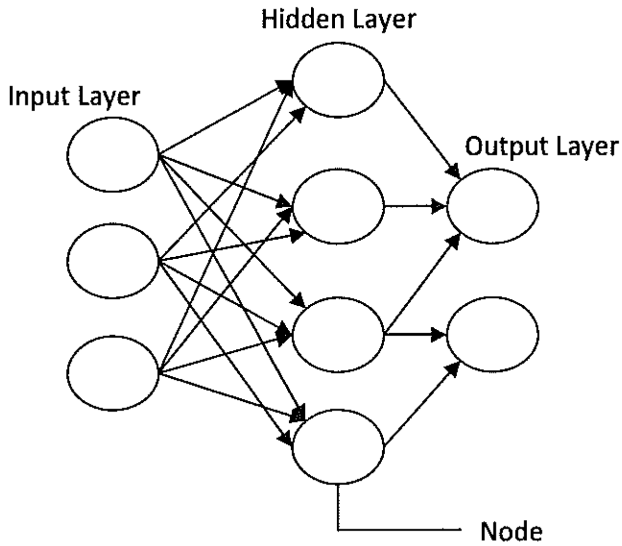

Figure 2 A Sample Feed-Forward Neural Network

One of the most popular neural network training methods is "backpropagation" (BPN). Training is the process of assigning the best weights for the connections between all units in a neural network. The goal of the training process is setting the weights in order to make the output as close to the desired output as possible. The backpropagation algorithm measures the overall error of the network by comparing the values produced on each training process to the actual value, and then a group of adjustment procedures are followed to lower the output error to an acceptable level. When the weights on the network no longer change significantly and the error no longer decreases, the training stops. This is the point that the network has learned to recognize the patterns in the input. After the training process the network model is "tested" by using a new set of data that was not used in the training process. The output that is obtained from here is compared to the answer which is already known and an accuracy measure is calculated. This ensures to examine how well the network has learnt to set the correct weights and recognizes the patterns (Turetken 1999).

Like artificial neural networks, decision tree models are popular and powerful data mining techniques for prediction. A decision tree is the structure which has the ability to divide a large collection of data into successively smaller parts by applying a group of decision rules. "Basically the decision tree approach begins by searching the independent variable that divides the sample in such a way that the difference with respect to the target variable (dependent variable) is greatest among the divided subgroups" (Koh and Low 2004). "This process continues repeatedly, and the data is split into smaller groups in such a way that each new generation of nodes has greater purity than its ancestors with respect to the target variable" (Berry and Linoff 2004).

There are several decision tree algorithms in the literature. The most popular of these tree algorithms are Kass' (1980) CHAID, Breiman's (1984) CART and Quinlan's (1986, 1993) ID3 and C4.5. Among these models CHAID (Kass 1980) and CART (Breiman 1984) like each other in the way that they build decision trees, but CHAID differs in the way that it chooses its splits. Instead of the entropy or Gini metrics for choosing optimal splits, CHAID uses chi square test. Moreover, CHAID differs from the other three methods in tree generating: CART, ID3 and C4.5 generate binary trees whereas CHAID can generate nonbinary trees. "CHAID works all types of categorical and continuous variables, but continuous predictor variables automatically categorized for the purpose of the analysis" (Koyuncugil and Ozgulbas 2007). In developing CHAID models, two groups of variables are used. These are the target variable and the predictor variables that will explain the target variable. Basically CHAID algorithm operates as follows:

- Step 1: The best predictor is chosen from the set of the predictor variables by using a chi square test of independence. The "best" in the previous sentence means the predictor with the most significant p-value.
- Step 2: By using this predictor the data set is divided into two or more subsets.

- Step 3: Each subset coming from Step 2 is divided into new subsets based on the criterion in Step 1, and this is repeated until it can no longer be split into statistically significant subsets.

CHAPTER 4

MODELS

In this part, a specific application of ANN (Multilayer Perceptron) and Decision Tree (CHAID Algorithm) methods to the data which was mentioned in the "Data and Variables" section is presented.

4.1 Artificial Neural Networks Model

The ANN model used in this study consisted of "10" input and "2" output nodes. The input nodes were the financial ratios which were previously mentioned, and the output nodes were the current status of the selected firms in the ISE market lists. To state it more clearly, output nodes indicated that whether the selected firms are still listed in ISE markets (going concern) or permanently de-listed from ISE markets (failed business). For simplicity a dummy variable called "Status" was used for creating the output nodes and within this variable "1" stood for the firms that have failed and "0" stood for the going concerns.

The ANN model used in this study was a multilayer perceptron which is a feedforward artificial neural network model was built by using SPSS Statistics 17.0 software. Multilayer perceptron is a modification of standard linear perceptron in that it uses three or more layers with non-linear activation functions. In this study sigmoid function (see Figure 3), which has an S-shaped curve, was used as the activation function. The form of this function is presented below:

$$\gamma(c) = \frac{1}{1+e^{-c}} \qquad (4.1)$$

"It takes real-valued arguments and transforms them to the range (0, 1)" (SPSS Statistics 17.0 User's Manual). The input data was rescaled before it was used in the network to ensure that all of the scale dependent variables could fit the requirements of activation function. This procedure was done by subtracting the mean of independent variables from the related independent variable and dividing the result by the standard deviation:

$$\frac{x - mean}{s} \qquad (4.2)$$

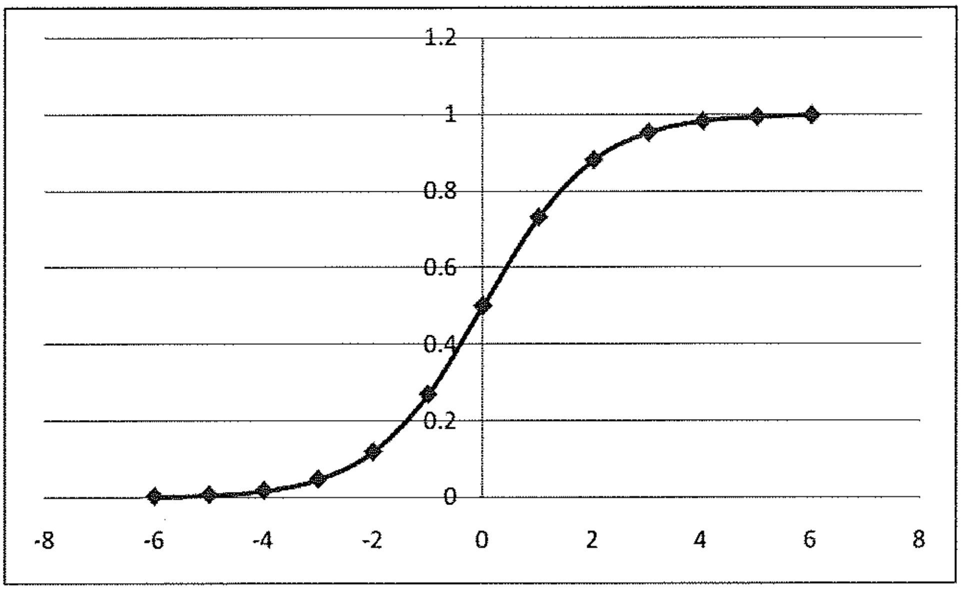

Figure 3 Sample Sigmoid Function Curve

The network was trained using backpropagation algorithm. This algorithm can be summarized in three steps (Berry and Linoff 2004):

1. The network gets a training example and using the existing weights in the network, it calculates the output or outputs.
2. Then the error is calculated by the backpropagation algorithm by taking the difference between the calculated result and the expected (actual result).
3. The error is fed back through the network and the weights are adjusted to minimize the error- hence the name backpropagation because the errors are sent back through the network.

The data set was split into two subsamples. The first subsample that consisted of 19 paired firms (38 firms in total) was used for the training and the second subsample that consisted of 8 paired firms (16 firms in total) was used for the testing of the network. In terms of percentage, that equals: 70% or the data was allocated for the training and 30% for the testing purposes.

The network was built using one hidden layer, and there were "10" hidden units (nodes) in the hidden layer. The number of hidden units was defined by trial and error. Starting with "1" hidden units, the number of hidden units was gradually increased and the optimum results were observed with 10 hidden units. "Sum of squares error" was used as the classification error for the network.

In Table 1, model summary for the ANN is presented. According to this summary, 15.8% of the training sample is misclassified and Sum of Squares Error (SSE) is 5.308, and 18.8% of the testing sample is misclassified and Sum of Squares Error (SSE) is 3.229.

Model Summary		
Training	Sum of Squares Error	5.308
	Percent Incorrect Predictions	15.8%
Testing	Sum of Squares Error	3.229
	Percent Incorrect Predictions	18.8%

Table 1 Model Summary for ANN Model

Table 2 shows the classification results for the ANN model. Looking at the table, it is observed that 84.2% of the data in the training sample and 81.3% of the data in the testing sample is classified correctly.

Classification				
		Predicted		
Sample	Observed	0	1	Percent Correct
Training	0	15	4	78.9%
	1	2	17	89.5%
	Overal Percentage	44.7%	55.3%	84.2%
Testing	0	7	1	87.5%
	1	2	6	75.0%
	Overal Percentage	56.3%	43.8%	81.3%

Table 2 Classification Results for ANN Model

Predicted by observed chart depicted in Figure 4 is a way of visualizing the classification results shown in Table 1. This chart displays clustered box plots of predicted pseudo-probabilities[4] for the combined training and testing samples. "The x axis corresponds to the observed response categories and the legend corresponds to the predicted categories" (SPSS Statistics 17.0 User's Manual). The portion of the box plots above the 0.5 mark on the y axis represents correct predictions shown in Table 1, and the portion below the 0.5 mark represents the incorrect predictions. As it can be seen from the classification results table, the network is almost in equilibrium in predicting the cases "0" and "1". Therefore, there is not a significant distance between the leftmost and the rightmost box plots. However, it indicates that the network does better in classifying the "1" cases (de-listed group), since the rightmost box plot has a greater pseudo-probability than the leftmost box plot.

[4] Pseudo-probabilites (also called as risk neutral probabilities) are not real probabilities, but theoretical probabilities under a series of assumptios that helps to simplfy the calculations.

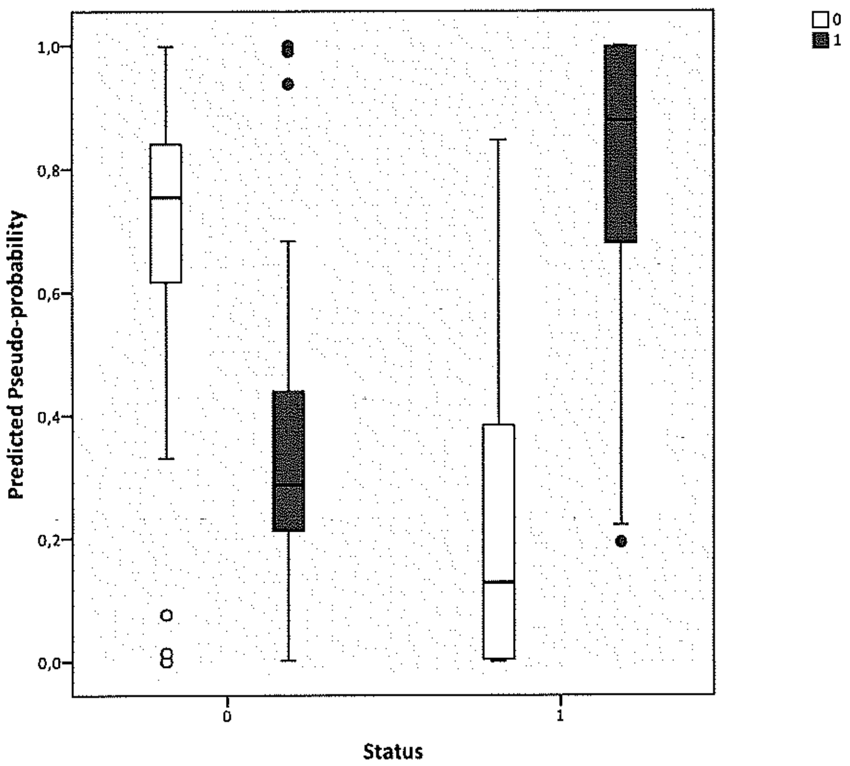

Figure 4 Predicted by Observed Chart for ANN Model

The ROC (Receiver Operating Characteristic) curve in Figure 5 gives a visual display of the tradeoff between sensitivity[5] and specificity[6]. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test. The accuracy of the model is measured by the area under the ROC curve. An area of "1" represents a perfect accuracy and an area of "0.5" represents a worthless model. According to the area measures presented in Table 3, ANN model has a significant accuracy since the area equals "0.835" for both case "0" and case "1".

[5] Sensitivity is the probability that a "positive" case is correctly classified.
[6] Specificity is the probability that a "negative" case is correctly classified.

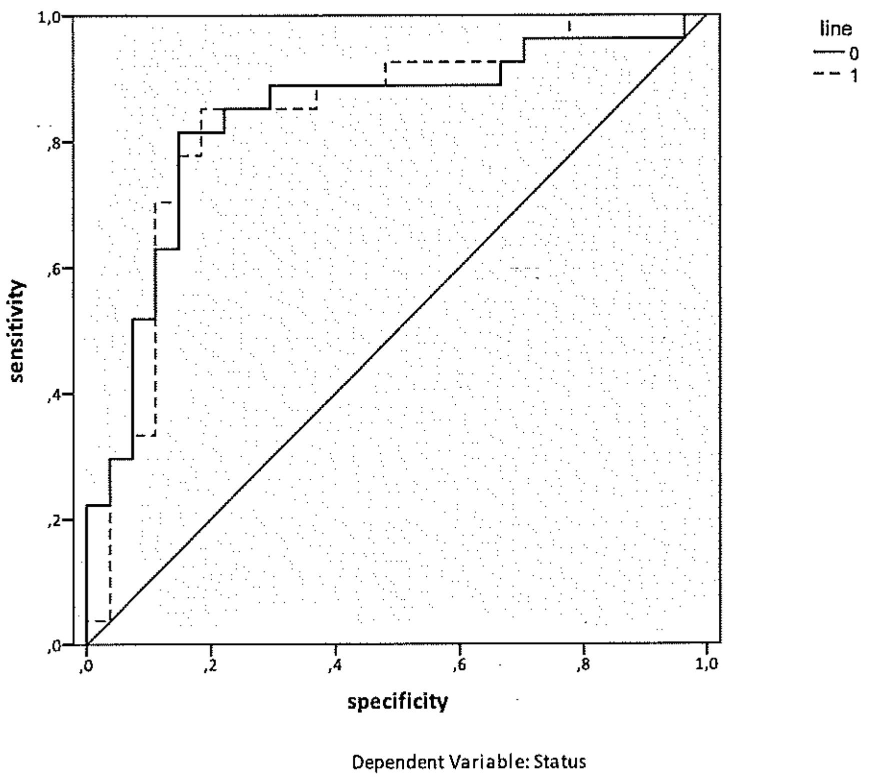

Figure 5 ROC Curve for ANN Model

Area Under the Curve		
		Area
Status	0	0.835
	1	0.835

Table 3 Area under the ROC Curve

"Independent variable importance" is a sensitivity analysis that computes the importance of each predictor in determining the neural network. In Table 4, importance values and the normalized importance of the independent variables in the network are presented. "Normalized importance is simply the importance values divided by the largest importance values and expressed as percentages" (SPSS Statistics 17.0 User's Manual).

	Importance	Normalized Importance
WC/TA	0.085	54.6%
RE/TA	0.027	17.6%
EBIT/TA	0.156	100.0%
Sales/TA	0.093	59.9%
TL/TA	0.089	57.2%
P-in-C/TA	0.079	50.5%
CA/CL	0.099	63.4%
ROA	0.123	79.1%
ROE	0.143	92.0%
QR	0.107	68.6%

Table 4 Independent Variable Importance of ANN model

Figure 6 shows the importance values and the normalized importance of the independent variables in a bar chart. According to this chart, “EBIT / TA” and “ROE” are the two most important predictors in determining the neural network model.

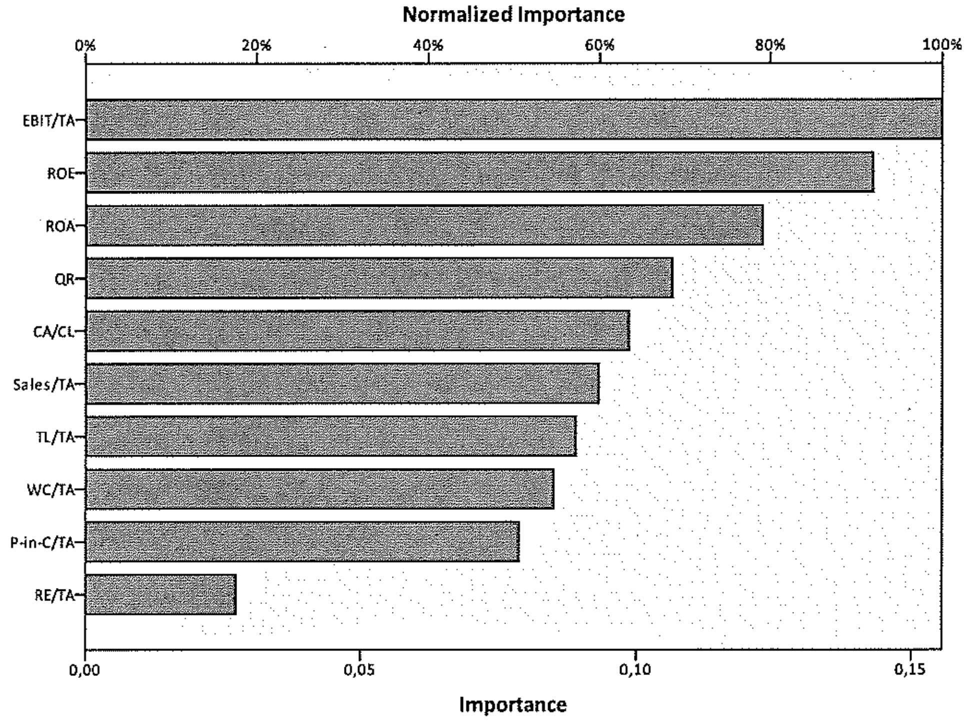

Figure 6 Independent Variable Importance Chart of ANN Model

4.2 Decision Tree Model

The second model used in this study was a decision tree model. CHAID algorithm was used to develop the model. Basically, CHAID algorithm operates using p-values adopted from chi square tests of independence. The decision tree model used in this study was also developed using SPSS Statistics 17.0 software. A short description of the steps of CHAID algorithm was presented in the "Methods" section. A more detailed description of the algorithm that was mentioned in the SPSS Statistics 17.0 User's Manual is as follows (It is also mentioned in the study of Koyuncugil and Ozgulbas (2007)):

1. For each predictor variable X, find the pair of categories that gives the largest p-value with respect to the dependent variable Y. The calculation of the p-value depends on the type of the dependent variable:
   a. If Y is a continuous variable use an F test.
   b. If Y is nominal categorical, the null hypothesis of independence of X and Y is tested. Therefore, a contingency table is formed using classes of Y as columns and categories of predictor X as rows. Use Pearson chi square statistic or likelihood ratio statistic.
   c. If Y is categorical ordinal the null hypothesis of independence of X and Y is tested against the row effects model (with the rows being the categories of X and columns the cases of Y). Use likelihood ratio statistic.
2. For the pair having the largest p-value, compare the p-value to a predetermined alpha level, α_{merge}
   a. If p-value is greater than α_{merge}, merge this pair into a single compound category. Then a new set of categories of X is formed, and you start the process over at step 1.
   b. If the p-value is less than α_{merge}, go on step 3.
3. Compute the adjusted p-value for the merged categories by applying Bonferroni adjustment[7] (generally preferred to use for large samples).
4. Select the predictor X that has the smallest adjusted p-value and compare its adjusted p-value to a predetermined alpha level, α_{split}
   a. If the adjusted p-value is less than or equal to α_{split}, split the node based on the set of categories of X.
   b. If the adjusted p-value is greater than α_{split}, do not split the node. The node is a terminal node.
5. Continue the tree-growing process until the stopping rules are met.

Independent variables of this model were 10 financial ratios and the dependent variable was a dummy variable that represented the current status of the selected firms. The dummy variable was the same variable that was used in the ANN model, and again "1" stood for the businesses that have failed and "0" stood for the going concerns.

5% significance level was selected for both splitting the nodes and merging the categories. Likelihood-ratio method was selected to calculate chi square for determining splitting and category merging. Split-sample validation was used for

[7] "Bonferroni adjustment is to divide the desired bound for the p-value by the number of comparisons being made, in order to get a confidence of 1-p for all comparisons" (Berry and Linoff 2004).

the validation of the model. The same subsamples that were used in the ANN model were used in the training and testing phases of this model, which meant again the training sample had the 70% and the testing sample had 30% of the total data set.

Table 5 shows the classification results for the Decision Tree model. Looking to the table, it is observed that 84.2% of the data in the training sample, and 62.5% of the data in the testing sample is classified correctly.

Classification				
		Predicted		
Sample	Observed	0	1	Percent Correct
Training	0	16	3	84.2%
	1	3	16	84.2%
	Overal Percentage	50.0%	50.0%	84.2%
Testing	0	4	4	50.0%
	1	2	6	75.0%
	Overal Percentage	37.5%	62.5%	62.5%

Table 5 Classification Results for Decision Tree Model

Figure 7 and Figure 8 show the decision tree diagrams for training and testing samples. According to these diagrams, the most important independent variable of the model is "TL / TA" ratio since it stands in the terminal node, and the other important independent variables are "ROA" and "QR" standing in the child nodes. According to the findings the main indicator of a going concern is having a "TL / TA" ratio less than or equal to 0.71, and the reverse is true for the failed businesses.

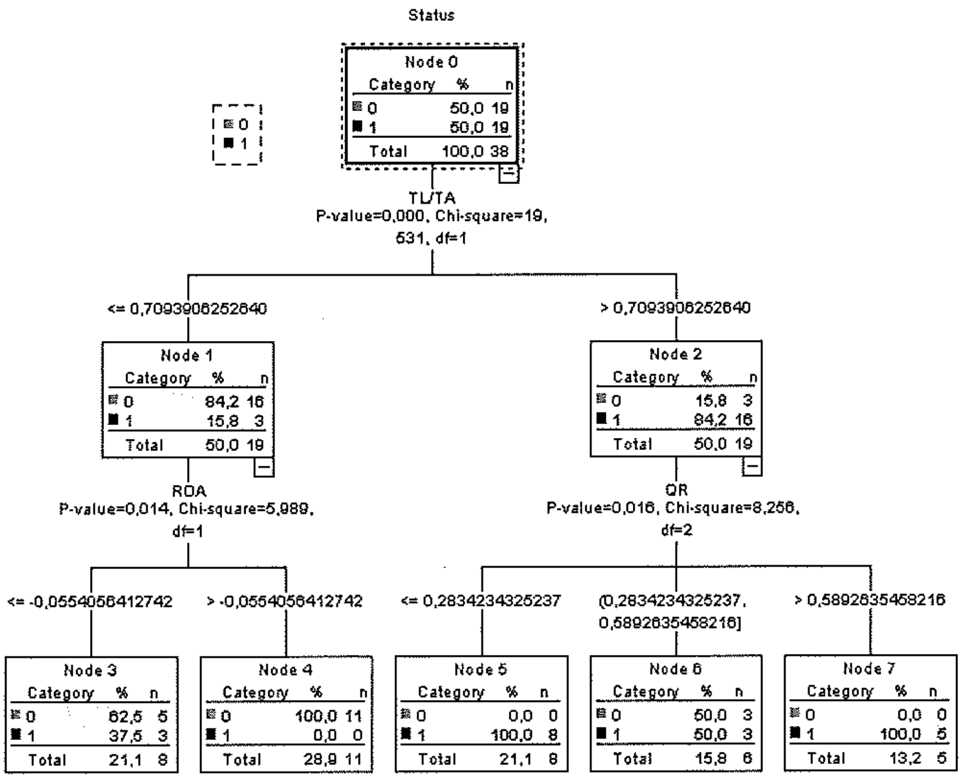

Figure 7 Decision Tree Diagram for Training Sample

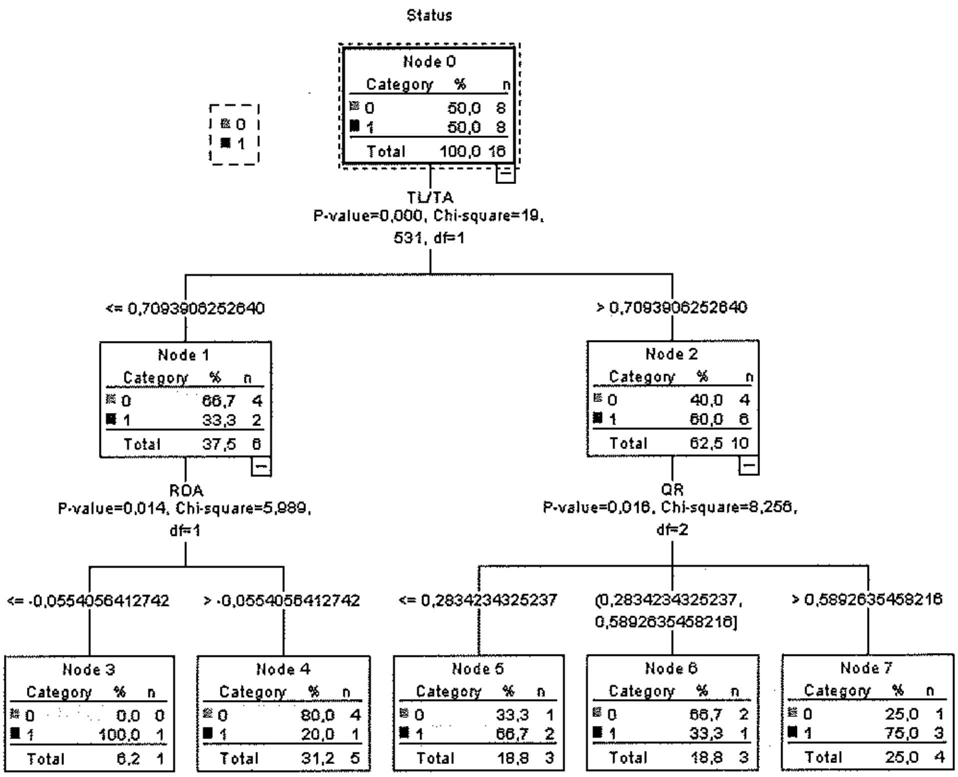

Figure 8 Decision Tree Diagram for Testing Sample

The risk estimation in Table 6 can be thought of as a summary of the classification results in Table 5. The risk estimate of 0.158 for the training sample indicates that the category predicted by the model is wrong for 15.8% of the cases, and similarly the risk estimate of 0.375 for the testing sample indicates that the category predicted by the model is wrong for 37.5% of the cases.

Risk		
Sample	Estimate	Standard Error
Training	0.158	0.059
Test	0.375	0.121

Table 6 Risk Estimate for the Decision Tree Model

CHAPTER 5

RESULTS

In this study, it was tried to figure out whether it was possible to predict business failures in Turkey, especially in ISE markets, using some certain financial ratios. Therefore, two non-parametric models were created and their classification performances were observed.

Table 7 shows the results of the models used in the study. Looking to the percentages of predictions it can be said that "Multilayer Perceptron" algorithm has a better predictive ability in predicting failed businesses (denoted with "1") than its predictive ability in predicting going concerns (denoted with "0") in the training sample. On the other hand, there is a balance between the predicted groups in the second case in training sample where "CHAID" decision three algorithm is used. As it can be seen from the table, overall classification percentages of the two models are equal for the training sample and it is 84.2%. Looking to the literature, it is generally expected that the classification percentages to decrease in the testing samples. This statement holds true in the ANN model with a slight decrease in the classification percentage from 84.2% to 81.3% in the testing sample. However, the rate of the decrease is dramatic for the decision tree model: the classification percentage decreases from 84.2% to 62.5%. Looking to these results it can be said that ANN model outperforms decision tree model with its better performance in the testing sample.

Classification								
		ANN Model			Decision Tree Model		
		Predicted			Predicted		
Sample	Observed	0	1	Percent Correct	0	1	Percent Correct
Training	0	15	4	78.9%	16	3	84.2%
	1	2	17	89.5%	3	16	84.2%
	Overal Percentage	44.7%	55.3%	84.2%	50.0%	50.0%	84.2%
Testing	0	7	1	87.5%	4	4	50.0%
	1	2	6	75.0%	2	6	75.0%
	Overal Percentage	56.3%	43.8%	81.3%	37.5%	62.5%	62.5%

Table 7 Classification Summary of the Models Used In the Study

Although the two models have approximately the same classification abilities, the methods that they use, indicate different ratios as the most important performance indicators except one. The "Multilayer Perceptron" algorithm indicates "EBIT / TA", "ROE" and "ROA" ratios as the most important predictors in determining the ANN model. On the other hand, "CHAID" algorithm presents "TL / TA", "ROA" and "QR" ratios as the most important predictors in determining the decision tree model. "ROA" is the only ratio that is selected by both models as one of the most important predictors.

CHAPTER 6

CONCLUSIONS AND DISCUSSION

This study explored the potential usefulness of financial ratios in predicting the business failures of the firms whose stocks were traded in the ISE markets. Comparative analysis of the classification results indicated that the ANN and decision tree models have similar performances in the training sample, but the ANN model outperformed the decision tree model in the testing sample.

The general trend in the business failure literature is to compare the results of the parametric models with the non-parametric ones, and the results of these comparative analyses indicate that the models constructed by non-parametric techniques outperform the models constructed by the parametric techniques most of the time. However, studies which compare non-parametric techniques are relatively rare. Therefore, in this study it was aimed to compare two non-parametric techniques to see whether there was a significant difference in their performance of predicting the business failures. The results of this analysis were as expected: the prediction performances of these two techniques were almost equal. In the literature, studies with similar results are common. Koh and Low (2004) compared three models constructed using logistic regression, ANN and decision tree techniques and found that all three models gave adequate results. Among these results the best one belonged to decision tree model. Likewise, Olmeda and Fernandez (1997) showed in their study about the hybrid classifiers that ANN and decision tree models had almost the same predictive ability when they were thought as single models (In this study ANN model was slightly more successful than the decision tree model).

There are a number of limitations about the study. One of them is about the difference between the classification percentages of the ANN and the decision tree models in the testing sample. This difference might be the effect of the sample size, because previous studies made in this topic mostly had larger sample sizes. Therefore, this study was a case specific one and a generalization might only be possible using larger sample sizes and making the all available tests to measure the method related effects.

Another limitation of the study might be the number of the financial ratios used to develop the models. Only ten selected financial ratios were used in the study as a result of the availability of the data and the computability of the ratios. Including more independent variables and implementing different models to select the most appropriate and predictive ones might be a better approach.

Since an ANN model is a "black box", it is difficult to interpret the exact relationships that the technique builds in the hidden layer. Because of this it sometimes becomes difficult to compare the power of ANN to the other techniques. Moreover, the necessary test statistics for effect assessment are not available yet. Therefore, ANN and the other data mining techniques do not replace traditional statistical methods.

In the study, the classification percentages were computed in a simple manner ignoring the misclassification costs that might occur in the practice. Determining the potential breakpoints and classification costs might be a point that increases the power of the models in practice.

Despite the limitations stated above, it can be said that data mining techniques like ANNs and decision trees are powerful tools as alternatives for commonly used statistic techniques in predicting business failures. For instance, MDA models can only be used if the data satisfy some certain statistical assumptions (multivariate normality and equality of variance-covariance matrices). On the other hand, ANN and decision trees do not depend on such assumptions. Therefore, they are generally more flexible than parametric techniques.

This study suggests that some key financial ratios such as "EBIT/TA", "ROE", "TL/TA", "ROA" are beneficial for managers (especially managers in Turkey) to notice and handle some potential problems about their firms. These ratios will also be significant indicators of performance for the other stakeholders mentioned in the "Introduction" part. Investors who plan to invest in the stocks of the firms, banks which loan credits, stock market specialists who make suggestions to their clients and etc. may support their analysis from some aspects using these suggestions. Despite the all benefits, financial ratio analysis is not a sufficient way alone in decision making, and it has to be supported by some other critical analyses such as sector and company analyses. However, it is beneficial for the decision makers, since it gives some clues about the potential threats and the opportunities.

REFERENCES

Acar-Boyacioglu, M., Kara Y., Baykan Ömer K., *Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey*, Expert Systems with Applications, Vol. 36, Issue 2, pp. 3355-3366, 2009

Altman, E., Margaine, M., Schlosser, M., Vernimmen, P., *Financial and Statistical Analysis for Commercial Loan Evaluation: A French Experience*, The Journal of Financial and Quantitative Analysis, Vol.9, No.2, pp. 195-211, 1974

Altman, E., Saunders, A., Credit risk measurement: Developments over the last 20 years. Journal of Banking and Finance 21(1998), pp. 1721-1742, 1998

Altman, Edward I. , Haldeman, Robert G., Narayanan P., *ZETA analysis A new model to identify bankruptcy risk of corporations*, Journal of Banking & Finance, Volume 1, Issue 1, pp. 29-54, 1977

Altman, Edward I., *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*, The Journal of Finance, pp. 589-609, 1968

Beaver, W.H., *Financial Ratios as Predictors of Failure*, Journal of Accounting Research, Vol.4, Empirical Research in Accounting: Selected Studies, pp. 71-111 1966

Berry, Michael J.A., Linoff, Gordon S., *Data mining techniques: for marketing, sales, and customer relationship management*", 2nd Ed., Wiley Pub., 2004

Blum, M., *Failing Company Discriminant Analysis*, Journal of Accounting Research, Vol. 12 Issue 1, pp. 1-25, 1974

Boritz, J.E., Kennedy, D.B., De Miranda e Albuquerque, A., *Predicting Corporate Failure Using a Neural Network Approach*, International Journal of Intelligent Systems in Accounting Finance & Management, Vol. 4 Issue 2, pp. 95-111, 1995

Breiman, L., Friedman, J.H., Stone, C.J., *Classification and Regression Trees*, London: Chapman and Hall

Brigham, Eugene F., Ehrhardt, Michael C., *Financial Management Theory and Practice*, 12th Ed., Thomson South-Western, 2008

Brockett, Patrick L., Golden, Linda L., Jang, J., Yang, C., *A Comparison of Neural Network, Statistical Methods, and Variable Choice for Life Insurers' Financial Distress Prediction*, Journal of Risk & Insurance, Vol. 73 Issue 3, pp. 397-419, 2006

Carter, C., Catlett, J., *Assessing credit card applications using machine learning*, IEEE Expert Intelligent Systems & Their Applications, Vol. 2, pp. 71-79, 1987

Charitou, A., Neophytou, E., Charalambous, C., *Predicting corporate failure: empirical evidence for the UK*, European Accounting Review, Vol. 13 Issue 3, pp. 465-497, 2004

Chase, Jr., Charles W., *Selecting the appropriate forecasting method*, Journal of Business Forecasting Methods & Systems, Vol. 16 Issue 3, p2, 1997

Chen, J., Marshall, Ben R., Zhang, J., Ganesh, S., *Financial Distress Prediction in China*, Review of Pacific Basin Financial Markets & Policies, Vol. 9 Issue 2, pp. 317-336, 2006

Chen, W., Du, Y., *Using neural networks and data mining techniques for the financial distress prediction model*, Expert Systems with Applications, Part 2, Vol. 36 Issue 2, pp. 4075-4086, 2009

Chung, Hyung-Min M., Tam, K. Y., *A Comparative Analysis of Inductive-Learning Algorithms*, International Journal of Intelligent Systems in Accounting Finance & Management, Vol. 2 Issue 1, pp. 3-18, 1993

Deakin, Edward B., *A Discriminant Analysis of Predictors of Business Failure*, Journal of Accounting Research, Vol. 10 Issue 1, pp. 167-179, 1972

Doumpos, M., Zopounidis, C., *Model combination for credit risk assessment: A stacked generalization approach*, Annals of Operations Research, Vol. 151 Issue 1, pp. 289-306, 2007

Edmister, R., *An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction*, The Journal of Financial and Quantitative Analysis, Vol. 7, No. 2, Supplement: Outlook for the Securities Industry, pp. 1477-1493, 1972

Fausett L., *Fundementals of Neural Networks: Architecture, Algorithms and Applications*, Prentice Hall, 1994

Fletcher, D., Goss, E., *Forecasting with neural networks: An application using bankruptcy data,* Information & Management, pp. 159–167, 1993

Freeman, N., *Forecasting methods facilitate planning accuracy*, Ophthalmology Times, Vol. 28 Issue 6, p55, 2003

Friedman, J.H., *A Recursive Partitioning Decision Rule for Nonparametric Classification*, IEEE Transactions on Computers, pp. 404-409, 1977

Frydman, H., Altman Edward I., Kao D., *Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress*, Journal of Finance, Vol. 40 Issue 1, pp. 269-291, 1985

Gentry, James A., Newbold, P., Whitford, David T., *Classifying Bankrupt Firms with Funds Flow Components*, Journal of Accounting Research, Vol. 23 Issue 1, pp. 146-160, 1985

Hosal-Akman, N., Simga-Mugan, F.N. Can, *Principles of Financial Accounting*, 3rd Ed., Gazi Kitabevi, 2007

Kass, G.V., *An Exploratory Technique for Investigating Large Quantities of Categorical Data*, Applied Statistics, Vol. 29, No. 2, pp. 119-127, 1980

Kiviluoto, K., *Predicting bankruptcies with the self-organizing map*, Neurocomputing, Vol. 21, Issues 1-3, pp. 191-201, 1998

Koh, Hian C., Low, Chan K., *Going concern prediction using data mining techniques*, Managerial Auditing Journal, Vol. 19 Issue 3, pp. 462-476, 2004

Koyuncugil, Ali S., Ozgulbas, N., *Detecting Financial Early Warning Signs in Istanbul Stock Exchange by Data Mining*, International Journal of Business Research, Vol. 7 Issue 3, pp. 188-193, 2007

Lin, Feng Y., McClean, S., *A data mining approach to the prediction of corporate failure*, Knowledge-Based Systems, Vol. 14, Issues 3-4, pp. 189-195, 2001

Messier, W.F., Hansen, J., *Inducing Rules for Expert System Development: An Example Using Default and Bankruptcy Data*, Management Science, Vol. 34, No. 12, pp. 1403-1415, 1988

Odom, M.D., Sharda, R., *A neural network model for bankruptcy prediction*, In Proceedings of the IEEE International Joint Conference on Neural Networks, pp. 163-168, 1990

Ohlson, James A., *Financial Ratios and the Probabilistic Prediction of Bankruptcy*, Journal of Accounting Research, Vol. 18 Issue 1, pp. 109-131, 1980

Olmeda, I., Fernandez, E., *Hybrid classifiers for financial multicriteria decision making: The case of bankruptcy prediction*, Computational Economics, Vol. 10, Issue. 4, pp. 317-35, 1997

Quinlan, J.R., *Induction of Decision Trees*, Machine Learning, Volume 1, Number 1, 1986

Salchenberger, Linda M., Mine Cinar, E., Lash, Nicholas A., *Neural Networks: A new Tool for Predicting Thrift Failures*, Decision Sciences, Vol. 23 Issue 4, pp. 899-916, 1992

Shirata, C.Y., *Financial Ratios as Predictors of Bankruptcy in Japan: An Empirical Research*, Proceedings of The Second Asian Pacific Interdisciplinary Research in Accounting Conference, pp. 437-445, 1998

SPSS Statistics 17.0 User's Manual

Tam, Kar Y., Kiang, Melody Y., *Managerial Applications of Neural Networks: The Case of Bank Failure Predictions*, Management Science, Vol. 38 Issue 7, pp. 926-947, 1992

Turetken, O., *Prediction of financial performance in Istanbul Stock Exchange: A comparative analysis*, Master's Thesis, 1999

Wilson, Rick L., Sharda, R., *Bankruptcy prediction using neural networks*, Decision Support Systems, Vol. 11, pp. 545-557, 1994

Zmijewski, Mark E., *Methodological Issues Related to the Estimation of Financial Distress Prediction Models*, Journal of Accounting Research, Supplement, Vol. 22, pp. 59-82, 1984

APPENDIX A

COMPANIES WITH STOCKS DE-LISTED FROM THE ISE MARKETS PERMANENTLY (*) (AS FROM YEAR 2000)

CODE	STOCK	DE-LISTING DATE
MEGES	MEGES BOYA	12/24/2008
ARAT	ARAT TEKSTİL	11/17/2008
ALFA	ALFA MENKUL DEĞ.	11/17/2008
EGIYM	EGESER GİYİM	11/17/2008
EGHOL	EGS HOLDİNG	11/17/2008
MEDYA	MEDYA HOLDİNG	11/17/2008
SABAH	SABAH YAYINCILIK	11/17/2008
SAPAZ	SABAH PAZARLAMA	11/17/2008
LIOYS	LİO YAĞ	9/19/2008
ABANA	ABANA ELEKTROMEKANİK	5/1/2008
RAKSE	RAKS ELEKTRONİK	6/15/2007
RKSEV	RAKS EV ALETLERİ	6/15/2007
UNTAR	ÜNAL TARIM	2/7/2007
KOTKS	KONİTEKS	2/7/2007

GORBN	GORBON IŞIL	12/22/2004
IKTFN	İKTİSAT FİNANSAL KİRALAMA	5/13/2004
FACF	FACTO FİNANS	5/13/2004
METAS	METAŞ	10/8/2003
CUKEL	ÇUKUROVA ELEKTRİK	6/18/2003
KEPEZ	KEPEZ ELEKTRİK	6/18/2003
SEZGD	SEZGİNLER GIDA	11/18/2002
AKTAS	AKTAŞ ELEKTRİK	8/16/2002
EGDIS	EGS DIŞ TİCARET	8/16/2002
GUMUS	GÜMÜŞSUYU HALI	8/16/2002
KOYTS	KÖYTAŞ TEKSTİL	8/16/2002
SOKSA	SÖKSA	8/16/2002
MDRNU	MUDURNU TAVUKÇULUK	5/7/2002
TPBNK	TOPRAKBANK	1/31/2002
EMEK	EMEK SİGORTA	1/30/2002
APEKS	APEKS DIŞ TİCARET	1/15/2002
INMDY	INTERMEDYA	1/15/2002
IHFIN	İHLAS FİNANS	11/7/2001
DEMIR	DEMİRBANK	9/20/2001
SVGSH	SEVGİ SAĞLIK HİZM.	7/9/2001

ESBNK	ESBANK	4/3/2001
YABNK	YAŞARBANK	4/3/2001
EMSAN	EMSAN BEŞYILDIZ	10/18/2000
EMPAS	EMSAN PAS.ÇELİK	10/18/2000

(*) Stocks delisted because of acquisitions are not included.

Source: www.imkb.gov.tr

APPENDIX B

PAIRED FIRMS AND ASSET SIZES

Code	Firm(*)	Asset Size	Code	Paired Firm(**)	Asset Size	Year(***)
MEGES	Meges Boya	24,427,493 TL	CBSBO	ÇBS Boya	41,551,835 TL	2006
ARAT	Arat Tekstil	30,036,122 TL	MTEKS	Metemteks	29,629,085 TL	2006
EGIYM	Egeser Giym	44,341,856 TL	BISAS	Bisaş Tekstil	46,340,337 TL	2006
MEDYA	Medya Holding	8,159,331 TL	VKFRS	Vakıf Girişim	4,263,669 TL	2006
SABAH	Sabah Yayıncılık	63,993,286 TL	BAKAB	Bak Ambalaj	72,403,051 TL	2006
LIOYS	Lio Yağ	159,704,290 TL	TUKAS	Tukaş Konserve	159,654,098 TL	2006
ABANA	Abana Elektromekanik	3,092,323 TL	EMKEL	Emek Elektrik	11,961,693 TL	2006
RAKSE	Raks Elektronik	129,355,688 TL	MUTLU	Mutlu Akü	133,387,296 TL	2005
RKSEV	Raks Ev Aletleri	15,716,401 TL	GEREL	Gersan Elektrik	15,090,846 TL	2005
UNTAR	Ünal Tarım	24,094,676 TL	FRIGO	Frigo Pak Gıda	28,726,893 TL	2005
KOTKS	Koniteks	17,339,797 TL	LUKSK	Lüks Kadife	19,522,516 TL	2005
GORBN	Gorbon Işıl	2,118,647 TL	DOGUB	Doğusan	8,521,456 TL	2002

IKTFN	İktisat Finansal Kiralama	45,420,304 TL	TEKFK	Tekstil Fin. Kir.	35,359,986 TL	2002
FACF	Facto Finans	12,697,105 TL	CRDFA	Creditwest Fact.	26,571,339 TL	2002
METAS	Metaş	4,544,068 TL	BURCE	Burçelik	5,975,466 TL	2001
CUKEL	Çukurova Elektrik	889,146,814 TL	AKENR	Ak Enerji	353,297,364 TL	2001
KEPEZ	Kepez Elektrik	74,720,711 TL	ZOREN	Zorlu Enerji	151,243,208 TL	2001
SEZGD	Sezginler Gıda	122,142,884 TL	PNSUT	Pınar Süt	78,101,984 TL	2000
AKTAS	Aktaş Elektrik	82,024,520 TL	AYEN	Ayen Enerji	117,463,283 TL	2000
EGDIS	EGS Dış Ticaret	158,078,549 TL	KORDS	Kordsa	122,962,106 TL	2000
GUMUS	Gümüşsuyu Halı	44,026,851 TL	ARSAN	Arsan Tekstil	49,315,711 TL	2000
KOYTS	Köytaş Tekstil	3,179,588 TL	CEYLN	Ceylan Giyim	6,176,708 TL	2000
SOKSA	Söksa	2,126,178 TL	BYSAN	Boyasan Tekstil	4,352,837 TL	2000
MDRNU	Mudurnu Tavukçuluk	33,440,372 TL	PENGD	Penguen Gıda	33,189,387 TL	2000
SVGSH	Sevgi Sağlık Hizm.	7,373,303 TL	ACIBD	Acıbadem Sağ.	7,668,332 TL	1999
EMSAN	Emsan Beşyıldız	2,190,295 TL	FMIZP	İzmit Piston	2,784,963 TL	1998
EMPAS	Emsan Pas. Çelik	3,971,619 TL	EMNIS	Eminiş Ambalaj	3,870,536 TL	1998

(*) Failed businesses

(**) Going Concerns

(***) Year that the asset size belongs to

APPENDIX C

THE DATA SET

	Firm Code	WC/TA	RE/TA	EBIT/TA	Sales/TA	TL/TA	P-in-C/TA	CA/CL	ROA	ROE	QR	Status
1	MEGES	-0.057801142	0	-0.143700911	0.800827166	0.89208078	0.640380902	0.93384962	-0.242248028	-2.24471625	0.836354742	1
2	ARAT	-0.122161543	0	-0.034957975	0.647193702	0.87708836	1.331729842	0.843887553	-0.036397009	-0.29612337	0.396315785	1
3	EGIYM	-0.378174405	0.03105161	-0.000353255	0	1.152184315	0.130238121	0.671398825	0.000713073	-0.004685591	0.671398825	1
4	MEDYA	-230.7212827	0.615328634	-0.765625393	0	235.0362517	3.799331097	0.002867688	-6.828015189	0.029175032	0.002867688	1
5	SABAH	-26.98612967	0.032904827	-1.082197342	0	27.73304823	0.515679098	0.022020813	-1.00269533	0.037507707	0.022020813	1
6	LIOYS	0.56843684	0.00311815	-0.02812508	0.672503149	0.710005786	0.313078628	3.104425901	-0.146177382	-0.504069995	1.133630154	1
7	ABANA	0.084811968	0.032995583	-0.092675636	0.037715012	0.057686083	0.864398706	2.680405964	-0.176165944	-0.186950379	1.754300872	1
8	RAKSE	-0.668956583	0.024855722	-0.007491592	0.023032833	2.151787774	0.042079325	0.057861737	-0.21295195	0.184888184	0.040480452	1
9	RKSEV	-1.270731384	0.01872407	5.43381E-05	0.207615217	3.666204432	0.01908834	0.112080787	-0.338788823	0.127067834	0.037551202	1
10	UNTAR	0.53877562	0.000500982	-0.101917121	0.353903493	0.425519853	1.245088334	2.608082116	-0.101244275	-0.176236334	0.562698128	1
11	KOTKS	0.056296045	0	-0.246451097	1.005623134	0.645474108	0.579276735	1.095919084	-0.176464638	-0.497748238	0.283423433	1
12	GORBN	-0.721676145	0.024081879	-0.16355863	0.49388926	1.970204097	0.135935812	0.559156012	-0.968705027	0.998454892	0.413348634	1
13	IKTFN	-2.396062871	0.123633981	0.293600303	0.645914237	2.898507945	0.132099512	0.07722942	-1.014241913	0.53423106	0.07722942	1
14	FACF	-3.11447326	0.020350623	-0.176984202	0.64713988	4.09262198	0.118137166	0.238889553	-1.067317314	0.345117289	0.238889553	1

15	METAS	-4.572094212	0.015848574	-0.233759706	0.005883495	4.75895123	0.155147326	0.037163273	-1.962581326	0.522108749	0.01540161	1
16	CUKEL	-0.432956397	2.44864E-05	0.144158483	0.537434474	1.301476828	0.000562337	0.183607438	-0.239810761	0.795453377	0.171826937	1
17	KEPEZ	0.330641072	0.000320915	0.422562387	1.753472849	0.709043106	0.00160598	1.596814355	0.266432783	0.915712219	1.554929357	1
18	SEZGD	0.141533714	0.016298821	0.080026291	1.787659746	0.79962032	0.03946198	1.180823205	0.028785926	0.143656911	0.596370807	1
19	AKTAS	0.262447973	0.005585744	0.222221904	2.458766305	0.802831044	0.012191476	1.566477838	0.158276879	0.802747459	1.461559924	1
20	EGDIS	-0.0536376	2.51014E-05	-0.020954595	2.932217014	0.895315796	0.088563566	0.939976051	0.002183522	0.020858181	0.925351783	1
21	GUMUS	0.013310037	4.08841E-07	0.052557359	0.780018062	0.709142337	0.255525884	1.025029385	-0.151630172	-0.521320878	0.589263546	1
22	KOYTS	-0.365158002	0.267416093	-0.048932755	0	3.350024594	0.314506156	0.506925664	-0.331482255	0.141054802	0.461626435	1
23	SOKSA	-0.528172147	0.113501786	-0.325533422	0.088279062	1.496301815	0.228579169	0.640159293	-0.340499714	0.686073884	0.494362969	1
24	MDRNU	-0.689156269	0.084795797	-0.562011631	1.038730729	1.404779438	0.098683113	0.433318828	-0.688983334	1.702120387	0.266284314	1
25	SVGSH	-1.151810525	0	-0.179792421	0.504265727	1.845068486	0.383933903	0.149909677	-1.224804948	1.449355844	0.133544085	1
26	EMSAN	-0.194813941	0.007523188	0.11014909	0.740900655	1.17395328	0.076163257	0.818502224	-0.145061738	0.83391206	0.768241335	1
27	EMPAS	0.238631148	0.006257901	0.073254005	0.624939855	0.929687868	0.197589447	1.321804642	0.097522446	1.386993157	1.269691245	1
28	CBSBO	-0.944246602	0	-0.049466119	0.503842995	2.144003821	0.473939358	0.40788045	-0.300773672	0.262913171	0.390920509	0
29	MTEKS	-0.034305818	0	-0.069266027	0.537011453	0.64946204	0.565096762	0.895359961	-0.187576093	-0.535109216	0.665965669	0
30	BISAS	-0.271287949	0.002618777	-0.187308931	0.976406667	0.953296542	0.135950673	0.660547595	-0.239830345	-5.135173136	0.440704014	0
31	VKFRS	0.565171921	0.011945346	-0.08609158	0	0.007006172	0.211085804	128.1948271	-1.256956391	-1.265824979	128.1948271	0
32	BAKAB	0.185232263	0.009922759	0.158739374	1.176999475	0.326126685	0.124304154	1.742475953	0.130117997	0.193089702	1.29628483	0
33	TUKAS	0.260906269	0.016786484	0	0	0.755663246	0.390531786	1.446124722	-0.102081802	-0.417791431	0.58660497	0
34	EMKEL	-0.139139167	0.001680949	-0.298334191	1.002997318	1.164551623	1.836630316	0.758669905	-0.241642801	1.468492356	0.407608378	0
35	MUTLU	0.191516829	0.047879102	0.016156351	1.071885369	0.349516561	0.425077962	2.015712182	0.008331476	0.012808129	1.132091824	0
36	GEREL	0.567435451	0	0	1.543951081	0.178525975	0.893389277	4.437698871	-0.055405641	-0.067446614	3.131627333	0
37	FRIGO	0.218592662	0	-0.026197438	0.815010172	0.623440273	0.417727041	1.595801553	0.006078764	0.016142894	0.583818229	0
38	LUKSK	0.334988072	0	0.066661029	0.635676109	0.372660163	0.152132031	10.15588416	0.082348966	0.131266916	7.393512854	0
39	DOGUB	-0.413518065	0.00013155	-0.047514063	0.191725334	0.709390625	1.009672291	0.333236201	-0.279003612	-0.960064044	0.176017154	0

40	TEKFK	-0.356511991	0.012334479	0.106456971	0.484446317	0.662896841	0.282805542	0.117886214	0.04393073	0.130318358	0.117886214	0
41	CRDFA	0.599326026	0.000740158	0.009968899	0.076024848	0.400961766	0.30686662	2.501751568	0.079035422	0.131937192	2.501751568	0
42	BURCE	0.190510497	0.003876852	0.056754235	1.427223249	0.749629234	0.108443425	1.458769025	0.005966731	0.023831581	0.928040593	0
43	AKENR	0.555604802	0.135646871	-0.010101205	0.499576156	0.310906718	0.038211437	3.639632202	0.283646532	0.411622837	3.594221019	0
44	ZOREN	0.194686389	0.053243601	0.019899393	0.608043668	0.436830281	0.066118672	1.870297121	0.08114777	0.144091145	1.844834075	0
45	PNSUT	0.084313748	0.015954601	0.112690108	1.498228176	0.705142689	0.066154376	1.132290977	0.05240018	0.177713686	0.850154671	0
46	AYEN	0.306415027	0	0.163079709	0.227361677	0.695326207	0.037713913	2.28504135	0.171279999	0.562175031	2.284977942	0
47	KORDS	0.227233925	0.035491137	0.099562942	0.788094675	0.50331616	0.195862293	1.859563792	0.074839756	0.150678862	1.062362667	0
48	ARSAN	0.145408448	0.055067076	0.036202966	0.927540232	0.424363627	0.170331114	1.391065663	0.001567675	0.002723377	0.988710615	0
49	CEYLN	0.229922477	0.03658826	0.117485236	1.224756618	0.585419936	0.116809796	1.473606175	0.048529087	0.117056007	0.417766785	0
50	BYSAN	-2.541542677	0.004635827	-0.003322661	0	2.634538348	0.172301421	0.004741344	-0.84858289	0.519157529	0.004741344	0
51	PENGD	0.165879472	0.037775057	0.122339711	0.630595889	0.730424367	0.125341272	1.259857197	0.004389958	0.016284697	0.319713903	0
52	ACIBD	0.011250687	0.084160936	0.443439851	1.335146679	0.499022864	0.011736581	1.030725924	0.26371263	0.526396539	0.936284861	0
53	FMIZP	0.270669327	4.2471E-05	0.109636509	1.928048126	0.374602972	0.078461307	2.184724318	0.043579377	0.069682737	1.578037368	0
54	EMNIS	0.066894869	0.020617558	0.162937381	0.977657875	0.616996199	0.090478425	1.208883977	-0.061845698	-0.161475415	0.669069449	0