Bankruptcy Prediction

using

Classification and Regression Trees

Bachelor Thesis Informatics & Economics

Faculty of Economics
Erasmus University Rotterdam

M.A. Sprengers

E-mail: masprengers@hotmail.com

21st August 2005

Abstract

This report is devoted to Bankruptcy prediction using Classification and Regression Trees (CART). Bankruptcy prediction has become increasingly important over the last few decades. The number of corporate bankruptcies has been growing ever since the economical depression of 1930. This situation brings forth many concerns for company shareholders, creditors, employees, customers and even national economies.

In this thesis a relatively new technique for bankruptcy prediction, CART, constructs an accurate classification model for bankruptcy prediction. This model is benchmarked with the Z-score model introduced by Altman (1968),the most common used classification model for this problem, which is based on discriminant analysis.

The data set used for this report consists of 122 Dutch companies. All of them were or still are listed on the Amsterdam Stock Exchange (AEX). 61 companies went bankrupt somewhere in between the period 1945-1999. The other 61 companies are "matched" companies, from the same industry group, same size and same period of listing on the AEX.

Keywords: Bankruptcy prediction, Classification and Regression Trees (CART), Discriminant Analysis, Artificial Intelligence

Acknowledgements

I hereby want to thank my supervisor Dr.Ir. Jan van den Berg for his assistance and helpful remarks during the writing of this thesis.

I am also very grateful to Drs. Frans Caljé for providing me with a suitable data set. Without this kind gesture, probably I would not have been as satisfied with my thesis as I am now.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This Chapter gives an introduction to this report and explains the motivation for choosing "Bankruptcy Prediction using Classification and Regression Trees" as the subject of this thesis. In section 1.1, the motivation and origin for the topic of this report is discussed. After formulating the motivation, the goal of this report will be described in section 1.2. In section 1.3, the methodologies used in this report will be described shortly. Finally, section 1.4 gives an overview of this report. In this overview, every chapter will be discussed briefly.

1.1 Motivation

Since the depression in 1930 corporate bankruptcy reached numbers never seen before. Corporate bankruptcy is an event which has a big impact on management, shareholders, employees, creditors, customers and other stakeholders. It will cause financial losses to most of the aforementioned parties. These events also have a negative influence, both socially and economically, on a nation[1]. For this reason it would be very helpful if we could say something about the probability by which these events happen. Therefore, accurate prediction of bankruptcy has become an important issue in the financial world. A major focus of the Basel II regulations to minimize credit risk is another reason why timely identification of corporate bankruptcy is desirable.

There are various factors that are related to corporate bankruptcy. High interest rates and heavy debt burden are factors which definitely have negative influence on the financial situation of a firm. Industry-specific characteristics and government regulations can also contribute to financial distress within any firm. Besides the factors described before, studies[2] pointed out that small and young private companies are relatively more vulnerable to financial distress than large and well-established public companies.

In the past few decades extensive research has been done on bankruptcy events and especially on bankruptcy prediction. Several bankruptcy prediction models have been introduced. It is essential to discover how to identify potential bankrupt corporations. Beaver (1966) introduced one of the classical works about ratio analysis for bankruptcy prediction. His model which was based on univariate analysis, formed a starting point

[1] Altman (1968)

[2] Dun and Bradstreet (1980)

for many other researchers. Multiple discriminant analysis (MDA) used by Altman (1968) are still being used as a standard tool for bankruptcy prediction, despite its limitations. Models based on newer techniques, such as recursive partitioning, artificial neural networks and genetic algorithms are more sophisticated. In general, most of these techniques outperform MDA. All of these techniques assume normal economic conditions.

1.2 Goal

The objective of this report is: "Constructing an accurate bankruptcy prediction model." The technique used for the construction of this prediction model is based on Classification and Regression Trees (CART).

For many years MDA models are used as standard tools for the bankruptcy prediction problem, despite the fact that many techniques outperform MDA in prediction corporate bankruptcy. This is mainly due to two facts. Firstly, because MDA models are easy to apply for this bankruptcy prediction problem. Secondly, extensive research conducted on the application of MDA models for the bankruptcy prediction problem makes these models comprehensible and common used.

This report compares the popular Z-score model [3], which makes use of MDA, with CART. CART is classification technique which uses historical data to construct decision-trees (supervised learning). These decision-trees can be used to classify new data. Before building these trees it is necessary to know the number of classes a priori. The bankruptcy prediction problem has two classes, namely bankrupt and non-bankrupt. The CART technique has the ability to select the input variables which are most important in determining the variable to be explained.

The goal of this report is not only to compare the Z-score model with CART, but also to get a better feeling for the input variables which are of most importance in determining the dependent variable to be explained. The knowledge that, hopefully, is acquired from this report will help future research. For example, it could contribute to structure identification for adaptive neuro-fuzzy inference systems (ANFIS).

1.3 Methodology

In this report a decision tree is constructed to perform a regression and classification for bankruptcy prediction. The building and testing of this tree, which is constructed by the CART algorithm, consists of three stages. In the first stage, the maximum tree has to be build to classify all the examples in the training set, growing the tree. During the second stage, this maximum tree has to pruned, pruning the tree. Reason that this maximum tree is based on the training data, which will bring a high degree of bias with it. In other words, overfitting and overspecializing toward the training set will be the result, instead of increasing the accuracy toward the test set. Finally, this pruned tree has to classify new data. To test the accuracy the tree is feeded with test data, of which, a priori, is known to which class it belongs.

[3] Introduced by Altman (1968)

This report will also describe other techniques to predict bankruptcy comprehensively. From these techniques, the Z-score model is the most popular one[4]. That's why this technique will act as a benchmark technique for the decision tree based on the CART algorithm.

The data set used in this report consists of 61 bankrupt and 61 non-bankrupt Dutch companies. Each bankrupt company has a comparable non-bankrupt company. The balance sheets from up to 5 years before failure are available from all of these companies. Various financial ratios result from these balance sheets and are used as the input variables for the decision tree and for the Z-score model.

1.4 Structure

In addition to this introductory chapter, this report consists of five chapters. In fact this report can be subdivided into three parts. The first part, consisting of Chapter 2 and 3, gives some financial background to corporate bankruptcy and introduces the models that will be used in this report. The second part, Chapter 4, describes the experimental setup. Results, conclusions and suggestions for future research will be discussed in the third and last part. (Chapter 5 and 6)

In Chapter 2, the financial background for this report is sketched. A through understanding of the reasons and stimulating circumstances under which corporate distress occurs will be created. Section 2.2 gives a literature review about the history of bankruptcy prediction models. This chapter will be concluded with some limitations and suggested improvements of the discussed techniques.

Chapter 3 is the methodology chapter and is divided into two sections. The first section describes the Z-score model of Altman (1968). The second section is the examination of the CART algorithm. This section starts with the technical explanation of the CART algorithm and proceeds with the application of a CART model for our bankruptcy prediction problem.

Chapter 4 will function as the experimental setup. This chapter starts by introducing the data set of bankrupt and non-bankrupt companies that will be used. Details about this data set will be given. Besides that, this chapter will give a short commentary to the composition of the balance sheet. The financial ratios that are important for this research, will be clarified and the framework of the research will be mapped.

The results of both the Z-score and the CART model will be described in Chapter 5. Positive and negative characteristics concerning these techniques solving this problem will be discussed. Also this chapter will show a table with the accuracy of both models with respect to the bankruptcy prediction problem.

Chapter 6 recapitulates findings of the report and attaches a conclusion to it. This chapter will also suggest new topics for future research.

[4]The Z-score model is extensively described in most of the articles written on the bankruptcy prediction subject

Chapter 2

What is Bankruptcy Prediction ?

This Chapter describes the fundamental idea behind bankruptcy prediction. Section 2.1 starts with background information about corporate distress. Section 2.2 describes the importance of bankruptcy prediction, the history of bankruptcy prediction and will be concludes with some limitations and suggested improvements of the most common techniques used for bankruptcy prediction.

2.1 Corporate Distress

At the end of the twentieth century, corporate distress reached levels not seen since the great depression of the 1930s.[1] The number of business failures and bankruptcies increased together with the increase in corporate distress. Four generic terms that are generally found in literature for corporate distress are *failure, insolvency, default and bankruptcy*. Their individual economic meaning are described in the following paragraphs.

Failure means that the realized rate of return on invested capital, with allowances for risk consideration, is significantly and continually lower than prevailing rates on similar investments. Somewhat different criteria has also been utilized, including insufficient revenues to cover costs and cases of the average return on investment being below the firm's cost of capital. A firm could be an economic failure for many years without failing to cover its current obligations because of the absence of legally enforceable debt.

Insolvency is a term used in a more technical way. It indicates lack of liquidity, so it is more cash based, which happens when a company cannot meet its financial obligations. Technical insolvency most often is the cause of formal bankruptcy declaration. Bankruptcy comes along when the insolvency of a company becomes critical, when the total liabilities of a company exceed a fair value valuation, for example stock based, of its total assets.

Default is another condition that is inescapably associated with distress. Defaults always occur between the debtor firm and a creditor class. A firm is not always immediately in default when it misses a loan payment or its interest payments. However, when a firm misses an interest payments or a principal repayments of publicly held bonds, and this problem is not fixed within 30 days, the security is immediately "in

[1] This section is based on the book Altman (1993)

default". In the last few decades these defaults on publicly held indebtedness have become a commonplace event.

Finally the term *bankruptcy* will be discussed . A firm can go bankrupt when the total liabilities exceed a fair value of the total assets of that firm, as discussed in the paragraph about insolvency. On the other hand a firm can be declared bankrupt by a Federal District Court. This Federal District Court can declare the firm bankrupt immediately or offer the firm to participate to a recovery program, which is called a "bankruptcy reorganization". When a firm value is worth more than its liquidation value, the company has to participate to a recovery program.

The firm's creditors and the owners of the firm are the two primary groups of interest when a firm is in corporate distress. These two groups both have an extremely large importance in the evaluation of the bankruptcy reorganization process. The goal of the reorganization process is to restructure the firm in a way that the firm's financial situation will stabilize and that no other financial problems will occur in the near future.

2.1.1 Causes of corporate distress and bankruptcies

It is important to identify the main reasons for corporate distress with bankruptcies as a consequence. Several studies about this subject have been done over the past decades. An example of these studies was done by a consulting firm, Buccino & Associates (1991)[2]. They surveyed over 1,300 managers, and the result pointed out that, by 88% of the respondents, the quality of management was identified as the primary difference in success or failure. Dun and Bradstreet (1980) identified earlier that lack of experience, unbalanced experience, or just plain incompetence was the cause of firm failures in more than 44% of the situations.

Another important issue to take into account is the relation between the age of a firm and the possibility to fail. Dun and Bradstreet (1980) showed that over 50% of all failures occur with firm with ages between two and five. After the age of five, firms tend to be more stabilized, experienced, established and as an indirect result of these reasons have better access to capital.

Other, mainly financial reasons for firm failure which had the upper hand during the 80s are the following:

Industries Some industries tend to be "sick". Firms which are active in these industries have a high possibility to fail in the near future,

Interest rates Because of high interest rates some firms fall into the position in which they cannot obey to their obligations anymore,

Competition International competition intensifies the charges for companies enormously. Scale advantages will bring with itself that small firms will take off against big firms, because these firms are more capable of doing business at a sharper price,

Debt to equity Companies, particularly in the United States, increased their leverage. Because of that, a lot of firms put themselves in the situation of more obligations. In times of corporate distress these persisting obligations could lead to failure,

[2] See Altman (1968)

Deregulation Deregulating of key industries leads to a far more competitive environment,

Formation rates High new business formation rates will cause higher frequency of firm failures. New companies just have the characteristic to have a higher failure possibility than established companies, as mentioned in the preface paragraph.

Eventually all aforementioned reasons and probably even much more reasons contribute to the chance of failure of a firm. In this report, the focus will primarily be on the financial causes of firm failures, mainly because of the fact that these causes are just more quantifiable. We hope to outline the problem in a way, in order that the models used in this report are able to act as a reliable bankruptcy predictors.

2.2 Bankruptcy Prediction

2.2.1 Why is bankruptcy prediction important

As mentioned in the preface section there are two primary groups of interest[3] when a firm is in corporate distress, but besides these two groups bankruptcy prediction also is of importance to bond holders, and to a lot of other major players in the financial and legal services. The reasons of importance to the owners and the creditors are easy to imagine, but for the other mentioned interested parties some further explanation will be given. For bond holders the default risk of a bond is an important factor influencing the value of a bond. When a company or government, who issued a bond, is not capable of meeting the obligations which come along with the bond, we say that the issuer has 'defaulted' on the bond [4]. So it is of great importance for the bond holder to know about the possibility of failure of the issuer of the bond. For legal and accounting companies, bankruptcies are big business[5]. Particularly the bankruptcy reorganizations processes are extremely profitable for these businesses. Besides them bankruptcy courts are extremely busy handling all new or current bankruptcies.

The bankruptcy prediction problem can be seen a classification problem[6]. Investors, auditors, and other individuals, who like to evaluate the risk of an investment, want to know if the company they are looking at is going to be bankrupt in the near future with a certain probability. The input of the classification problem can be modeled as a vector of financial and/or strategic ratios. Given this vector a classification technique has to be able to assign one of the two possible outputs to the company, bankrupt or not bankrupt.

2.2.2 History of bankruptcy prediction models

Factors which can contribute to the understanding of corporate bankruptcy can be found in the fields of Economics and theory of Business Management. Probably because of the diversity of the problem, we still have not seen a fully successful model to

[3]The owners of the firm and the firm's creditors
[4]Tan (2000) page 11
[5]Altman (1993)
[6]see Trippi and Turban (1996) page 230

predict bankruptcy[7]. A possible explanation could be the fact that all firms are unique and information about firms is limited and most of the time not uniform. Within the bankruptcy prediction models we are not only looking for the explanatory factors, we also want to get a feeling for the reasons why a particularly outcome of bankruptcy was observed.

According to Altman (1993), Smith and Winakor (1935) were the ones who did the first studies about bankruptcy prediction. They were followed by Merwin (1942). Both their studies pointed out that failing firms show significantly different ratios[8] than successful firms do. This basic principle was an enormous breakthrough and offered considerable perspectives for further research. In addition, Hickman (1965) did some research about ratios of large asset-size corporations that experienced difficulties in meeting their fixed indebtedness obligations. Another academic who studied on ratio analysis and bankruptcy prediction models was Beaver (1966). His study is considered as one of the classical works on this subject. Beaver questioned the use of multivariate analysis. Instead, Beaver strongly believes in the use of univariate analysis of financial analysis to predict corporate bankruptcy. Neter (1966) on the other hand, a discussant of Beaver, strongly supports the use of multivariate analysis. Beaver found that up until 5 years before failure of a firm a number of ratios differ from a matched non-failed-firm. Beaver viewed a firm as “reservoir of liquid assets, which is supplied by inflows and drained by outflows. The solvency of the firm can be defined in terms of the probability that the reservoir will be exhausted, at which point the firm will be unable to pay its obligations as they mature”. By this framework Beaver stated four propositions:

- The larger the reservoir, the smaller the probability of failure,
- The larger the net liquid-asset flow from operations, the smaller the probability of failure,
- The larger the amount of debt held, the greater the probability of failure,
- The larger the fund expenditures for operations, the greater the probability of failure.

The ratios which Beaver used were categorized into six groups. These groups together consist of 30 ratios that were expected to capture relevant aspects (see table: A.1). By a univariate discriminant analysis, these ratios were applied on 79 pairs of bankrupt/nonbankrupt firms. The best discriminators were working capital funds flow/total assets and net income/total assets which correctly identified 90% and 88% of the cases.

The studies discussed before make use of several different ratios. These ratios tell us something about the probability of bankruptcy. Most of these ratios measure profitability, liquidity, and solvency. The aforementioned studies didn’t make clear which ratios have the most explaining power. All these studies mentioned different ratios as being the most effective ones. So questions for further research will be;

[7] Bernhardsen (2001)

[8] Beaver (1966) quoted: “A financial ratio is a quotient of two numbers, where both numbers consist of financial statement items”

1. which ratios are most important in the prediction of bankruptcy,
2. what weights should we attach to these ratios,
3. how do we attach these weights objectively.

The bankruptcy prediction model which nowadays still is extremely popular and well known is the Z-score model designed by Altman (1968). This Z-score model, introduced in 1968, is based on five variables and uses multiple discriminant analysis (MDA) which showed very strong predictive power. MDA is a statistical technique capable of classifying observations into groupings based on the characteristics of the observations. The Altman technique and his application of MDA will be discussed in further detail in chapter 3. Various studies validated the results of the study of Altman, and mainly because of that reason MDA became a common used approach in bankruptcy prediction. Although this Z-score model was extremely popular, there was a need to update the model because of the following reasons[9]:

- The Z-score model was focussed on relatively small firms. Because of the dramatic increase in size of the bankrupt firms the need was born for a model which was more able to predict business failure with these ”bigger” firms .
- We would like to have a model which should behave as customary as possible with respect to the temporal nature of the data
- Up until now models only concentrated on specific industries. The updated model must be capable of coping with different industries.
- The models seen up until now only look at past failures. The new model has to be applicable to the data which will appear in the future as well.
- The updated model would enable us to test and assess several of the advances and controversial aspects of the discriminant analysis. For discussions about the use of discriminant analysis especially with respect to failure prediction see H.D.Platt and M.B.Platt (1990, 1991), Weiss (1981), Zavgren et al. (1988) and many others.

For these reasons Altman et al. (1977) came up with the ZETA model which can be applied to larger firms, not limited to specific industries. The ZETA model appeared to do quite well for bankruptcy classification, it showed 90% of the sample one year prior and 70% accuracy up to 5 years.

In the succeeding years several academics used the models introduced by Beaver (1966) and Altman (1993) as a basis for their studies. One of them was Deakin (1972). Deakin proposed an alternative model for bankruptcy prediction which was based on the models used by Beaver and Altman. Deakin tried to capture best of both models, mainly because he believed that Beaver’s empirical results showed greater predictive ability, but the method of Altman had more intuitive appeal. He searched for the linear combination of the 14 ratios used by Beaver which best predicts firm failure in each of the five years prior to failure. Wilcox (1971), Edmister (1972), Libby (1975) and Scott

[9]For further details see Altman (1968)

(1981) where other people who based their studies on the models introduced by Beaver and Altman. Sung et al. (1999) quoted that according to Scott: "the ZETA model is perhaps the most convincing multi-dimensional model since it has high discriminating power, is reasonably parsimonious, and includes accounting and stock market data as well as earnings and debt variables". Ohlson (1980) introduced a logit analysis model for bankruptcy prediction. Ohlson based his study on two unpublished papers of White and Turnball (1975a,b) and on a paper by Santomero and Vinso (1977), the first studies which logically and systematically develop probabilistic estimates of failure. He examined the probability of firm failure by the effect of the following four factors: the size of the firm, measures of the firm's financial structure, measures of performance, and measures of the current liquidity. Instead of using 5 independent variables, as Altman did, Ohlson used 9 independent variables to predict the probability of failure. This model had some success, nevertheless the model have never been a common used approach by practitioners so far.

Classical classification techniques, discriminant analysis, form the basis for bankruptcy prediction models in the studies described so far. The last three decades other methods like recursive partitioning, neural networks, genetic programming and other decision tree techniques gained popularity for the bankruptcy prediction problem as well. The aforementioned Artificial Intelligence (AI) techniques will be described globally in the following paragraphs.[10] These techniques offer good alternatives for the classical techniques discussed so far and will form an introduction to the technique which which will be explored in this thesis.

Recursive partitioning

Recursive partitioning is a supervised learning technique in the form of inductive learning. Supervised learning means that it uses of examples of which the dependent variables are already known. The training of a supervised learning model is based on these dependent variables. Inductive learning is a technique used for building decision trees which are able to learn from examples by a process of generalization. A decision tree partitions an input space of a dataset into subsets. The recursive partitioning procedure then recursively partitions each subset into sub-subsets. By following this procedure there will occur a tree with a root in the top and mutually exclusive regions, leafs, in the bottom of the tree. All of these leafs have a label, in the scope of this report they can have the label bankrupt or nonbankrupt. Inductive learning have the goal to find the hypothesis that fits the data, which has as an advantage that it only requires little prior knowledge.[11] The ID3 algorithm introduced by Quinlan (1986) is a popular inductive partitioning algorithm. This algorithm learns trees by constructing them top-down. Every time the algorithm needs to decide which variable (financial ratios in the scope of this report) is the most valuable. This is done by using a statistical test (see formulas B.2 and B.1) to determine how well it alone classifies the training examples.(for further information about this ID3 algorithm see appendix B.1.1)

In the bankruptcy prediction problem, decision trees are constructed by recursively partitioning the training dataset into subsets until the final nodes (leafs) only consists of one of the two types, bankrupt or nonbankrupt. When the tree is built, a new firm

[10] Technical details about the techniques can be found in Appendix B

[11] see chapter 12 in Mitchell (1997)

can be evaluated by following the tree. The leaf in which the firm will fall, points out to what group, bankrupt or nonbankrupt, it belongs to.

Artificial neural networks

Artificial Neural Networks (ANNs) or simply neural networks offer another suitable classification possibility for the bankruptcy prediction problem. These ANNs perform their classification task in the same way as a human would decide whether water is hot or cold, in a response to impending signals of financial health of a firm. ANNs have proven to be good classifiers in many real-world classification problems due to their nonlinear nonparametric adaptive-learning properties.

ANNs are networks existing of a number of layers of interconnected simple logic units or nodes.[12] These networks have been invented in the 1950s and were inspired by the way scientists believed the human brain worked. The use of of ANNs however, was limited strongly by the lack of suitable training methods. This changed in the mid-1980s with the reformulation of the backpropagation algorithm by Rumelhart et al. (1986). The logical units in feedforward neural networks - as opposed to recurrent ones - are called perceptrons. These perceptrons model a human brain's neuron that 'fires' on the output side when a certain threshold is reached. In perceptrons the input x is a weighted linear combination of the outputs of perceptrons in the previous layer and a so called 'bias' (always equal to 1). The output is computed by using a nonlinear, differentiable activation function called a 'transfer function' or the identity function $f(x) = x$. The following activation functions are most commonly used.[13]

Logistic function:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2.1}$$

Hyperbolic tangent function:

$$f(x) = \tanh(\frac{x}{2}) = \frac{1 - e^{-x}}{1 + e^{-x}} \tag{2.2}$$

Information on explanatory factors would be taken at input nodes via input layers, when using ANNs for the bankruptcy prediction problem. From these input nodes, weighted interconnections are made to hidden layer nodes, which collect and process the information and determine the probability of failure. The first attempt to use ANNs for bankruptcy prediction was done by Odom and Sharda (1990).

Genetic algorithms

Genetic Algorithms (GAs) are stochastic derivative free optimization techniques which can search effectively through very large spaces, in many different ranges of applications. GAs are motivated by the analogy of biological evolution.[14] (Darwin's theory of evolution, survival of the fittest)
GAs have a number of advantages which contribute to their popularity:

[12] This paragraph is based on Jang et al. (1997)

[13] for further information about the backpropagation algorithm see appendix B.2.1 or the books Jang et al. (1997) & Mitchell (1997)

[14] Holland (1992) was the first who proposed and investigated the GAs.

1. The possibility that GAs get trapped in a local minimum is small, because they are stochastic and they use many points instead of one in the search space simultaneously;

2. GAs are able to optimize continuous and discrete functions, or even combinations;

3. GAs can be used on several computers at the same time, which will increase the speed of the optimization;

4. GAs use probabilistic instead of deterministic rules in the optimization;

5. GAs work with strings of characters representing the parameters, instead of working with the parameters themselves;

6. The nature of GAs is inductive, which means that it doesn't have to know any rules of the problem, because it works by its own internal rules;

7. GAs can be used to identify or even estimate parameters for use in other AI models like neural networks or fuzzy inference systems.

Every GAs works with a collection of hypothesis, called a population, which is evaluated every generation. These hypothesis are represented by bit strings, called chromosomes.[15] In each generation these chromosomes are evaluated according to their fitness value, which is usually equal to the output of the objective function. The chromosomes which have the highest fitness value immediately go, unaltered, to the new population. Others are used create offspring individuals by utilizing genetic operators such as crossover and mutation.[16] GAs are heavily used for variable selection for example in neural networks within the bankruptcy prediction.

[15] explanation about terms used in this paragraph see appendix B.3

[16] for further information about GAs see appendix B.3.1 or the books Jang et al. (1997) & Mitchell (1997)

2.2.3 Limitations and improvements

This section describes the (dis-)advantages of both aforementioned statistical and artificial intelligence prediction models. To make a clear comparison between the prediction models is difficult, mainly because each application has different goals and circumstances that need to be treated differently. For this reason every application requires different techniques.

In sign of this report, division has been made between the different prediction models, resulting into discriminant analysis, decision trees, neural networks and genetic algorithms. First the classical, statistical, discriminant analysis will be discussed, which will be followed with the ones which make use of artificial intelligence. For each of these techniques some of the disadvantages will be highlighted.

Discriminant analysis

Discriminant analysis are extremely popular ever since Beaver (1966) introduced the multivariate analysis approach for bankruptcy prediction. Based on his work Altman (1968) introduced his Z-score model, which also makes use of discriminant analysis and is seen as the basic tool for bankruptcy prediction. Although discriminant analysis is so heavily used, there are some disadvantages connected to it. The most prominent disadvantages are taken up in table 2.1[17].

Discriminant analysis
-Requires that the decision set used to distinguish between distressed and viable firms need to be linearly separable
-Does not allow for a ratios signal to vacillate depending on its relationship to another ratio or set of ratios
-Reduction of dimensionality
-Difficulty in interpreting relative importance
-Violations of normality and independence
-Difficulty in specifying classification algorithm
-Difficult to interpret time-series prediction test

Table 2.1: Disadvantages of discriminant analysis.

Decision trees

Recursive partitioning is a supervised learning technique which also gained popularity in the world of bankruptcy prediction. Mainly because decision trees are able to generate understandable rules and are capable to deal with continuous and categorial variables. Decision trees can cope with missing values in a data set. Nevertheless there are at least three demonstrable weaknesses, quoted in table 2.2.

Artificial neural networks

ANNs are less heavily used as the aforementioned techniques, but they also catched up popularity for bankruptcy prediction problems. ANNs can handle a wide range

[17] Source of information Bernhardsen (2001) and Sung et al. (1999)

of problems and produce really good results for complicated problems, and is like decision trees capable of coping with continuous as well as with categorial variables. Some of the disadvantages are cited in tables 2.2.

Genetic algorithms

Most of the advantages of genetic algorithms are described in section 2.2.2 and B.3.1, the most apparent disadvantages can be seen in the following table[18], which comprises of the main disadvantages of the decision trees, ANNs and GAs.

Decision trees	ANNs	Genetic algorithms
-Error-prone with too many classes	-Black boxes, difficult to understand	-Difficulty in encoding
-Computationally expensive to train	-Cannot explain the results	-No guarantee of optimality
-Trouble with non-rectangular regions	-May converge on an inferior solution	-Computationally expensive

Table 2.2: Disadvantages of the different Artificial Intelligence techniques.

2.2.4 Conclusion

Discriminant analysis and decision trees are the two techniques to be used within this report. The discriminant analysis model will function as a benchmark model. The reason for choosing this discriminant analysis as the benchmark technique was rather obvious, because this technique simply is the most common used and most examined technique for bankruptcy prediction. Besides that, this technique proved to be robust and has showed respectable prediction accuracies for this classification problem. The reasons for choosing a decision tree based technique was less obvious. Classification and Regression Trees (CART) was selected as the main technique for this report. A number of factors was determinative for choosing this technique. CART trees are able to cope with missing values, the final models are easy to interpret and show a good overall picture of the data set. Moreover, relatively little research has been done on CART models for the bankruptcy classification problem.

[18]Source of information Bernhardsen (2001) and Sung et al. (1999)

Chapter 3

Methodology

This Chapter describes the relevant techniques used in this research. The first section will extensively discuss the traditional Z-score model of Altman (1968). This Z-score model functions as the benchmark model within our research. “Classification and Regression Trees” (CART) will be described in Paragraph 3.2, and is considered as the main technique to predict bankruptcies within this research. First the technical background of this, decision tree based, technique will be described. Subsequently a sketch of a CART model application for the bankruptcy prediction problem will be given.

3.1 Classical bankruptcy prediction: Altman Technique

This section will be dedicated to the bankruptcy prediction model introduced by Altman (1968)[1]. As mentioned before in subsection 2.2.2, Altman’s Z-score model is extremely popular ever since the introduction in 1968. The Z-score model is based on multiple discriminant analysis (MDA), a statistical technique used to classify an observation into one of the, a priori determined, groupings dependent on the observation’s individual characteristics. This MDA model is primarily used for predicting qualitative dependent variables, so very suitable for the bankruptcy prediction problem[2]. The MDA technique has the advantage to take several characteristics into account at the same time, as well as their inter-relational behavior. Another strength of MDA is the ability to reduce the ”a priori” groupings into a small dimension. MDA reduces the dimension in our problem to the simplest form, where the a priori groupings are bankrupt and nonbankrupt. Because of this, the MDA model used for bankruptcy prediction can treated as “simple” discriminant analysis (DA).

3.1.1 Discriminant analysis

Before exaggerating about the DA application for the bankruptcy prediction problem, we need to have a clear description of discriminant analysis on its own. Discriminant analysis, like analysis of variance, is an analysis of dependence method which actually

[1] based on chapter 8 of Altman (1993)

[2] bankrupt, nonbankrupt

is a variant of canonical correlation.[3] However in the case of discriminant analysis the dependent variables are categorical, which divide the set of observations into mutually exclusive and collectively exhaustive groups. Simple discriminant analysis, which has only two groups, only needs a single dichotomous dependent variable to indicate group membership. For multiple discriminant analysis (MDA) we need $n-1$ dichotomous variables to indicate group membership across n groups. Discriminant analysis make use of information about independent variables, and turn that information into the clearest possible separation between or among groups. The Fisher approach, a well known discriminant analysis approach, finds the linear combination of independent variables that produces the best discriminant score. If we look at the scatter plot 3.1 [4] and keep the two plots of figure 3.2 in mind, we see that figure 3.4 represents the best linear combination to discriminate between the two groups.

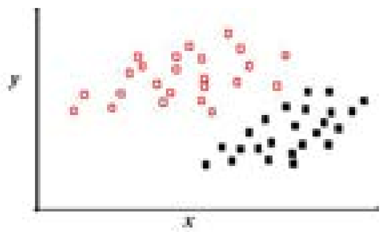

Figure 3.1: Scatter plot showing two groups.

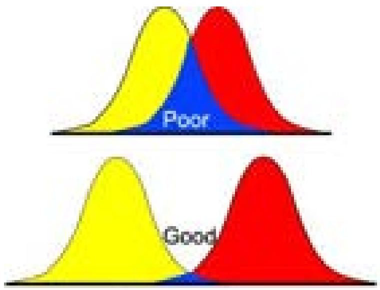

Figure 3.2: Visualization of two possible linear combinations of two independent variables.

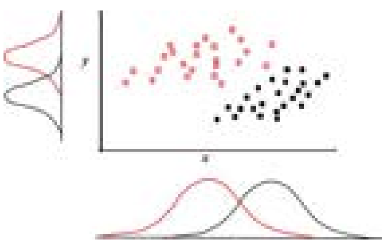

Figure 3.3: Scatter plot of two groups using x or y to discriminate between both groups.

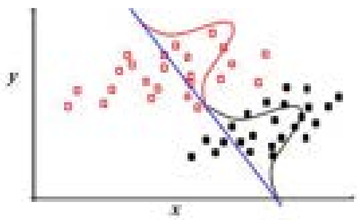

Figure 3.4: The best linear combination of x and y to discriminate between both groups.

[3] This section is based on Lattin et al. (2003) Chapter 12

[4] Source of information:
http://www.doe-mbi.ucla.edu/~parag/multivar/dawords.htm

When we look at the technical background and when **k** denotes the linear combination, then the discriminant scores are given by;

$$\mathbf{t} = \mathbf{X}\mathbf{k} \tag{3.1}$$

The linear combination of **k** has to maximize the ratio of the between-group sum of squares to the within-group sum of squares of the discriminant scores **t**. This ratio is proportional to:

$$\frac{\mathbf{k}'\mathbf{d}\mathbf{d}'\mathbf{k}}{\mathbf{k}'\mathbf{C}_W\mathbf{k}} \tag{3.2}$$

where $\mathbf{d} = (\bar{\mathbf{x}}_{(2)} - \bar{\mathbf{x}}_{(1)})$ is a vector describing the difference between the means of the two groups, and $\mathbf{C}_W$ is the pooled within-group covariance matrix of **X**. So the smaller the within-group variation the larger the objective function. In fact, equation 3.2 is maximized by choosing **k** as follows:

$$\mathbf{k} \propto \mathbf{C}_W^{-1}\mathbf{d} \tag{3.3}$$

Formula 3.1 calculates the discriminant function scores for all the observations in the data set.

A cutoff score, t_c, can be used to categorize observations. All observations with discriminant function scores $t > t_c$ are assigned to one group, the others are assigned to the other group. The following formula will do to calculate the cutoff score for a two group discriminant analysis

$$t_c = \frac{(\bar{t}_{(1)} + \bar{t}_{(2)})}{2}, \tag{3.4}$$

where $\bar{t}_{(1)} = \bar{\mathbf{x}}'_{(1)}\mathbf{k}$ and $\bar{t}_{(2)} = \bar{\mathbf{x}}'_{(2)}\mathbf{k}$ are the discriminant function scores of the two group centroids. This formula is only appropriate when the two groups are of equal size. When the groups sizes differ the following formula for the cutoff score is

$$t_c = \frac{(n_1\bar{t}_{(1)} + n_2\bar{t}_{(2)})}{n_1 + n_2} \tag{3.5}$$

This equation minimizes the expected probability of misclassification.

The objective in MDA is no different than the objective in DA. The only difference is the number of groups exceeds two, which has more than one dependent variable as a consequence[5]. Letting **W** denote the within-group sum of squares matrix and **A** denote the across-group sum of squares matrix, then the objective for this discriminant problem is given by

$$\frac{\mathbf{k}'\mathbf{A}\mathbf{k}}{\mathbf{k}'\mathbf{W}\mathbf{k}}, \tag{3.6}$$

When we take the derivative and solve the first order condition for **k** we get

$$\mathbf{W}^{-1}\mathbf{A}\mathbf{k} = \lambda\mathbf{k} \tag{3.7}$$

Because the bankruptcy prediction problem only has two groups, namely bankrupt and non-bankrupt, the MDA model constructed for this report can be seen as a two group discriminant analysis. For this reason Chapter 4 and 5 continues with formulas 3.1 to 3.5.

[5]For three groups we need two indicator variables. For instance, Y_1 and Y_2 which indicate two groups and if both of them are false $\Rightarrow$ third group.

3.1.2 Discriminant analysis application for bankruptcy prediction

The DA determines the discriminant coefficients, while the independent variables are the input (actual) values[6]. These individual variable values transform the following discriminant function:

$$Z = V_1X_1 + V_2X_2 + \ldots + V_nX_n \qquad \begin{cases} V_1, V_2, \ldots, V_n = & \text{discriminant coefficients, and} \\ X_1, X_2, \ldots, X_n = & \text{independent variables,} \end{cases} \tag{3.8}$$

into a single discriminant score, or Z-score. Some of these variables might have a certain degree of correlation or collinearity with each other. This brings careful selection of the variables as a consequence. The DA technique, with its ratio analysis, has the potential to reformulate a problem properly. The discriminant score in fact is the value which follows from applying a discriminant function formula to the data for a given case. The Z-score is the discriminant score for standardized data. A Z-score measures the placement of a specific value in terms of the number of standard deviations away from the mean. There is a corresponding Z-score for each value in a particular data set. The definition for the Z-score given by Shiffler and Adams (1995) is:

> "The Z-score, denoted by Z, that corresponds to a value of X is the distance between X and its mean in units of standard deviation."

Altman's Z-score model is a linear analysis of five ratios, which are all weighted objectively. The summation of the five ratios formS the Z-score. This Z-score forms the basis for the classification of a firm. It classifies a firm into one of the a priori known groupings.

Altman based his research for this Z-score model on a data set of 66 firms, 33 bankrupt and 33 non-bankrupt firms. The bankrupt firms went bankrupt in the period 1946 - 1965 and all were manufacturers. Every bankrupt firm was matched with a non-bankrupt firms with the same characteristics. It's obvious to see that this data set might not be the best data set for this research. A number of points for discussion are:

- Small data set. Only 33 bankrupt firms and 33 paired non-bankrupt firms form this data set. It is difficult to say if this data set is reliable enough to form the basis for his research.
- Long time period. It would be better to have a smaller time period, because average ratios can drift over time.
- Small mean asset size. The mean asset size is \$6.4 million dollar, with a range from \$0.7 million and \$25.7 million dollar. It would be better to have a more synchrone data set with respect to this asset size.
- Only one specific industry. The data set, and with that the Z-score model is based on only one industry; manufacturers. This produces a reliable model for this industry, but as a result it is risky to use this model for other industries.

[6] This subsection is based on the books Altman (1993)(Chapter 8) and Trippi and Turban (1996)

In spite of these problems, this date set functioned as the input for the Z-score model. Unfortunately it simply wasn't possible to built a date set which was more suitable, thanks to the lack of available data.[7]

Balance sheets and income statements were collected for these 66 companies. Altman did a comprehensive literature review to collect the variables (ratios) which are significant indicators for bankruptcy prediction. A list of 22 helpful ratios originated, which were classified into five groups or categories, including liquidity, profitability, leverage, solvency, and activity. The best[8] five ratios were selected for the Z-score model. Several steps had to be taken to get to the final profile of variables. First, observe the statistical significance of the various functions and determine the relative contribution of the independent variables. Second, look at the inter-correlations between the variables. Third, observe the predictive accuracy. Fourth, take a judgement by an analyst. The following discriminant function formed the Z-score model of Altman (1968) after taking these steps.

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5 \tag{3.9}$$

where,

X_1 = Working capital / total assets,

X_2 = Retained earnings / total assets,

X_3 = EBIT / total assets (where EBIT is earnings before interest and taxes),

X_4 = MVE / total debt (where MVE is the market value of equity and total debt is book value of total liabilities)

X_5 = Sales / total assets

Altman assigned all firms with a Z-score higher than 2.99 to the category of non-bankrupt, whereas he assigned firms with a Z-score below 1.81 to the category of bankrupt firms. Firms with a Z-score between 1.81 and 2.99 were allocated to the gray area, what means that the model cannot link a conclusion to these firms. However, for simple predictive purposes, Altman classified the firms as bankrupt if $Z \leq$ 2.675 (the 1968 model's cutoff score) and as non-bankrupt if $Z >$ 2.675.

Altman used an accuracy matrix to test the accuracy of his Z-score model. (see table 3.1[9]) Where H stands for correct classifications, M_1 for a Type I error and M_2 represents a Type II error. To see how accurate the model is, you have to sum the H diagonal and divide this number by the total number of firms in the data set. This will give the accuracy percentage.

[7] The data was derived from Moody's Industrial Manuals and selected annual reports

[8] The best overall job together in the prediction of corporate bankruptcy. This means that the individual prediction power of each variable might not be the most significant.

[9] copied from Altman (1993) page 191

	Predicted Group Membership	
Actual Group Membership	Bankrupt	Non-Bankrupt
Bankrupt	H	M_1
Non-Bankrupt	M_2	H

Table 3.1: The accuracy matrix used to measure the accuracy of the Z-score model of Altman.

Altman's model proved to be extremely accurate, the correct classifying percentage was 95% with the initial sample of 33 firms in each group.[10] Type I and Type II error respectively were 6% and 3%.[11] From these numbers you can conclude that his model proved to be accurate. But remember the fact that this model is only based on the small data set as discussed before.

The second test for the Z-score model of Altman was to observe the degree of discriminating power. For this test, the model was exposed to a data set consisting of a compilation of two financial statements prior to bankruptcy. This gave a reduction in accuracy, which is explicable because the indications are less clear. Notwithstanding it correctly classified 83 % of the total sample.[12]

Until now the same data set has been used for training the model as well as for testing the model. This can give a distorted picture of the accuracy, for example because of overfitting toward the training data set. In order to test the accuracy more rigorously two new data sets were introduced. The first consisted of 25 bankrupt firms with more a less the same asset size as the bankrupt firms in the initial group. The result for this data set was astonishing, it even exceeded the initial discriminant sample. (96 % versus 94 %) The second new data set consisted of 66 non-bankrupt firms which suffered from a two or three years of negative profits. Again the performance of the model on this data set was surprisingly high, 79 % of the sample firms was correctly classified as non-bankrupt[13].

3.2 Classification and Regression Trees

The "Classification and Regression Trees" (CART) methodology is introduced by Breiman et al. (1984) and is a classification technique which uses supervised learning[14] to built a decision tree. Decision trees are represented by a set of questions which split a data set into smaller and smaller parts. As a result decision trees partition a data set into mutually exclusive regions. When decision trees are used for classification problems, these trees are called classification trees. If it concerns a regression problem then they are often called regression trees. CART is a technique which combines both of them. The CART algorithm only asks yes or no questions like: "Is the (working capital / total assets) $> x$?" [15] The algorithm searches for all possible input

[10]When using the financial statement 1 year before bankruptcy

[11]Type I error of 6% means that 2 out of 33 bankrupt firms were predicted as non-bankrupt

[12]For details about the accuracy of the prediction model of Altman see tables 8.4, 8.5 and 8.7 on pages 191-193 in the book of Altman (1993)

[13]For further details see Altman (1993) page 194

[14]Make use of historical data, from which the values of the dependent variables are known on forehand.

[15]Where x is certain number which act as a boundary line.

variables in order to find the best split.[16]

CART is a classification technique which can function directly as a model, or can be used for structure identification for other techniques. CART can, for example, identify the most relevant input variables, which can be helpful for other models which can only handle a few input variables.[17]

3.2.1 Technical background: Classification and Regression Trees

Recursive partitioning, or in other words decision-tree induction of sample data, is a heavily explored topic within Artificial Intelligence, ever since the introduction.[18] ID3 and C4, both proposed by Quinlan (1986), are good examples of recursive partitioning algorithms. At the same time Breiman et al. (1984) introduced the CART algorithm, which was built to approach similar problems. According to Jang et al. (1997) the fundamentals of ID3 and CART are the same, the main difference between CART and ID3 is that CART induces strictly binary trees and uses resampling techniques for error estimation, while ID3 partitions according to attribute values.

The following sections will describe the three stages of the CART algorithm. It first grows the tree to the maximum tree is reached, which means splitting the learning sample up to the last observations. Then the tree pruning process will be explained. This implies cutting of some insignificant nodes and/or even subtrees to counteract overfitting. The section will be concluded with tree testing. Different ways to test a tree will be discussed in this part.

Tree growing

Tree growing is the most time consuming stage of the CART algorithm. It simply means that the tree gets bigger by splitting the data into subsets. This action is done recursively in order to split the entire data set into disjoint subsets. Every time when the algorithm wants to grow the tree it searches for the best split, which means that it searches for the splitting variable that best reduces an error measure.

Now we will have a closer look at the construction of classification trees and proceed afterwards with the construction of regression trees.

Classification trees are used when we know the class, to which the samples belong to, up forehand. In order to grow a classification tree properly, we need to have an "impurity function". This function acts like an error measure $E(t)$ that quantifies the performance of a node t in separating data into classes. This measure should get a value of zero when the given data all belong to one class. When the data is evenly distributed through all possible classes, this measure reaches its maximum value.

The fundamental idea is to select the splits in such a way that the data in the subsets is "purer" than the data in the parent subset. So the purity of the descendent nodes has to be greater than their parent nodes. Figure 3.5[19] illustrates

[16] The question which splits the data into two parts with the highest homogeneity.

[17] For instance when building a fuzzy system. This system can only use a few variables, otherwise the input-space would increase to an unacceptable level.

[18] This subsection is based on Chapter 14 of Jang et al. (1997)

[19] Copied from Breiman et al. (1984)

this idea clearly. Lets say t is a node, and some split s is going to be happening which divides t into t_L and t_R. Here a proportion p_L is going to node t_L and a proportion p_R is going to node t_R.

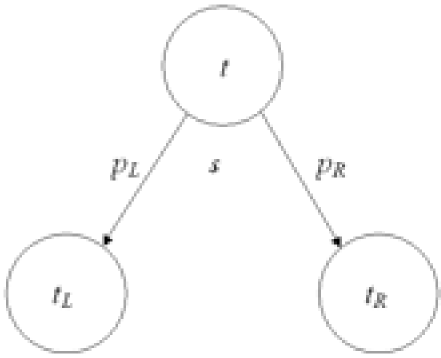

Figure 3.5: Visualization of a split from parent node to a left and a right child-node.

The change in impurity which follows from this illustration is[20];

$$\Delta i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R), \tag{3.10}$$

and measures the "goodness" of the split. And therefore the CART algorithm solves for every node the following maximization problem;

$$\arg\max[i(t) - p_L i(t_L) - p_R i(t_R)] \tag{3.11}$$

When we look at the impurity function for a J-class problem, we can see that the impurity function ϕ is a J-place function that maps its input arguments $p_1, p_2, \ldots, p_j$, with $\sum_{j=1}^{J} p_j = 1$, into a non-negative real number, such that

$$\phi(1/J, 1/J, \cdots, 1/J) = \text{maximum}, \tag{3.12}$$
$$\phi(1,0,0,\cdots,0) = \phi(0,1,0,\cdots,0) = \phi(0,0,0,\cdots,1) = 0.$$

Which means that the impurity is largest when the all J classes are equally mixed over the subsets, and is smallest when the subsets only contain one class. The impurity measure of node t is expressed as

$$E(t) = \phi(p_1, p_2, \ldots, p_J), \tag{3.13}$$

when using this impurity function ϕ, where p_j stands for the percentage that each class is represented in node t. The function of the impurity measure of the total tree is expressed as

$$E(t) = \sum_{t \epsilon \widetilde{T}} E(t), \tag{3.14}$$

[20] Technical background is based on Breiman et al. (1984), Jang et al. (1997) (Chapter 14) and Timofeev (2004)

where $\widetilde{T}$ is the set of terminal nodes in tree T.

In theory there are several impurity functions, the *Entropy function* and the *Gini splitting rule*, or Gini diversity index, are the most popular ones and will be described shortly. The *Entropy function* is expressed as

$$\phi_e(p_1, \dots, p_J) = -\sum_{j=1}^{J} p_j \ln p_j, \tag{3.15}$$

and the *Gini splitting rule*, the most broadly used rule, is expressed as

$$\phi_g(p_1, \dots, p_J) = \sum_{j \neq 1} p_i p_j = 1 - \sum_{j=1}^{J} p_j^2, \tag{3.16}$$

where $\sum_{j=1}^{J} p_j = 1$ and $0 \leq p_j \leq 1$ for all j, the preceding two functions are always positive unless one of p_j is unity and all the others are zero. Both formulas reach their maxima when $p_j = 1/J$ for all j.

In a binary tree, which are used by the CART algorithm, the impurity change due to splitting is expressed as follows:

$$\Delta E(s,t) = E(t) - p_l E(t_l) - p_r E(t_r), \tag{3.17}$$

which gives the following change in impurity when we apply the *Gini splitting rule* to the maximization problem 3.10:

$$\Delta i(s,t) = -\sum_{j=1}^{J} p_{j|Parent}^2 + p_{Left} \sum_{j=1}^{J} p_{j|Left}^2 + p_{Right} \sum_{j=1}^{J} p_{j|Right}^2 \tag{3.18}$$

When we go back to formula 3.17 we can say, in symbols, that the tree growing procedure is trying to find the "best" split s^* for a root node t_1 in the way that this split gives the largest decrease in impurity:

$$\Delta E(s^*, t_1) = \max_{s \epsilon S} \Delta E(s, t_1), \tag{3.19}$$

where S is a set of all possible ways of splitting the cases in node t_1. After defining the best split s^* node t_1 is splitted into t_2 and t_3, this procedure will repeat over and over until the change in impurity will be smaller than a certain threshold. Until now we have only discussed input variables that are numerical or ordered. A typical question for these variables could be;

$$\text{Is} \quad x \leq s_i?$$

Where x is a numerical value and the split value s_i regularly is the average of the x values of two data points that are adjacent in terms of their x coordinates alone. The number of possible splits is equal or less than the size of the dataset minus one. The same procedure is applicable to another category of input variables, namely category variables. The question to ask here is;

$$\text{Is} \quad x \text{ in } S_1?$$

Because of the fact that splitting a node depends on how to put the possible labels of a variable into disjoint sets. S_1 usually is equal or less than half the size of S, and is a non-empty proper subset of S.

Another impurity function is the *Twoing splitting rule* which will maximize the following change in impurity;

$$\frac{p_l p_r}{4}[\sum_j p_j(t_l)p_j(t_r)]^2 \tag{3.20}$$

This *Twoing splitting rule* searches for two classes that make up together more than 50% of the data. This has as a result that this *Twoing splitting rule* will produce more balanced trees than the *Gini index* does. The main disadvantage of this *Twoing splitting rule* is the fact that it works slower than the *Gini index*.

Regression trees have the same goal as classification trees. Splitting nodes recursively and minimize an error measure is the goal of both types of trees. The difference is that regression trees do not have pre-assigned classes. Because of that reason the aforementioned impurity functions[21] can not be applied to regression trees. The splitting procedure in regression trees usually minimize the squared error by the following formula:

$$E(t) = \min_{\boldsymbol{\theta}} \sum_{i=1}^{N(t)} (y_i - d_t(\mathbf{x}_i, \boldsymbol{\theta}))^2, \tag{3.21}$$

where $\{\mathbf{x}_i, y_j\}$ is a typical data point, $d_t(\mathbf{x}, \boldsymbol{\theta})$ is a local model for t and $E(t)$ is the mean-squared error of fitting the local model d_t to the data set in the node. The change in error works approximately the same as the change in impurity function with the classification trees. The change in error function for regression trees are expressed as

$$\Delta E(s,t) = E(t) - E(t_l) - E(t_r). \tag{3.22}$$

The goal of this function is the find the split with the highest decrease in error. The following formula shows this

$$\Delta E(s^*,t) = \max_{s \epsilon S} E(t,s). \tag{3.23}$$

For now we have discussed how to construct a maximum tree by tree growing procedures. The problem of these maximum trees is the fact that they are too specific and too much adapted on a training data set. This gives the problem of overfitting the data. For this reason the tree has to be pruned. This procedure is described in the next section.

Tree pruning

After having constructed a maximum tree, the next step in the CART procedure has to be taken: “Tree pruning”. The global idea behind this step is to counteract overfitting.

[21] like the Gini diversity index, Entropy function and the Twoing splitting rule

This maximum tree is built by a training data set, so it is made to fit this training data set. This has as a big disadvantage that this tree will be too specific for the training set, and are because of this situation biased toward this training data set. Besides that, these trees are frequently too large and as a direct result of this are extremely complex to read. Too large trees have a higher misclassification rate than generalized smaller trees. On the other hand these trees can not be too small, because then they loose their prediction ability. As a result of above mentioned reasons, maximum trees are mostly overspecialized toward training data sets, and therefore these trees can not generalize good enough for new data points.

The goal of tree pruning is cutting the tree in such a way that it leaves a tree with the right size.[22] So after having constructed the maximum tree T_{max}, we have to find the weakest subtrees in it. This weakest subtree can be deleted. This will have a smaller, less complex, more general tree as a result. An effective way to find the tree with the right size is based on the principles of *minimal cost-complexity* or *weakest-subtree shrinking*.

Breiman et al. (1984) stated the following definition for *minimal cost-complexity*: *For any subtree $T \subset T_{max}$, define its complexity as $|\widetilde{T}|$, the number of terminal nodes in T. Let $\alpha \geq 0$ be a real number called the complexity parameter and define the cost-complexity measure $E_\alpha(T)$ as*

$$E_\alpha(T) = E(T) + \alpha\widetilde{T}. \tag{3.24}$$

From this formula follows that $E_\alpha(T)$ is a linear combination of the "cost" and the "complexity" of a tree. The α parameter describes the complexity cost per terminal node. If T_{max} is so large that every leaf in the tree contains only one case, then every case is correctly classified.[23] For this reason the α "complexity" parameter functions as a penalty parameter. As a result of this penalty parameter the minimized tree contains less terminal nodes than T_{max}.

We want to search for a smaller tree than T_{max} which minimizes $E_\alpha(T)$. This leads to the following minimization problem[24]

$$E_\alpha(T(\alpha)) = \min_{T \subset T_{max}} E_\alpha(T) \tag{3.25}$$

where the tree is T_{max} when α is zero and when α is sufficiently large T will consist of only the root, so T_{max} will be pruned completely. Although α runs through an continuum of values, only a limited number of subtrees of T_{max} possible with progressively fewer nodes than T_{max}. Though this pruning process is easy to describe, a few critical question are left:

- Is there one unique $T \subset T_{max}$ which minimizes $R_\alpha(T)$?
- In minimizing sequence of trees $T_1, T_2, \ldots,$ is each subtree obtained by pruning upward from the previous subtree, i.e., does the nesting $T_1 \succ T_2 \succ \ldots \succ \{t_1\}$ hold?

[22]The right size means that the tree is complete enough to make proper and accurate predictions without losing its generality.

[23]At least for the training data set.

[24]Has to be done for each value of α.

Next to these questions the most important issue is to that of finding an effective pruning algorithm. Every time we need to search for the next minimizing tree of T starting with the T_{max}. A direct search through all possible subtrees is computationally very expensive. To find the next minimizing tree for a tree T, we follow this procedure:

a) Find, for each internal node t in T, a value for α that makes $T - T_{max}$ the next minimizing tree. This α, denoted by α_t, is expressed as

$$\alpha_t = \frac{E(t) - E(T_t)}{|\widetilde{T}_t| - 1}, \tag{3.26}$$

where α_t is equal to the ratio between the change in error measures and the change in the number of terminal nodes before and after the shrinking,

b) Then we have to choose the internal node with the smallest α_t. This node will function as the target node for the shrinking procedure.

c) Repeat step a) and b) until the tree only consist of a single root node.

Only a few candidate trees will be the result of this pruning procedure, which reduces the problem enormously. The next step is to select one of these few trees as the optimum-sized tree. This decision is made by testing the trees. The one which gives the smallest error is the optimum-sized tree. This testing procedure can be done in two ways. The first option is to check the trees on an independent test data set. This is an easy and computational cheap option. The second option, cross-validation, is a bit more time consuming but more reliable, because is makes more effective use of the available data.[25]

Most of the time these testing procedures apply both to the training and to the test data sets. In fact nearly in every situation you encounter the same pattern.[26] The error measure decreases, when the tree complexity becomes larger for the training data set. This sounds logical, because then the tree is overspecializing toward the training set. The same pattern happens for the error measure when looking at the test data set for the smaller, low complex trees. The error measure will decrease when the number of terminal nodes increase until a certain number of nodes are reached. After this number of nodes the error measure will increase in case of adding an extra node. This is the point where the tree looses generality. Therefore, it is the goal to find this critical point, number of terminal nodes, where the error measure decrease change-over to increase in case of adding another node.

Tree testing & classification of new data

Before starting to classify new data we need to test the accuracy of the tree extensively. Roughly this testing of a constructed tree can be done in three ways:

Holdout method The idea behind this technique is extremely simple. It simply splits a data set into two individual data set, a training and a test data set.[27] So it

[25] Cross-validation will be discussed in the following section and in chapter 4 and 5.

[26] In a situation when you plot the error measure on the y-axis, and the number of terminal nodes on the x-axis

[27] Where the test set usually is smaller than the training set.

leaves the test set out of the training of the tree. In this way the test data set can function as an independent data set to determine the accuracy of the tree. The advantages of this method is its simpleness and low computational costs. The main disadvantage is the possibility of high variance in the evaluation. Mainly, because this evaluation totally depends on the way how the data set is separated.

K-fold cross validation This technique is an improvement over the holdout method described before. This technique splits the data set into k subsets, so the holdout method can be repeated k times. Every iteration one subset is used as the test data set, while $k-1$ subsets form the training data set. Then the average error of the k tests is computed. Because of this technique some of the disadvantages of the holdout method are banned. For instance the fact that this technique uses several different test data sets, it reduces the change to score high variances. As a disadvantage there is the problem of high computational cost, because it has to train and test the tree k times. The number which they usually assign to k is 10. A variant to this k-fold cross validation is to randomly divide the data set into a training and test data set k different times.

Leave-one-out cross validation This is the same as k-fold cross validation, only now it assigns the size of the total data set to k. This means that every iteration the tree is trained by $N-1$ data points and tested by 1 data point. In this way the data set is optimally utilized.

It depends on the data set and on the situation which of the aforementioned techniques to elect as the best technique. But as you can see the last two techniques make more use of a data set than the holdout method and obtain smaller variances in the evaluation.

Finally, after the testing procedure, the tree can be used for classification of new data. When a new observation arrives at the tree, it will be confronted with several question which will lead this observation to a terminal node. This terminal node will assign a class to the observation.[28]

3.2.2 Strengths and weaknesses of CART

So far we have broadly discussed the technical details of CART. In this section the advantages and disadvantages of the CART method will be made explicit.

Strengths of CART

Most of the advantages of CART will be highlighted in the following summary:

- CART is nonparametric and does not make any assumptions about variables to be selected. CART can be used as a tool to identify the most significant variables.

- CART can handle outliers easily, and is not affected to collinearities and heteroscedasticity.

[28] Bankrupt or non-bankrupt are the two classes used in this report to which the observations can be assigned to.

- CART does not make any distributional assumptions. Therefor none of the variables in CART follow a statistical distribution.
- CART can handle continuous, interval and even categorical explanatory variables.
- CART is able to cope with outliers.
- CART is invariant to monotone transformation of independent variables.[29]

Weaknesses of CART

Next to the strengths that the CART method possess, it has a few weaknesses as well. A few of these weaknesses are:

- It is not based on a probabilistic model, instead it is purely based on historical data. The accuracy of these models is also based on historical accuracy.
- Sometimes CART can produces unstable decision trees.

As you can see the strengths of CART win from the weaknesses. In particular the fact that CART can be used as a tool to identify the most significant variables, is extremely powerful. In this way the results of this report can be used for other research, where the most significant variables have to be selected in advance.

3.2.3 Application of CART for bankruptcy prediction

CART is an appropriate method for the bankruptcy prediction problem. In fact, the bankruptcy prediction problem is very simple to understand and to model. The CART model only has to classify each observation into two classes, bankrupt or non-bankrupt for this problem. Especially when the most significant variables are not known in advance, CART can offer good results. CART can handle many different variables at high speed, for this reason there is no restriction to the number of input variables.

In Chapter 4 the construction of a CART model, for the bankruptcy prediction, will be clarified. Besides that all encountered problems will be described in this Chapter. First of all the data set, the financial statements and the the financial ratios to be used will be clarified carefully. This is crucial for the construction of this prediction model. Secondly, the construction of the models, the Z-score model and the CART model, will get attention. In this section the complete preparation of both these models will be done for this economical classification problem.

[29] This means that the transformation of explanatory variables to logarithms or squares or square roots has no effect on the tree produced. The splitting variables stay the same, only the values will change.

Chapter 4

Experimental Setup

The fundamental research of this report will be outlined in this Chapter. This Chapter starts with the financial setup in section 4.1. This section consists of four subsections, in which the data set and financial ratios, derived from this data set, will be explained. The architecture of the models will be sketched in section 4.2.

4.1 Financial Setup

This section pays attention to the data collection and to the derivation of the financial ratios. Before we go deeper into these financial ratios, the balance sheet and in the profits and losses accounts (P&L), which are used for data representation, will be illustrated.

4.1.1 Data set

The data set used for the research of this report consists of 122 Dutch N.V.'s[1]. (61 bankrupt and 61 non-bankrupt firms.) All of these companies[2] were listed, during the period 1945-1999, at the official Amsterdam Stock Exchange (AEX). The motivation for this data set is based on a number of reasons and restrictions. For instance, the companies in the data set had to be located in one country with a focus on only one industry group. Besides that, the companies had to be listed, otherwise it just would have been to hard to get all the relevant financial information about these companies. Only with the combination of these restrictions my research and models can give reliable results.

For the period 1945-1999, 134 AEX listed companies we re declared bankrupt. These companies were allocated to five industry groups.[3] The fifth industry group, business trade, industries and others, was best represented with 65 bankrupt companies. As mentioned before, the restriction was to look at one industry group, but on the other hand data set had to be representative as well, so therefore the focus of this re-

[1]This data set originated from the research done by Caljé (2000) about "Costs of Financial Distress and Bankruptcy in the Netherlands"

[2]See appendix C and tables C.1,C.2 and C.3

[3]1) Bank, credit and insurance, 2) Railway, 3) Mine and Petroleum, 4) Cultural, 5) Business trade, industries and others.

port was on this fifth industry group, which in fact is a combination of several smaller groups.

Only 61 of the 65 bankrupt companies have been taken along in this report, mainly because of the lack of financial information about the other 4 companies. For each of these 61 companies the data set provides a ”matched” financially healthy company. This means that every bankrupt company is paired with another non-bankrupt listed company with the same core business, size and period of listing on the AEX. The financial data[4] of the foregoing five years to the possible failures were collected for this data set. These five years will give a reliable idea behind the process which happens before a possible firm failure.[5]

4.1.2 Financial statement analysis

The financial statement of a firm encapsulates the balance sheet and the profits and losses account. These financial statements must obey to strict accounting regulations. These accounting regulations vary enormously among different countries, which does not matter in the scope of this report because all the companies absorbed into the data set originate from the Netherlands. Besides these variations, financial statements of different industry sectors can diverge as well. Especially the financial statement of industrials and of financial companies differ, among other things like the fact that industrials have a lot of inventory on their balance where financial companies have not. The next section describes the ”normalized” financial statements used for this report.

4.1.3 Balance sheet and profits & losses accounts

In this section the “normalized” financial statement will be explained. The term “normalized” is used, because the data set consists of 122 companies from a time period between 1945 and 1999, where the financial statements were multifarious. For this reason it was necessary to standardize and translate these financial statements into one uniform financial statement, otherwise it simply was not possible to calculate comparable values/ratios.

The financial statements and the stock quotes of the foregoing 5 years to the possible failures of all the 122 companies were collected, which add up to 610 financial statements and stock quotes. The underlying thought behind the foregoing 5 years is the fact that a single financial statement, which follows from a single annual report, is a recording of single moment. A firm can for instance increase their liquidity just before the publication of the annual report. When you have a look at five preceding years it is possible to look at the average, which gives a more reliable portrait of the financial situation of the firm. Tables 4.1 and 4.2 illustrates the standardized financial statements used for this report.[6]

[4]Financial data consists of five years of balance sheets, P&L’s and stock quotes.

[5]Several sources of information were consulted for the collection of the data. These sources of information are; Publications in the financial press, annual reports, data research done before by other researchers, yearbook of Dutch companies, archive Amsterdam Exchange, manual of the official pricepaper of the Amsterdam Exchange, van Oss securitiesbook.

[6]To see how the completed standardized financial statements look like, see table C.4 and C.5 in Appendix C.2., where the balance sheet and profits & losses account of the bankrupt firm Air Holland are illustrated.

Consolidated balance sheet

Assets/Period	Liabilities/Period
Tangible fixed assets	Issued and paid up capital
Financial fixed assets	Reserves
Other fixed assets	Loss
Total Fixed assets	**Shareholders' equity**
Inventories	
Receivables	Provisions
Cash and cash equivalents	Long-term debt
Total Current assets	**Long-term liabilities**
	Short-term bank debt
Intangible assets	Other short-term debt
	Accounts payable
	Current liabilities
Other assets	**Other liablilities**
Total assets	**Total Liabilities**

Period	Period
Highest share price	Number of outstanding shares
Lowest share price	
Average share price	

Period	Nominal value share
Number of employees	

Table 4.1: Standardized balance sheet.

Consolidated profits & losses account

Period
Net sales (excluding turnover tax)
Cost of goods sold
Gross margin
Wages and salaries
Operating costs
Operating income (EBITDA)
Depreciation
Income from equity participations
Interest expenses
Result on ordinary activities
Extraordinary profits and losses: provisions
Other extraordinary profits and losses
Result before tax
Income tax expense
Result after tax
Minority interest
Net result

Table 4.2: Standardized profits & losses account.

4.1.4 Financial ratios

The financial performance of a company can be measured by looking at several financial ratios. The definition for a financial ratio given by Beaver (1966) is:

> "A financial ratio is a quotient of two numbers, where both numbers consist of financial statement items."

These financial ratios can be allocated to five groups, including liquidity, profitability, leverage, solvency and activity. Although financial ratios provide a fast and easy way to compare different companies, some caution has to be taken when using them. As described before, financial statements of companies can differ enormously, not only because of industry differences but also because of different notations in the financial statements.[7] The ratios used within this report are described in table 4.3[8]. Section 4.2

[7]This is the reason why the financial statements of our data set is normalized.

[8]These ratios follow from the data base and Master thesis made by Caljé (2000)

describes architecture of both two models constructed in this report, and will assigns a selection of the financial ratios to each of the models.

#	Abbreviation	Description of the fraction
1	WC/TA	Working capital/total assets
2	RE/TA	Retained earnings/total assets
3	EBIT/TA	Earnings before interest and taxes/total assets
4	MVE/TL	Market value equity/total debt
5	S/TA	Sales/total assets
6	LIQ/TA	Cash and cash receivables /total assets
7	(LTL+CL)/(LTL+CL+MVE)	(Long-term liabilities + current liabilities)/ (long-term liabilities + current liabilities + market value equity)
8	(LTL+CL)/TL	(Long-term liabilities + current liabilities) /total liabilities
9	MVE/TFA	Market value equity/total fixed assets
10	Employee change-ratio	((# employees year x) - (# employees year x-1))/(# employees year x-1)
11	(TFA+Inv.)/TA	(Total fixed assets + inventories)/total assets
12	IA/TA	Intangible assets/total assets
13	EBITDA growth-ratio	((EBITDA year x)-(EBITDA year x-1)) /EBITDA year x-1
14	Cash turnover rate	((Inventories + receivables - accounts payable) /(Sales)) * 12
15	(W&S+OC)/S	(Wages & salaries + operation costs)/Sales
16	MVE/(LTL+CL+MVE)	Market value equity/(long-term liabilities + current liabilities + market value equity)

Table 4.3: The 16 financial ratios used in this report. The abbreviation and the description of the fraction.

4.2 Architecture of the models

This section encapsulates the selection of the financial ratios, and a comprehensive explanation about the accuracy testing methods used for each of the two models. The construction and testing procedures of the models are accomplished in MATLAB 6.5[9], chiefly because MATLAB includes all the techniques used for this report. Besides that this program is rather simple to use. Subsection 4.2.1 starts with the architecture of a discriminant analysis based on the research done by Altman (1968). Subsection 4.2.2 follows with an overview of the architecture of the CART model constructed within this report.

[9]MATLAB, short for ”matrix laboratory”, refers to both the numerical computing environment and to its core programming language. Created by The MathWorks, MATLAB allows one to easily manipulate matrices, plot functions and data, implement algorithms, create user interfaces, and interface with programs in other languages. Although it specializes in numerical computing, an optional toolbox interfaces with the Maple symbolic engine, making it a full computer algebra system.

4.2.1 Classical model: Altman Z-score model

As mentioned before the Altman Z-score model forms a benchmark model within this report, because of that reason this Z-score model, as described in Altman (1993), keeps the same characteristics as Altman assigned to it in 1968. Therefore variable selection will not be utilized. Variable numbers 1-5 from table 4.3 form the independent variables.

To construct the discriminant analysis model based on the variables used by Altman (1968) the following action plan is very useful.

1. **Find the discriminant coefficients** They are determined by utilizing equations 3.2 and 3.3 as stated in subsection 3.1.1. Before identification of the discriminant coefficients we need to calculate the differences between the two group means and the pooled within-group covariance matrix.

2. **Create the discriminant function scores** Equation 3.1 finds the discriminant scores.

3. **Determine the cutoff score** The cutoff score has to be calculated to be able to classify the observations and to check the accuracy of the model which follow from the first step. (Equations 3.4 and 3.5 identify the correct cutoff score)

4. **Determine the accuracy** In this step the overall accuracy of the model will be be calculated. This is the fraction between the total number of correctly classified observations and the total number of observations.

5. **Create an accuracy matrix** This accuracy matrix, as reported in table 3.1, illustrates the classification of the observations perfectly.

6. **Test the accuracy** The accuracy, or hit rate, is a simple and intuitive approach to measure the goodness of fit for the classification model. Though, the question is; "is this accuracy percentage reasonable?" For this reason the accuracy should be compared to some sort of benchmark. To test the accuracy of the model the frequently used benchmark method ,the proportional chance criterion, is implemented. This method creates the following accuracy matrix based on the relative frequency with which each group appears in the data.

	Expected number classified by chance		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	np^2	$np(1-p)$	np
Non-Bankrupt	$np(1-p)$	$n(1-p)^2$	$n(1-p)$
All	np	$n(1-p)$	n

Table 4.4: The accuracy matrix based on the relative frequency with which each group appears in the data.

Where p is the change that an observation will be assigned to group 'bankrupt' and $1-p$ is the change the observation will be assigned to the group 'non-bankrupt'. The expected correctly classified observations which follow from the matrix 4.4 is

$$h_{expected} = np^2 + n(1-p)^2 \tag{4.1}$$

With this expected hits we can calculate expected hit ratio

$$hr_{expected} = \frac{h_{expected}}{n}, \tag{4.2}$$

which obviously return 50 percent for $hr_{expected}$ when the the group sizes are equal. With this $h_{expected}$ and $hr_{expected}$ we can do a *t*-test. This *t*-test calculates if the chance of coincidence of a certain accuracy will be negligible. In fact the accuracy is compared with the expected accuracy, based on the probability of classifying an observation in the right way. Before we can accomplish this *t*-test we need to calculate the last missing variable. The standard deviation of the number of expected hits from the proportional chance criterion. This standard deviation is given by

$$s_{CPRO} = \sqrt{nhr_{expected}(1 - hr_{expected})} \tag{4.3}$$

We can now execute the *t*-test, which is given by

$$t = \frac{h_{actual} - h_{expected}}{s_{CPRO}} \tag{4.4}$$

This aforementioned action plan will be done twice for the research of this report. The first time the model will be defined by the total data set. After the determination of this model, the performance will be tested on the same data.(Test data set and training data set are the same for this operation procedure) This procedure can produce potential bias, bias which is a result of overfitting toward the training data set. The second time leave-one-out cross validation will be practised.[10] This method is more time consuming, in fact it runs n discriminant analysis, but the accuracy percentage following from this procedure is far more valuable than from the previously mentioned operation procedure.

In Chapter 5 the results for the discriminant analysis prediction model will be described extensively. Besides that the encountered problems will be tagged, and the possible solutions for these problems will be expounded.

4.2.2 Classification and Regression Trees

The aforementioned procedure will partly be applicable for the CART model. Nevertheless there are some difference, for instance we can use more ratios with the CART model, to be exact we can use all the ratios as described in tables 4.3. Moreover the way to construct the model differs from the way to construct a discriminant analysis[11].

To make the benchmark of the two models more valuable, the CART model will be constructed twice. The first construction is based on the five ratios used by Altman (1968), the second construction will use most of the ratios as mentioned in table 4.3.

[10] See subsection 3.2.1 for a detailed explanation about leave-one-out cross validation.

[11] See section 3.2 for explanation about the CART model construction procedure.

For this report the Gini diversity index, as described in equation 3.16, is used for the impurity function. First of all the maximum trees are built. The pruning process starts after having constructed these maximum trees, which have an accuracy of 100% toward the training set. When these pruned trees are determined the action plan used for the discriminant analysis will carry on at step 4. From this point each observation will follow the tree and will be classified to one of the two possible groups. Each observation will encounter the different financial ratios which function as boolean questions. An example of such a question can be; WC/TA $\leq x$[12]? The CART algorithm will determine which ratio will settle in which node, and which value will be allocated to these financial ratios.

With the classifications of the individual observations we can determine the accuracies and accuracy matrices. Besides that the accuracy will be tested, in the same way as the accuracy of the discriminant analysis is tested. The CART models will be constructed twice, as we also did for the discriminant analysis. The first one is tested with the same data as used for training, the second way makes use of leave-one-out cross-validation.

The realization of the CART models and the test results will be described in the next Chapter. The encountered problems will be mentioned and solved whenever possible.

[12] x is a certain number, threshold, determined by the CART algorithm

Chapter 5

Results

In this Chapter, the results of the experiments, as described in the previous chapters, are discussed. The first section describes the the procedure and results for the discriminant analysis based on the Z-score model of Altman. The composition of section 5.2 is exactly the same, except for the fact that this section describes the results of the Classification and Regression Trees for the bankruptcy prediction problem. Finally, the Chapter will be concluded with a short comparison of both models.

5.1 Discriminant analysis: Based on the Z-score model of Altman

As described in subsection 3.1.1 the Z-score model of Altman is based on discriminant analysis. In this section we follow the action plan, which is discussed expansively in subsection 4.2.1, to construct and test the discriminant score model based on the same variables used by Altman (1968). But before we can start with this action plan, some problems about the data set have to be discussed.

The data set, as described in subsection 4.1.1, struggles with a lot of missing values. This has as a consequence that 26 out of the 122 companies in the data set misses information about their sales.[1] Another data set was created to cope with this problem. This new data set consists of 55 bankrupt and 41 non-bankrupt companies. Both data sets are used to create a discriminant analysis score function, and are tested in the following paragraphs.

5.1.1 DA test results for the data set consisting of 122 companies

Before starting to build and test the model it is helpful to visualize the data. Figure D.1[2] illustrates bivariate scatterplots and univariate histograms for each variable[3] for this data set. This figure helps to get a feeling for the data set.

Step one of the action plan starts at this point. (Produce the discriminant values, see tables 5.1) As mentioned in section 4.2.1 this action plan will be executed twice.

[1] Sales is a variables used for the fifth ratio in table 4.3, Sales / Total Assets

[2] Unfortunately it is not possible to show a 5 dimensional plot in this report. Figure D.1 functions as a substitute. Figure D.3 shows a multivariate plot for the data set.

[3] Financial ratio

The first results, which are shown below, were calculated by using the same training data set as test data set. The second results are based on leave-one-out cross-validation.

Discriminant coefficient values					
	Independent variables				
	WC/TA	RE/TA	EBIT/TA	MVE/TL	S/TA
k	2.2513	5.6977	11.0038	-0.4706	-0.0313
k determined by Altman (1968)	1,2000	1,4000	3,3000	0,6000	1,0000

Table 5.1: Discriminant coefficient values for the complete data set compared to the values for **k** determined by Altman.

The values determined for **k** are compared with the values defined by Altman (1968). The Altman values used in table 5.1 differ from the values mentioned on page 25. When using proportions like 0.10 for 10% instead of 10 for the independent variables, Altman proposed the values used in table 5.1.(See Altman (2000) page 13) As you can see, these value differ significantly, but overall the values are in proportion. The value defined for the fifth discriminant coefficient (S/TA) is substantially out of proportion, which in fact is as expected, because of the missing values in the data set for this financial ratio.

With these coefficients the discriminant function scores (**t**) can be determined. In order to assign a group tag to each of the observation, the correct cutoff score had to be defined. The cutoff score (t_c) for this data set is determined by formula 3.4.[4] If $t > t_c$ then this observation will fall into the non-bankrupt group. The calculated t_c for this data set is 1.7167. Accuracy matrix 5.2 was produced with this cutoff score, from which follows an accuracy of 77.05 %.

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	49	12	61
Non-Bankrupt	16	45	61
All	65	57	122

Table 5.2: Accuracy matrix for the total data set.

At this point only step 6 is left, in this step the accuracy of the model will be tested. Tables 5.3 shows the expected classifications.

[4]Formula 3.4 because both groups of the training data have the same size

	Expected number classified by chance		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	30.50	30.50	61.00
Non-Bankrupt	30.50	30.50	61.00
All	61.00	61.00	122.00

Table 5.3: The accuracy matrix based on the relative frequency with which each group appears in the data.

We can compare these expected classifications (table 5.3) with the actual classifications from table 5.2. This leads to a standard deviation of 5.5227 and a value of 5.9754 for the t-test[5], which in fact is significant compared with the expected accuracy[6].

The test result presented so far were gained by using the same training set as test set. This methodology can have overfitting toward the training set as a consequence. For this reason the same test was done for leave-one-out cross-validation. The results from this test are shown in the tables 5.4 and 5.5.

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	46	15	61
Non-Bankrupt	18	43	61
All	64	58	122

Table 5.4: Accuracy matrix for the total data set using leave-one-out cross-validation.

Results (leave-one-out cross-validation)			
	accuracy	s_{CPRO}	t-test
Values	72.95%	5.5227	5.070

Table 5.5: Test results for the complete data set based on leave-one-out cross-validation. (Where s_{CPRO} is the standard deviation)

Of course the standard deviation is the same as with the first test. On the other hand the accuracy and t-test results are lower, although the t-test still turns out significant compared with the expected accuracy.

5.1.2 DA test results for the data set consisting of 96 companies

For the adjusted data set we follow exactly the same procedure as for the complete data set. The only difference can be seen, as expected, in the test results. This data set obviously show better test results than the test results seen in the previous subsection.

[5] see formula 4.3 and 4.4

[6] See 4.2 for further details about this t-test.

Table 5.10 show accuracies of 89.58% and 85.42% respectively for the two ways of training and testing. (Test set is data set and leave-one-out cross-validation)

The accuracy of the leave-one-out cross-validation procedure points out that discriminant analysis for bankruptcy prediction, as introduced by Altman (1968), is robust to different data sets and works respectably accurate.

Table 5.6 shows the values determined for **k**. Again these values are compared with the values defined by Altman (1968). Also for this data set the values of **k** differ a lot, but are in proportion, even for the fifth discriminant coefficient.

Discriminant coefficient values					
	Independent variables				
	WC/TA	RE/TA	EBIT/TA	MVE/TL	S/TA
k	1.7819	5.8929	14.0613	-0.5744	0.4805
k determined by Altman (1968)	1,2000	1,4000	3,3000	0,6000	1,0000

Table 5.6: Discriminant coefficient values for the adjusted data set compared to the values for **k** determined by Altman.

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	53	2	55
Non-Bankrupt	8	33	41
All	61	35	96

Table 5.7: Accuracy matrix for the adjusted data set.

	Expected number classified by chance		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	31.5104	23.4896	55.00
Non-Bankrupt	23.4896	17.5104	41.00
All	55.00	41.00	96.00

Table 5.8: The accuracy matrix based on the relative frequency with which each group appears in the data. For the this adjusted data set, with different groups sizes.

Predicted group membership			
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	48	7	55
Non-Bankrupt	7	34	41
All	55	41	96

Table 5.9: Accuracy matrix for the adjusted data set using leave-one-out cross-validation.

Results for adjusted data set			
	accuracy	s_{CPRO}	t-test
Values test set is training set	89.58%	4.8979	7.5500
Leave-one-out cross-validation	85.42%	4.8979	6.7334

Table 5.10: Test results for the adjusted data set . (Where s_{CPRO} is the standard deviation)

5.2 Classification and Regression Trees

The test results for bankruptcy prediction using Classification and Regression Trees, the main subject of this report, are described in this section. Four tests are accomplished, two tests for each of the two data sets[7]. The first tests make use of the five financial ratios as described by Altman (1968), the second tests are based on the financial ratios as presented in tables 4.3. Unfortunately it was not possible to make use of all 16 financial ratios from tables 4.3, simply because the data set was incomplete for 5 financial ratios[8]. Because of this situation only 11 financial ratios could be used.

The first subsection shows the results for the total data set, the second subsection shows the results of the adjusted data set. As mentioned in subsection 4.2.2 the testing procedure of CART is largely the same as for Discriminant Analysis. For this reason the composition of the following subsection will show similarity with the previous two subsections.

5.2.1 CART test results for the data set consisting of 122 companies

First of all the CART algorithm has to built the maximum trees[9] as described in 3.2.1. The impurity function used for the tree growing process within this report is the Gini diversity index or Gini splitting rule. The “TREEFIT” function within MATLAB was used to construct these trees. Within the maximum tree the impure nodes must have 1 or more observations to be split. At the same time other trees were constructed, where the impure nodes must have 10 or more observations to be split, see figure E.3 and E.4. When we test the maximum tree and use the same test as training data set, obviously the accuracy will be 100%. This result is useless because it does not say anything

[7] Two for the complete data set and two for the adjusted data set. The complete data set consists of all 122 companies, the adjusted data set consists of 96 companies

[8] 5) S/TA, 10) Employee change ratio, 12) IA/TA, 14) Cash Turn , 15) (W&S+OC)/S , 16) W&S/Gross Margin

[9] See Appendix E figure E.1 and E.2

about the prediction capability of the tree. The other trees[10] are more useful, because these trees already show a little bit more generality.

To find the tree with the optimum generalization the "TREETEST" function within MATLAB was used. This function is able to find the number of terminal nodes which minimizes the error costs, and is able to produce a plot which shows the relationships between the *number of terminal nodes* and *the residual variance*. Figures 5.1 and 5.2 show that the best number of terminal nodes are only 2 for both trees. This can indicate that these CART models can be oversimplified for this prediction problem.

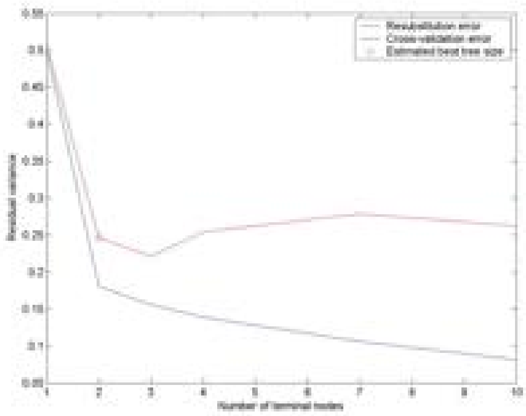

Figure 5.1: Best number of terminal nodes for the tree using the financial ratios introduced by Altman, when using the complete data set. Shows the relationship between the *number of terminal nodes* and *the residual variance*.

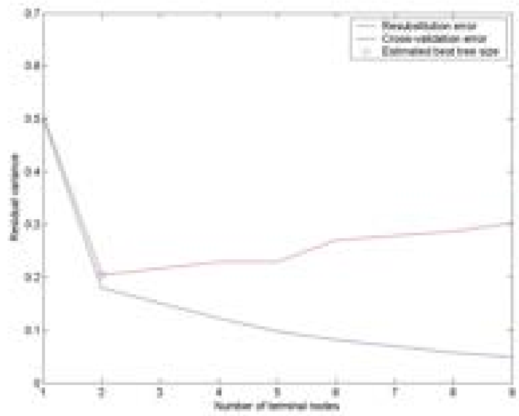

Figure 5.2: Best number of terminal nodes for the tree using the 11 financial ratios defined in table 4.3, when using the complete data set. Shows the relationship between the *number of terminal nodes* and *the residual variance*.

Figures E.5 and E.6 in Appendix E show the pruned trees for the complete data set. The first pruned tree is based on the 5 ratios as introduced by Altman (1968) the second is based on the 11 financial ratios as proposed in this report. As you can see, both trees are identical.

So far the trees have been constructed, and we have arrived at the point where the validation process of these trees can be started. In the following paragraphs the validation and test results are shown. For the validation process the "TREEVAL" function within MATLAB was used, this function returns a list of classes to which the tree assigns the individual observations to. From this list the prediction accuracy can be computed and tested.

[10] figure E.3 and E.4.

Test results using 5 financial ratios

This paragraph reports the test results for the complete data set using the 5 financial ratios introduced by Altman (1968). The composition of the paragraph is almost the same as section 5.1.2, except for the fact that there are 3 ways of training and testing. The first way makes use of tree E.3, the second makes use of the pruned tree E.5 and the third way makes use of leave-one-out cross-validation.

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	54	7	61
Non-Bankrupt	3	58	61
All	57	65	122

Table 5.11: CART accuracy matrix for the complete data set before pruning, using 5 ratios.

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	53	8	61
Non-Bankrupt	14	47	61
All	67	55	122

Table 5.12: CART accuracy matrix for the complete data set after pruning, using 5 ratios.

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	47	14	61
Non-Bankrupt	11	50	61
All	58	64	122

Table 5.13: CART accuracy matrix for the complete data set using leave-one-out cross-validation with 5 financial ratios.

Table 5.14 shows that the leave-one-out cross-validation method has an accuracy of 79.51% when making use of 5 financial ratios. To test the accuracy again the t-test was used in the same way the accuracy of the preface discriminant analysis was tested. The standard deviation is identical to the one calculated for the discriminant analysis, see 5.3 and 4.3. The t-test turned out to be significant.

Results for the complete data set using 5 ratios			
	accuracy	s_{CPRO}	t-test
Values test set is training set before pruning	91.80%	5.5227	9.2346
Values test set is training set after pruning	81.97%	5.5227	7.0618
Leave-one-out cross-validation	79.51%	5.5227	6.5186

Table 5.14: Test results for the complete data set using 5 ratios. (Where s_{CPRO} is the standard deviation)

Test results using 11 financial ratios

The tables below show the results for the same data set, the only difference is the number of financial ratios used, 11 instead of 5[11]. The prediction accuracy for the leave-one-out cross-validation with these 11 ratios is 78.13%, where the t-test again turned out to be significant.

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	59	2	61
Non-Bankrupt	4	57	61
All	63	59	122

Table 5.15: CART accuracy matrix for the complete data set before pruning, using 11 ratios..

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	53	8	61
Non-Bankrupt	14	47	61
All	67	55	122

Table 5.16: CART accuracy matrix for the complete data set after pruning, using 11 ratios.

[11] See table 4.3

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	53	8	61
Non-Bankrupt	13	48	61
All	66	56	122

Table 5.17: CART accuracy matrix for the complete data set using leave-one-out cross-validation with 11 financial ratios.

Results for the complete data set using 11 ratios			
	accuracy	s_{CPRO}	t-test
Values test set is training set before pruning	95.08%	5.5227	9.9589
Values test set is training set after pruning	81.97%	5.5227	7.0618
Leave-one-out cross-validation	78.13%	5.5227	6.2130

Table 5.18: Test results for the complete data set using 11 ratios. (Where s_{CPRO} is the standard deviation)

5.2.2 CART test results for the data set consisting of 96 companies

This subsection expounds the results for the adjusted data set, the composition of this subsection is exactly the same as the previous subsection. Figures 5.3, 5.4, E.11 and E.12 show that, just as with the total data set, the number terminal nodes which minimize the error is 2. The prediction accuracy for the leave-one-out cross-validation is 79.17% and 78.13% respectively when using 5 or 11 financial ratios.

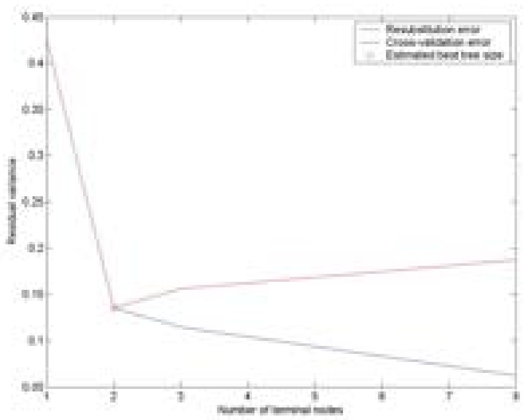

Figure 5.3: Best number of terminal nodes for the tree using the financial ratios introduced by Altman, when using the adjusted data set. Shows the relationship between the *number of terminal nodes* and *the residual variance*.

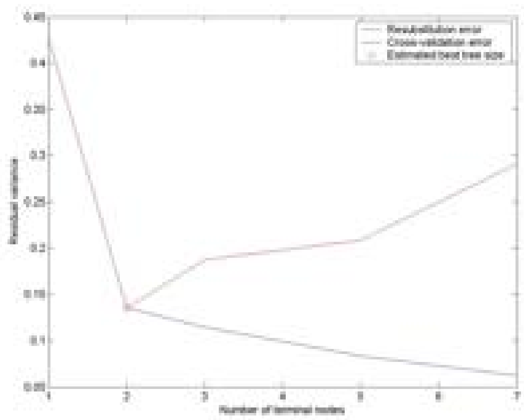

Figure 5.4: Best number of terminal nodes for the tree using the 11 financial ratios defined in table 4.3, when using the adjusted data set. Shows the relationship between the *number of terminal nodes* and *the residual variance*.

Test results using 5 financial ratios

The accuracy of the leave-one-out cross-validation for this test is 79.17%.

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	50	5	55
Non-Bankrupt	1	40	41
All	51	45	96

Table 5.19: CART accuracy matrix for the adjusted data set before pruning, using 5 ratios.

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	49	6	55
Non-Bankrupt	7	34	41
All	56	40	96

Table 5.20: CART accuracy matrix for the adjusted data set after pruning, using 5 ratios.

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	44	11	55
Non-Bankrupt	9	32	41
All	53	43	96

Table 5.21: CART accuracy matrix for the adjusted data set using leave-one-out cross-validation with 5 financial ratios.

Results for the adjusted data set using 5 ratios			
	accuracy	s_{CPRO}	t-test
Values test set is training set before pruning	93.75%	4.8979	8.3667
Values test set is training set after pruning	86.46%	4.8979	6.9375
Leave-one-out cross-validation	79.17%	4.8979	5.5083

Table 5.22: Test results for the adjusted data set using 5 ratios. (Where s_{CPRO} is the standard deviation)

Test results using 11 financial ratios

The last test results for this report are shown in this paragraph. The accuracy of the leave-one-out cross-validation is 78.13%, which is acceptable but not astonishing.

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	55	0	55
Non-Bankrupt	6	35	41
All	61	35	96

Table 5.23: CART accuracy matrix for the adjusted data set before pruning, using 11 ratios.

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	49	6	55
Non-Bankrupt	7	34	41
All	56	40	96

Table 5.24: CART accuracy matrix for the adjusted data set after pruning, using 11 ratios.

	Predicted group membership		
Actual Group Membership	Bankrupt	Non-Bankrupt	All
Bankrupt	42	13	55
Non-Bankrupt	8	33	41
All	50	46	96

Table 5.25: CART accuracy matrix for the adjusted data set using leave-one-out cross-validation with 11 financial ratios.

Results for the adjusted data set using 11 ratios			
	accuracy	s_{CPRO}	t-test
Values test set is training set before pruning	93.75%	4.8979	8.3667
Values test set is training set after pruning	86.46%	4.8979	6.9375
Leave-one-out cross-validation	78.13%	4.8979	5.3042

Table 5.26: Test results for the adjusted data set using 11 ratios. (Where s_{CPRO} is the standard deviation)

5.3 Comparison and summary

The following tables show the difference in prediction accuracy for all tested models. The Discriminant Analysis with the adjusted data set shows the best results, although the CART models show results which are more steady. CART shows significantly better results for the complete data set, which indicates that CART models can handle missing values more easily than discriminant analysis models[12]. From these results we can conclude that the model proposed by Altman (1968) works quite accurate, 85.42% is a relatively high percentage. Especially when we look at the quality of the data set, as described in section 4.1.1, this percentage is surprisingly high. The results for the CART models show respectable results as well. The fact that the results of the CART models using 11 ratios did not surpass the models which only used 5 ratios confirm how well and precise the 5 ratios were identified by Altman (1968). To compare our results with other results, the article written by Rahimian et al.[13] was used. This article showed test results for five different techniques with accuracies between 74.54% and 81.81%, which indeed are comparable to our results.

Results using Discriminant Analysis			
	accuracy	s_{CPRO}	t-test
Complete data set	72.95%	5.5227	5.070
Adjusted data set	85.42%	4.8979	6.7334

Table 5.27: Leave-one-out cross-validation test results for both data sets using Discriminant Analysis. (Where s_{CPRO} is the standard deviation)

Results using Classification and Regression Trees			
	accuracy	s_{CPRO}	t-test
Complete data set 5 ratios	79.51%	5.5227	6.5186
Complete data set 11 ratios	78.13%	5.5227	6.2130
Adjusted data set 5 ratios	79.17%	4.8979	5.5083
Adjusted data set 11 ratios	78.13%	4.8979	5.3042

Table 5.28: Leave-one-out cross-validation test results for both data sets using CART. (Where s_{CPRO} is the standard deviation)

[12] This was already mentioned on forehand as a strength of decision trees techniques on page 19.

[13] Source of information: Trippi and Turban (1996)

Chapter 6

Conclusions and Future Research

6.1 Conclusions

This report was devoted to bankruptcy prediction using CART models. In order to construct an accurate model, the "bankruptcy" event had to be stated precisely. Bankruptcies became common events since the economical depression of 1930. Bankruptcies have negative effects on company shareholders, creditors, employees, customers and even on national economies. For this reason bankruptcy prediction models were developed over the years. The Z-score model introduced by Altman (1968), which is based on discriminant analysis, is described in most literature as the most common used model for this problem.

The goal of this report was to accomplish a benchmark between a CART model and the Z-score model based on the financial input ratios as proposed by Altman (1968). The data set used for this report consisted of 122 companies, which all were or still are listed on the Amsterdam Stock Exchange. 61 of these companies went bankrupt during the period 1945-1999. The other half of the data set consists of 61 "matched" companies, which means that these companies are from the same industry, have the same size and were listed in the same period. The main problem of this data set is the fact that the time period is far to long. Financial situation can have complete different effects in different periods. Unfortunately each observation had to be treat similarly. This resulted in poorer or less reliable prediction power of our prediction models. Besides that, the data set struggled with a lot of missing values. For that reason 26 companies had to be deleted from the data set in order get a proper useful data set.

The Z-score model showed an accuracy of 85.42%, the CART model did not reach more than 79.17% prediction accuracy. Nevertheless both accuracies showed relatively accurate results, especially when we take the moderately poor data set into account. Above all this report showed the power of the Z-score model for this problem, and also proved the easiness to construct this model. For that reason it is very comprehensible that this model is so widely used over the past few decades for this prediction problem. The CART models produced for this report showed poorer results, but on the other hand they have a lot of potential. The construction process of CART models is harder to understand than the construction process of the Z-score model, although the output models are simple to read and give a helpful visual portrait of the classification problem.

6.2 Future research

There are three main issues for future research:

1. In this report most of the so far used bankruptcy prediction models were discussed. Subsection 2.2.2 presented a history/literature review for these models. Most of these models were based on the ideas of Beaver (1966), others were based on Artificial Intelligence (AI). These AI models are relatively new within the world of bankruptcy prediction, although a lot of these models show perfect characteristics for these problems. Future research about the quality of AI models for bankruptcy prediction could point out when to use which model.

2. CART is a classification technique which can function directly as a model, or can be used for structure identification for other techniques. They can perfectly identify the most relevant input variables for models which uses other techniques. A fuzzy system, also an AI technique, is a perfect example where the identification of these most relevant input variables is very helpful. Fuzzy systems prefer to have only a few input variables. The most important variables identified by the CART models, constructed for this report, can act as the input variables for new research for fuzzy systems for bankruptcy prediction.

3. Most of the discussed techniques for bankruptcy prediction also work for bond rating. Bond rating more a less is an expansion to bankruptcy prediction. Bond rating is about the chance that a firm can go bankrupt, therefore the addition of a probability parameter can turn these bankruptcy prediction models into bond rating models. A combination of a CART model, for input variable identification, and probabilistic fuzzy systems can be an excellent topic for future research.

Bibliography

E. Altman, R. Haldeman, and P. Narayanan. ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations. *Journal of Banking and Finance*, 1:29–54, 1977.

E.I. Altman. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4):569–608, September 1968.

E.I. Altman. Predicting financial distress of companies: Revisiting the z-score and zeta models. *Update of Journal of Finance 1968 and Journal of Banking & Finance 1997*, 2000.

E.I. Altman. *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting & Avoiding Distress and Profiting from Bankruptcy, Second Edition*. John Wiley & Sons, 1993.

W. Beaver. Financial Ratios as Predictors of Failures. In Empirical Research in Accounting, selected studies supplement to the. *Journal of Accounting research*, 1(4): 71–127, 1966.

E. Bernhardsen. A Model of Bankruptcy Prediction. , pages 1–54, 2001.

L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. *Classification and regression trees*. Chapman & Hall, Inc., 1984. ISBN 0-412-04841-8.

J.F. Caljé. Costs of Financial Distress and Bankruptcy in the Netherlands. Master's thesis, Erasmus University Rotterdam, October 2000.

E.B. Deakin. A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research*, 10(1):167–179, 1972.

Dun and Bradstreet. *Business Failure Record*. Dun & Bradstreet, 1980.

R.O. Edmister. An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. *The Journal of Financial and Quantitative Analysis*, 7(2):1477–1493, 1972.

H.D.Platt and M.B.Platt. Development of Stable Predictive Variables: The Case of Bankruptcy Prediction. *Journal of Business Finance and Accounting*, 17,1:31–44, 1990.

H.D.Platt and M.B.Platt. A Note on the Use of Industry-Relative Ratios in Bankruptcy Prediction. *Journal of Business Finance*, 1:1183–1194, 1991.

John H. Holland. *Adaptation in natural and artificial systems*. MIT Press, Cambridge, MA, USA, 1992. ISBN 0-262-58111-6.

J.-S.R. Jang, C.-T. Sun, and E. Mizutani. *Neuro-Fuzzy and Soft Computing*. Prentice Hall, 1997.

J. Lattin, J. Douglas Carroll, and P.E. Green. *Analyzing Multivariate Data*. Thompson, Duxbury, 2003. ISBN 0-534-34974-9.

R. Libby. Accounting Ratios and the Prediction of Failure: Some Behavioral Evidence . *Journal of Accounting Research*, 13(1):150–161, 1975.

T.E. McKee and M. Greenstein. Predicting bankruptcy using recursive partitioning and a realistically proportioned data set. *Journal of Forecasting*, 19:219–230, 2000.

T.M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.

J. Neter. Discussion of Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, 4(3):112–118, 1966.

M. Odom and R. Sharda. A neural network model for bankruptcy prediction. *Proceedings of the IEEE International Conference on Neural Network*, 2:163–168, 1990.

J.A. Ohlson. Financial Ratios and the Probabilistic Prediction of Bankruptcy . *Journal of Accounting Research*, 18(1):109–131, 1980.

J.R. Quinlan. Introduction of decision trees. *Machine Learning*, 1(1):81–106, 1986.

E. Rahimian, S. Singh, T. Thammachota, and R. Virmani. Bankruptcy prediction by neural network. Technical report.

D.E. Rumelhart, G.E. Hinton, and R.J. Williams. *Learning internal representations by error propagation, Chapter 8, Parallel distributed processing: explorations in the microstructure of cognition*, volume 1. MIT Press, 1986.

A. Santomero and J.D. Vinso. Estimating the Probability of Failure for Commercial Banks and the Banking System. *Journal of Banking and Finance*, 1977.

J. Scott. The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models. *Journal of Banking and Finance*, 5(3):317–344, 1981.

R.E. Shiffler and A.J. Adams. *Introductory Business Statistics with Computer Applications*. International Thompson Publishing, 2nd edition, 1995. ISBN 0-534-20946-7.

T.K. Sung, N. Chang, and G. Lee. Dynamics of Modeling in Data Mining: Interpretive Approach to Bankruptcy Prediction. , 1999.

R. Tan. Credit Rating Prediction. , supervisor: dr.ir. J. van den Berg, 2000.

R. Timofeev. Classification and regression trees (cart) theory and applications. Master's thesis, Humboldt University Berlin, 2004.

R.R. Trippi and E. Turban. *Nueral Networks in Finance and Investing*. McGraw-Hill, 1996.

L. Weiss. Bankruptcy Prediction: A Methodological and Empirical Update. *Freeman School of Business*, 1981.

R.W. White and M. Turnball. The Probability of Bankruptcy: American Railroads. *Working paper, Institute of Finance and Accounting, London Graduate School of Business*, 1975a.

R.W. White and M. Turnball. The Probability of Bankruptcy for American Industrial Firms. *Working paper*, 1975b.

J.W. Wilcox. A Simple Theory of Financial Ratios as Predictors of Failures. *Journal of Accounting Research*, 9(2):389–395, 1971.

C.V. Zavgren, M.T. Dugan, and J.M. Reeve. The Association between Probabilities of Bankruptcy and Market Responses: A Test of Market Anticipation. *Journal of Business Finance and Accounting*, 15:19–45, 1988.

Appendix A

Appendix A: Tables

A.1 Beaver's list of ratios tested

GROUP I (CASH-FLOW RATIOS)
1. Cash flow to sales
2. Cash flow to total assets
3. Cash flow to net worth
4. Cash flow to total debt

GROUP II (NET-INCOME RATIOS)
1. Net income to sales
2. Net income to total assets
3. Net income to net worth
4. Net income to total debt

GROUP III (DEBT TO TOTAL-ASSET RATIOS)
1. Current liabilities to total assets
2. Long-term liabilities to total assets
3. Current plus long-term liabilities to total assets
4. Current plus long-term plus preferred stock to total assets

GROUP IV (LIQUID-ASSET TO TOTAL-ASSET RATIOS)
1. Cash to total assets
2. Quick assets to total assets
3. Current assets to total assets
4. Working capital to total assets

GROUP V (LIQUID-ASSET TO CURRENT DEBT RATIOS)
1. Cash to current liabilities
2. Quick assets to current liabilities
3. Current ratio (current assets to current liabilities)

GROUP VI (TURNOVER RATIOS)
1. Cash to sales
2. Accounts receivable to sales
3. Inventory to sales
4. Quick assets to sales
5. Current assets to sales
6. Working capital to sales
7. Net worth to sales
8. Total assets to sales
9. Cash interval (cash to fund expenditures for operations)
10. Defensive interval (defensive assets to fund expenditures for operations)
11. No-credit interval (defensive assets minus current liabilities to fund expenditures for operations)

Figure A.1: List of ratios tested by Beaver (1966). The component of these ratios are defined in the following manner: cash flow - net income plus depreciation, depletion, and amortization; net worth - common stockholders' equity plus deferred income taxes; cash - cash plus marketable securities; quick assets - cash plus accounts receivable; working capital - current assets minus current liabilities; fund expenditures for operations - operating expenses minus depreciation, depletion, and amortization; and defensive assets - quick assets.

Appendix B

Appendix B: Algorithms

B.1 Recursive Partitioning Algorithms

B.1.1 ID3 algorithm

ID3, which stands for Interactive Dichotomizer 3, is a recursive partitioning algorithm developed by Quinlan (1986). The central question, as mentioned before in section 2.2.2, in the ID3 algorithm is selecting the variable that is most useful for classifying examples[1]. Information gain, is a statistical property term used to measure how well a variable separates the training examples. Entropy, a measure to describe the disorder in data, is used to describe information gain precisely. Entropy is used to assess the information value of each variable to develop the optimum splitting point for each variable. The variable which causes the highest reduction in entropy will be selected as the splitting point variable. This splitting point will partition the dataset into subsets. The same procedure will repeat recursively, with the variables left, until no feasible partitioning is possible. The algorithm thus creates a decision tree, where every node in the tree is divided into two branches which will lead to a leaf. This leaf has a label which can be seen as the dependent variable. The formula for entropy is:

$$Entropy(S) \equiv \sum_{i=1}^{c} -p_i \log_2 p_i \tag{B.1}$$

Given a collection S, consisting of positive and negative examples, where n is the number of possible classes, in the scope of this report $c = 2$ ($i = 1 \Rightarrow$ bankrupt, and $i = 2 \Rightarrow$ nonbankrupt), and p_i is the proportion of S belonging to class i.

So far we have calculated the entropy, which characterizes the impurity of the training examples in the dataset. Now we need a quantitative way of seeing the effect of splitting the dataset with a certain variable. We can use a measure called information gain, which calculates the reduction in entropy that would result on splitting the data on an attribute, A.

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \epsilon Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{B.2}$$

[1]this section is based on Mitchell (1997) and McKee and Greenstein (2000)

where $Values(A)$ is the set of all possible values for attribute A, and S_v is the subset of S for which attribute A has value v, so mathematically $S_v = \{s \epsilon S | A(s) = v\}$. The first part of the information gain formula (B.2) is the entropy for the total collection S, the second part is the expected entropy after partitioning collection S into subsets using attribute A. This expected entropy is the sum of the of entropies of all the subsets S_v, weighted by the fraction of examples $\frac{|S_v|}{|S|}$ that belongs to S_v. To summarize the ID3 algorithm[2]:

ID3(*Examples, Target_Attribute, Attributes*)
Examples are the training examples. Target_Attribute is the attribute whose value is to be predicted by the tree. Attributes is a list of other attributes that may be tested by learned decision tree. Returns a decision tree that correctly classifies the given examples.

- Create *Root* node for the tree
- If all *Examples* are positive, return the single-node tree *Root*, with label = +
- If all *Examples* are negative, return the single-node tree *Root*, with label = -
- If *Attributes* is empty, return the single-node tree *Root*, with label = most common value of *Target_Attribute in Examples*
- Otherwise Begin
  - $A \leftarrow$ the attribute from *Attributes* that best[3] classifies *Examples*
  - The decision attribute for *Root* $\leftarrow A$
  - For each possible value, v_i, of A,
    - Add a new tree branch below *Root*, corresponding to the best $A = v_i$
    - Let *Examples*$_{vi}$ be the subset of *Examples* that have value v_i for A
    - If *Examples*$_{vi}$ is empty
      - Then below this new branch add the leaf node with label = most common value of *Target_Attribute* in *Examples*
      - Else below this new branch add the subtree ID3(*Examples*$_{vi}$, *Target_Attribute*, *Attributes* - $\{A\}$)
- End
- Return *Root*

---

[2] literally copied from Mitchell (1997) page 56
[3] with the highest information gain as defined in equation B.2

B.2 Artificial Neural Networks

B.2.1 Backpropagation algorithm

Error propagation towards the input nodes - backpropagation - is by far the most popular training algorithm for ANNs.[4] It is only applicable in the case of continuous, differentiable activation functions like the logistic function. The 'net input' x_j of a node j is defined as the weighted sum of of the incoming signals plus a bias term.

$$x_j = \mathbf{x}_j^T \cdot \mathbf{w}_j, \tag{B.3}$$

where $\mathbf{x}_j$ is the vector of inputs and $\mathbf{w}_j$ is the vector of weights. The unit's output is then computed by using:

$$f(x_j) = f(x) = \frac{1}{1 + e^{-x_j}} \tag{B.4}$$

First a squared error measure for the p^{th} input-ouput pair is defined as

$$E_p = \sum_k (d_k - x_k)^2, \tag{B.5}$$

where d_k is the desired output for node k and x_k is the actual output for node k when the input part of the p^{th} data pair is presented. To find the gradient vector, an error term $\bar{\epsilon}_i$ for node i is defined as

$$\bar{\epsilon}_i = \frac{\partial^+ E_p}{\partial \bar{x}_i} \tag{B.6}$$

By the chain rule, the recursive formula for $\bar{\epsilon}_i$ can be written as

$$\bar{\epsilon}_i = \begin{cases} -2(d_i - x_i)\frac{\partial x_i}{\partial \bar{x}_i} = -2(d_i - x_i)x_i(1 - x_i) & \text{if node } i \text{ is a output node,} \\ \frac{\partial x_i}{\partial \bar{x}_i} = \sum_{j,i<j} \frac{\partial^+ E_p}{\partial \bar{x}_i}\frac{\partial \bar{x}_i}{\partial x_i} = x_i(1 - x_i)\sum_{j,i<j} \bar{\epsilon}_j w_{ij} & \text{otherwise,} \end{cases} \tag{B.7}$$

Next, the weights are updated using

$$\Delta \mathbf{w} = -\eta \frac{\partial^+ E}{\partial \mathbf{w}} = -\eta \nabla_{\mathbf{w}} E, \tag{B.8}$$

where $E = \sum_p E_p$.

This algorithm effectively minimizes the sum of squared errors E_p and therefore it creates the best possible model, given the training data and the chosen network structure.

[4] This paragraph is based on Jang et al. (1997)

B.3 Genetic Algorithm

Genetic algorithms are stochastic derivative free optimization methods based on the evolutionary theory of Darwin. The first research about these GAs was done by Holland (1992). Some of the advantages of GAs are discussed in section B.3.1, in this section some of the terms commonly used in GAs or equivalent techniques will be explained.

Chromosome A bit string which represent one possible solution.

Population A pool of chromosomes. GAs work with several chromosomes (possible solutions) simultaneously to increase the speed and to create offsprings for the new population.

Generation Each generation creates a new population. So every generation exists of one population which main role is to evaluate to a new population.

Fitness When a new population is created all chromosomes are evaluated. This evaluation simply means that the bit string, chromosome, is evaluated by its objective function.

Elitism Elitism is a form of selection. A small percentage of a population immediately flows to the next generation. Only the most fit chromosomes are allowed to do this, so only the elitism percentage part of the current population.

Selection The ones which will participate in creating the offspring for the next generation. Only the ones with high fitness will participate.

Crossover Crossover comes along in the part where the offspring is created. Chromosomes, which are not fit enough for the elitism percentage will be joined together with other chromosomes. (See figure B.1 for visual explanation)

Mutation When crossover does not accomplish the best solution in the solution space, for example a local minimum/maximum instead of a global minimum/maximum, the mutation operator can produce the satisfactory solution. It randomly mutate (flips) a bit in the bit string so a new chromosome will be the result.(see figure B.2 for a visual explanation)

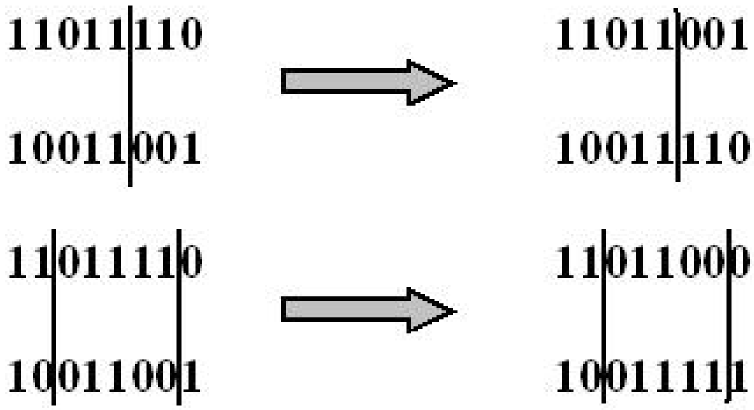

Figure B.1: One-point and 2-points crossover operators within GAs

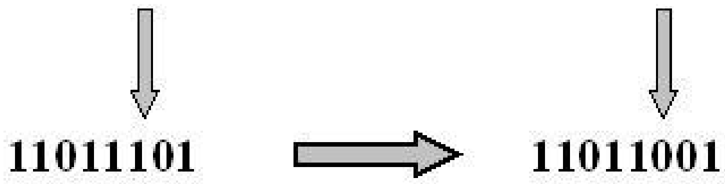

Figure B.2: Mutation operator within GAs

B.3.1 The algorithm

The summary of a standardized GA:[5] [6]

GA(Fitness, Fitness_threshold, p, r, m)
Fitness: A function that assigns an evaluation score, given a hypothesis.
Fitness_threshold: A threshold specifying the termination criterion.
p: The number of hypothesis to be included in the population.
r: The fraction of the population to be replaced by Crossover at each step.
m: The mutation rate.

- *Initialize population:* P $\leftarrow$ Generate p hypothesis at random
- *Evaluate:* For each h in P, compute *Fitness(h)*
- While $[\max_h$ *Fitness(h)*$]$ ¡ *Fitness_threshold* do
  Create a new generation P_s:
  1. *Select:* Probabilistically select $(1-r)p$ members of P to add to P_s. The probability Pr(h_i) of selecting hypothesis h_i from P is given by

[5] literally copied from Mitchell (1997) page 251
[6] see figure B.3 for a visualization of a standardized GA

$$Pr(h_i) = \frac{Fitness(h_i)}{\sum_{j=1}^{p} Fitness(h_j)}$$

2. *Crossover:* Probabilistically select $\frac{r \cdot p}{2}$ pairs of hypothesis from P, according to Pr(h_i) given above. For each pair, $\langle h_1, h_2 \rangle$, produce two offspring by applying the Crossover operator. Add all offspring to P_s.
3. *Mutate:* Choose m percent of the members of P_s with uniform probability. For each, invert one randomly selected bit in its representation.
4. *Update:* $P \leftarrow P_s$
5. *Evaluate:* For each h in P, compute *Fitness(h)*

- Return the hypothesis from P that has the highest fitness.

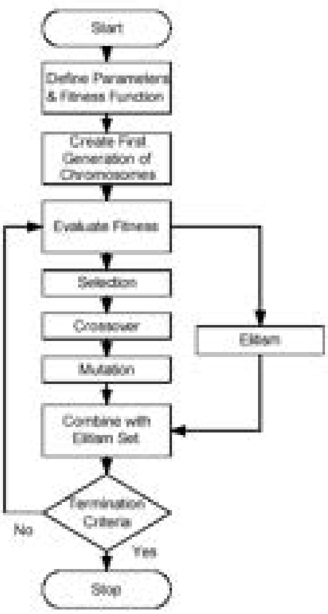

Figure B.3: Visualization of a standardized GA, copied from http://www.ece.concordia.ca/~kharma/ResearchWeb/html/research/ayo.html

Appendix C

Appendix C: Data set

C.1 Lists of bankrupt and non-bankrupt firms

[1]

Bankrupt firms	Presumably bankrupt	Non-bankrupt firms
C.V. Tieleman & Dros	1954	N.V. Hero Conserven Breda
Fridorfabrieken	1957	N.V. Haarlemsche Machinefabriek v/h Gebr. Figee
Klinker Isoliet B.V.	1957	N.V. Maats. Tot Expl. Van Steenfabrieken Udenhout
Ned. Kaiser-Frazer fabrieken NV	1959	N.V. Machinefabriek Breda voorheen Backer & Rueb
N.V. Allan & Co. Meubelen en Spoorwegmaterieel	1959	Orenstein en Koppel Spoorwegmaterieel
Arch N.V.	1964	Landr & Glinderman N.V.
Motorenfabriek Pluvier N.V.	1964	N.V. Gazelle Rijwelfabriek
Ubbink-Davo N.V.	1964	Kon. Fabr. Diepenbroek & Reigers N.V., Dru.
N.V. Verffabriek Avis	1966	Varossieau & Cie N.V.
Tricotagefabrieken v/h Frans Beeren & Zn. N.V.	1967	N.V. Stoomspinnerij Twenthe
Kunstzijdespinnerij Nijma N.V.	1967	N.V. Kon. Fabr. Van Verbandstoffen Utermohlen & Co.
Mulder-Vogem N.V.	1969	N.V. Kon. Metaalwaren Fabr. v/h J.N. Daalderop & Zonen

Table C.1: List of companies used for the data set of this report.

[1] This data set is created by Caljé (2000)

Bankrupt firms	Presumably bankrupt	Non-bankrupt firms
Hoogestraten Conserven N.V.	1969	Klene's Suikerwerkfabrieken N.V.
Entcomayer Handelsmaatschappij N.V.	1970	Hagemeijer Handelsmaatschappij
A.N. de Lint's beheermaatschappij N.V.	1970	Mees bouwmaterialen N.V.
Tricobest N.V. Kousenfabrieken	1971	N.V. Kon. D.S. van Schuppen & Zonen
N.V. Veneco	1972	N.V. Kon. Ned. Fabr. van Wollen dekens v/h Zaalberg
Kon. Ned. Textiel Unie N.V.	1972	Hatma-texoprint N.V.
Gebr. Gerzon's Modemagazijnen	1973	N.V. Kemo Hunkemoller Lexis
N.V. Gerofabriek	1973	N.V. Kon. Ned. Fabr. van Kempen & Begeer
Machinefabriek Reineveld N.V.	1973	N.V. Kon. Ned. Machinefabriek v/h E.H. Begemann
N.V. Ver. Ned. Kleermakerijen Gebr. Ibelings	1974	George Droge Textiel N.V.
Batava Margarine Fabrieken (Batava Marg. Works)	1975	Kwatta International N.V.
Enkes N.V.	1975	VMF Stork N.V.
Verenigde bedrijven Nederhorst N.V.	1975	Bredero N.V.
NV Kon. Delftsche Fabrk. De Porceleyne Fles	1976	Schuttersveld N.V.
Alg. Vruchten Import Maatsch. (AVIM N.V.)	1977	A.L. van Beek N.V.
Ver. Ned. Tapijtindustrie (VENETA) N.V.	1977	Tapijtfabriek Desseaux N.V.
Vulcaansoord N.V.	1977	Netam N.V.
N.V. Kon. Ned. Grofsmederij	1978	Koninklijke Hoogovens N.V.
Koninklijke Scholten-Honig N.V.	1978	De erven de Wed. J. Van Nelle N.V.
Noordelijke Industrie voor Vezelverwerking N.V.	1979	Drentsch Overijselsche Houthandel N.V.
Ned. Bontweverij N.V.	1980	Leidsche Wolspinnerij

Table C.2: List of companies used for the data set of this report.

Bankrupt firms	Presumably bankrupt	Non-bankrupt firms
Vandervliet-Werning Beheer N.V.	1980	Bam Holding N.V.
Duikers Apparatenfabriek	1981	Batenburg N.V.
G. Dikkers & Co. N.V.	1981	Grasso's Koninklijke Machinefabrieken N.V.
Asselbergs Holland N.V.	1981	VMF Stork N.V.
Papierfabrieken van Gelder & Zn. N.V.	1981	Kon. Ned. Papierfabrieken N.V. (KNP)
Bergos N.V.	1982	Tapijtfabriek Desseaux N.V.
OGEM Holding N.V.	1982	HBG
Vihamij Buttinger N.V.	1982	Nedap N.V.
Rijn-Schelde-Verolme N.V. (RSV)	1983	IHC Holland N.V.
N.V. Kon. D.S. van Schuppen & Zonen	1986	Gamma
Rademakers	1986	NAEFF
Leidsche Wolspinnerij N.V.	1987	Macintosch
Verenigde bedrijven Bredero N.V.	1987	HBG
Textlite Holding N.V.	1990	De Drie Electronics
Chamotte Unie	1991	ASR
Melia	1991	Center Parcs
Air Holland N,V,	1991	KLM
Infotheek Groep N.V.	1991	Volmac
Medicopharma N.V.	1991	Norit N.V.
Homburg Holding N.V.	1992	Alanheri
HCS Technology NV	1992	Getronics N.V.
United Dutch	1993	Burgman Heybroek
Verto	1993	Twentsche kabel N.V.
DAF N.V.	1993	DSM
Palthe N.V.	1993	Holland Colours N.V.
Berkel's Patent N.V.	1993	NKF
van Besouw N.V.	1995	Gamma N.V.
Wyers Beheer N.V.	1995	Blydenstein Willink N.V.
Fokker N.V.	1996	Schuttersveld N.V.
Tulip Computers N.V.	1998	LCI Technology N.V.

Table C.3: List of companies used for the data set of this report.

C.2 Financial statements

C.2.1 Balance sheet

Consolidated balance sheet (in thousands of NLG)

Assets/Period	1991	1990	1989	1988	1987	Liabilities/Period	1991	1990	1989	1988	1987
Tangible fixed assets	100845	200148	102292	104906	47679	Issued and paid up capital	16755	16755	16255	16255	11455
Financial fixed assets						Reserves	7164	7164	6590	6590	
Other fixed assets	35	28339	30439	24112	35	Loss	-19666	13668	9310	3785	-1240
Total Fixed assets	**100880**	**228487**	**132731**	**129018**	**47714**	**Shareholders' equity**	**4253**	**37587**	**32155**	**26630**	**10215**
Inventories	87952	1865	1980	2841	611						
Receivables	20508	16954	15557	9108	16647	Provisions	13048	9625	5629	169	6006
Cash and cash equivalents	6164	2884	9491	19988	77	Long-term debt	63929	142823	54674	80230	17363
Total Current assets	**114624**	**21703**	**27028**	**31937**	**17335**	**Long-term liabilities**	**76977**	**152448**	**60303**	**80399**	**23369**
						Short-term bank debt	86100	18942	17030	3380	13042
Intangible assets	**6499**	**3478**	**2142**	**1805**	**820**	Other short-term debt	36152	18540	25102	33438	11413
						Accounts payable	18521	7421	6182	4800	2130
						Current liabilities	**140773**	**44903**	**48314**	**41618**	**26585**
Other assets						**Other liabilities**		**18730**	**21129**	**14113**	**5700**
Total assets	**222003**	**253668**	**161901**	**162760**	**65869**	**Total Liabilities**	**222003**	**253668**	**161901**	**162760**	**65869**

Period	1991	1990	1989	1988	1987	Period	1991	1990	1989	1988	1987
Highest share price	15	33,6	37,4	35	35	Number of outstanding shares	3351000	3351000	650200	650200	458200
Lowest share price	0,2	9	31,3	35	35						
Average share price	**7,6**	**21,3**	**34,35**	**35**	**35**						

Period	1991	1990	1989	1988	1987	Nominal value share	fl 5,-	Aanpassen ratio's ivm IPO in 1989.
Number of employees	275	228	176	135	116		fl 25,-	

Table C.4: Balance sheet Air Holland. (Air Holland was declared bankrupt in 1991)

C.2.2 Profits & losses account

Consolidated profit & losses account (in thousands of NLG)

Period	1991	1990	1989	1988	1987
Net sales (excluding turnover tax)	135808	134422	110992	81074	59770
Cost of goods sold					
Gross margin	**135808**	**134422**	**110992**	**81074**	**59770**
Wages and salaries	17075	13848	12496	8312	6825
Operating costs	121757	102272	77169	60053	42053
Operating income (EBITDA)	**-3024**	**18302**	**21327**	**12709**	**10892**
Depreciation	9169	5725	4492	6189	5971
Income from equity participations					
Interest expenses	4797	2231	2435	1495	1365
Result on ordinary activities	**-16990**	**10346**	**14400**	**5025**	**3556**
Extraordinary profits and losses: provisions	32014				1114
Other extraordinary profits and losses					
Result before tax	**-49004**	**10346**	**14400**	**5025**	**2442**
Income tax expense	-7242	3073	4169		
Result after tax	**-41762**	**7273**	**10231**	**5025**	**2442**
Minority interest					
Net result	**-41762**	**7273**	**10231**	**5025**	**2442**

Table C.5: Profit & losses account Air Holland.

Appendix D

Appendix D: Scatters and plots

D.1 Discriminant analysis

D.1.1 Bivariate scatterplots between the five input variables for the Altman model

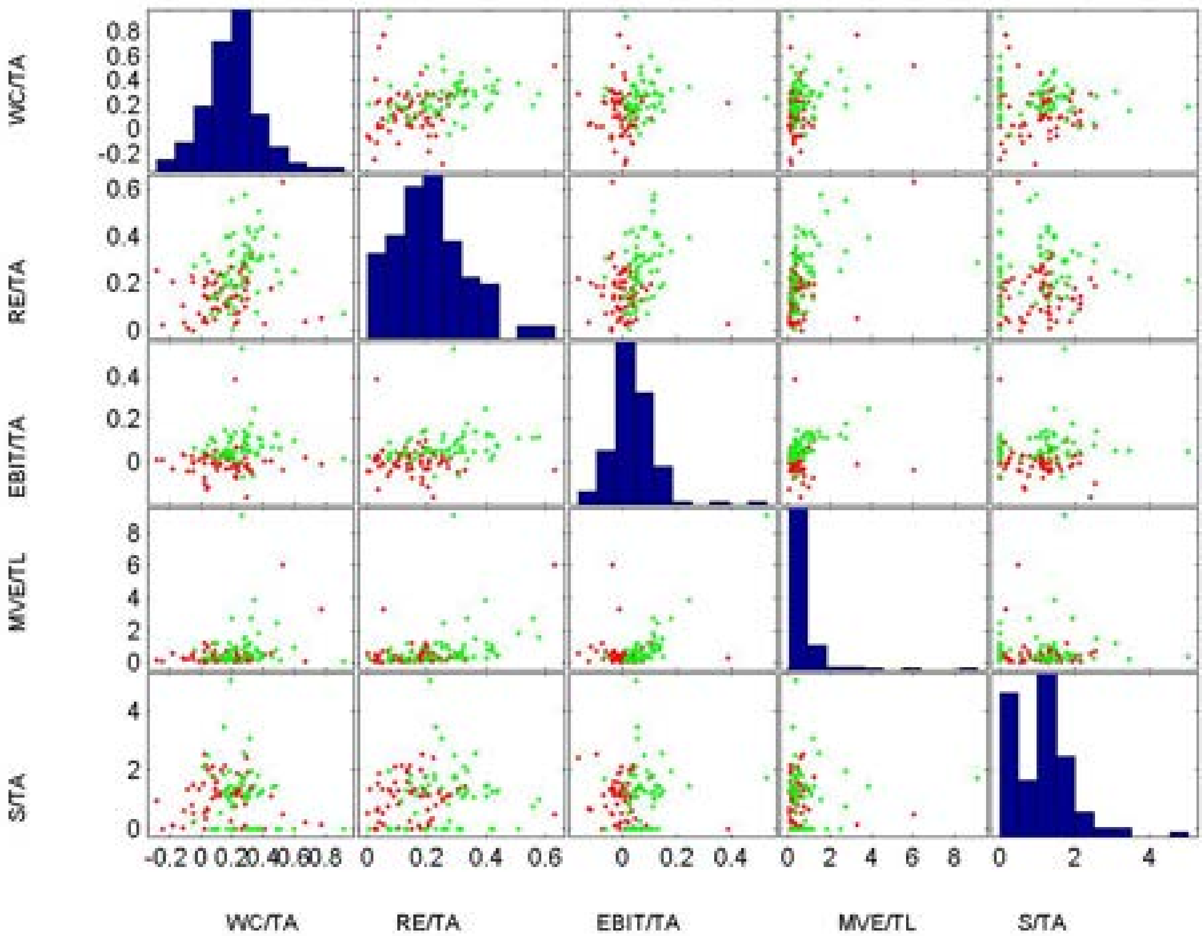

Figure D.1: Bivariate scatterplots for complete data set between our five variables, along with a univariate histogram for each variable.(Red is *Bankrupt*, green in *Non-Bankrupt*)

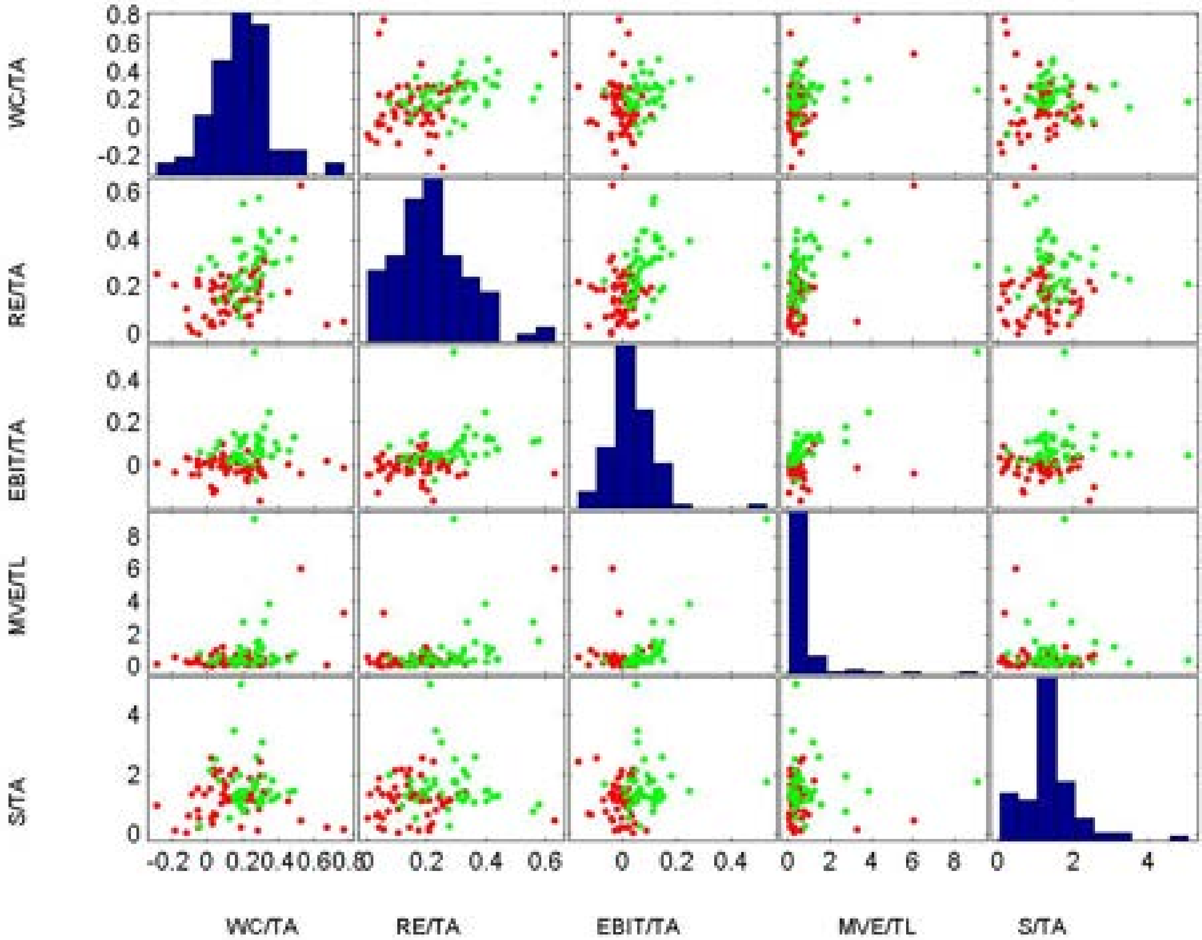

Figure D.2: Bivariate scatterplots for adjusted data set between our five variables, along with a univariate histogram for each variable. (Red is *Bankrupt*, green in *Non-Bankrupt*)

D.1.2 Multivariate plots for the model based on the variables used by Altman

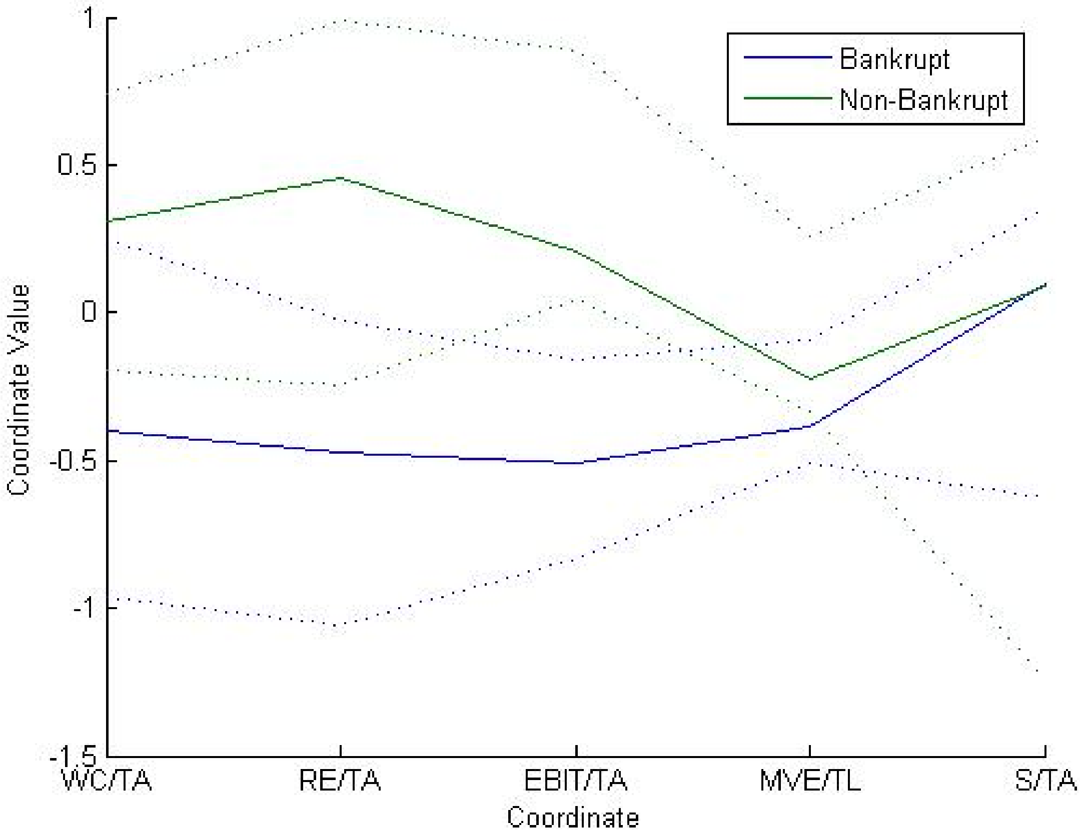

Figure D.3: Multivariate plot for complete data set. In this plot, the coordinate axes are all laid out horizontally, instead of using orthogonal axes as in the usual Cartesian graph. Only the median and quartiles (25% and 75% points) for each group are shown.

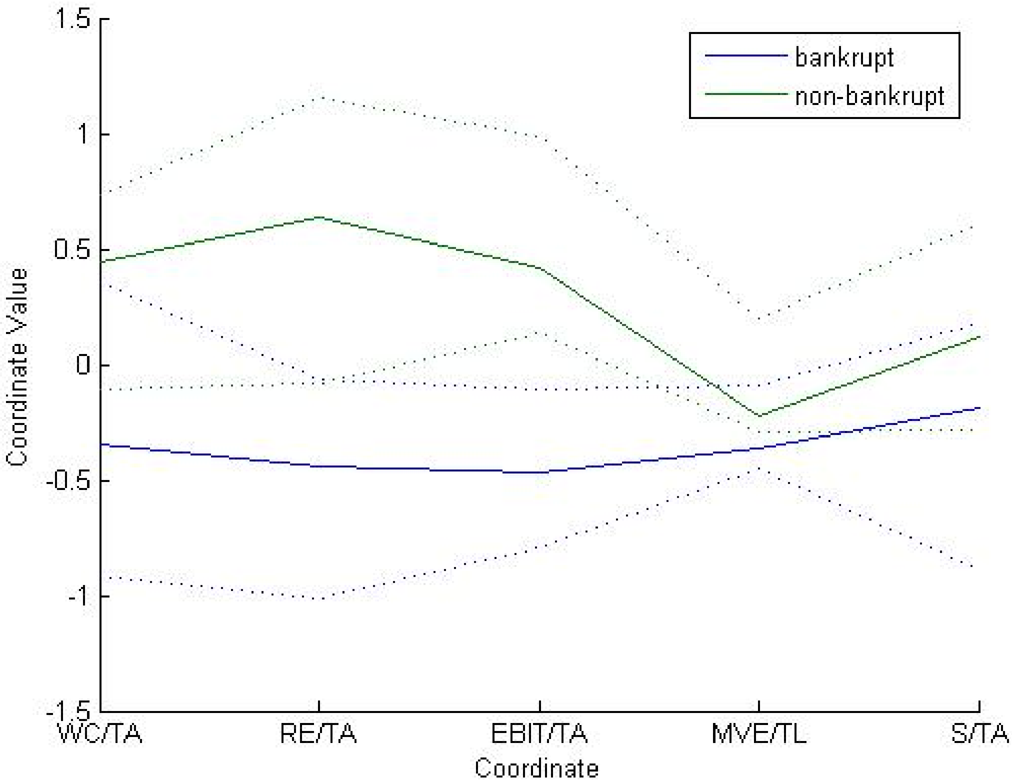

Figure D.4: Multivariate plot for adjusted data set. In this plot, the coordinate axes are all laid out horizontally, instead of using orthogonal axes as in the usual Cartesian graph. Only the median and quartiles (25% and 75% points) for each group are shown.

Appendix E

Appendix E: Classification and Regression Trees

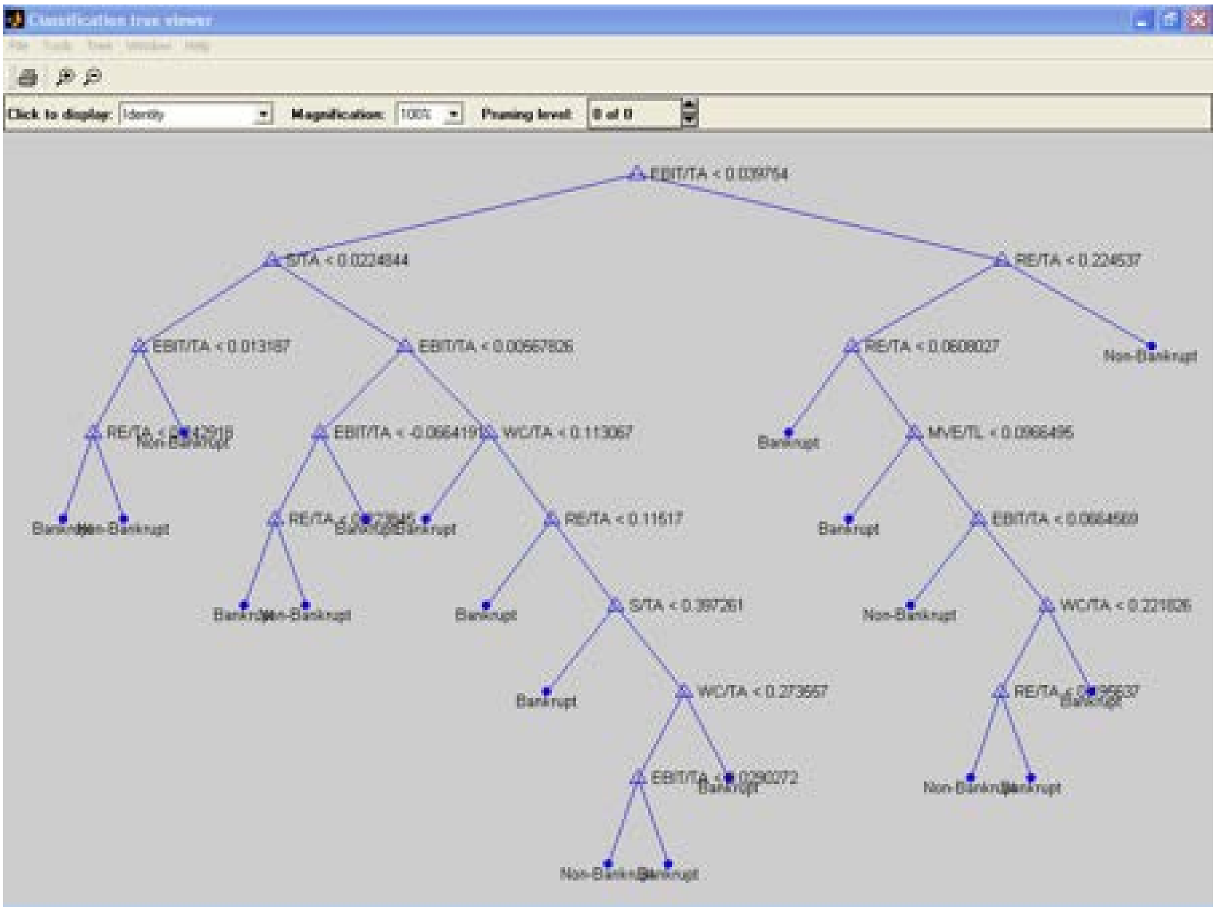

Figure E.1: Maximum tree for the complete data set using 5 ratios.

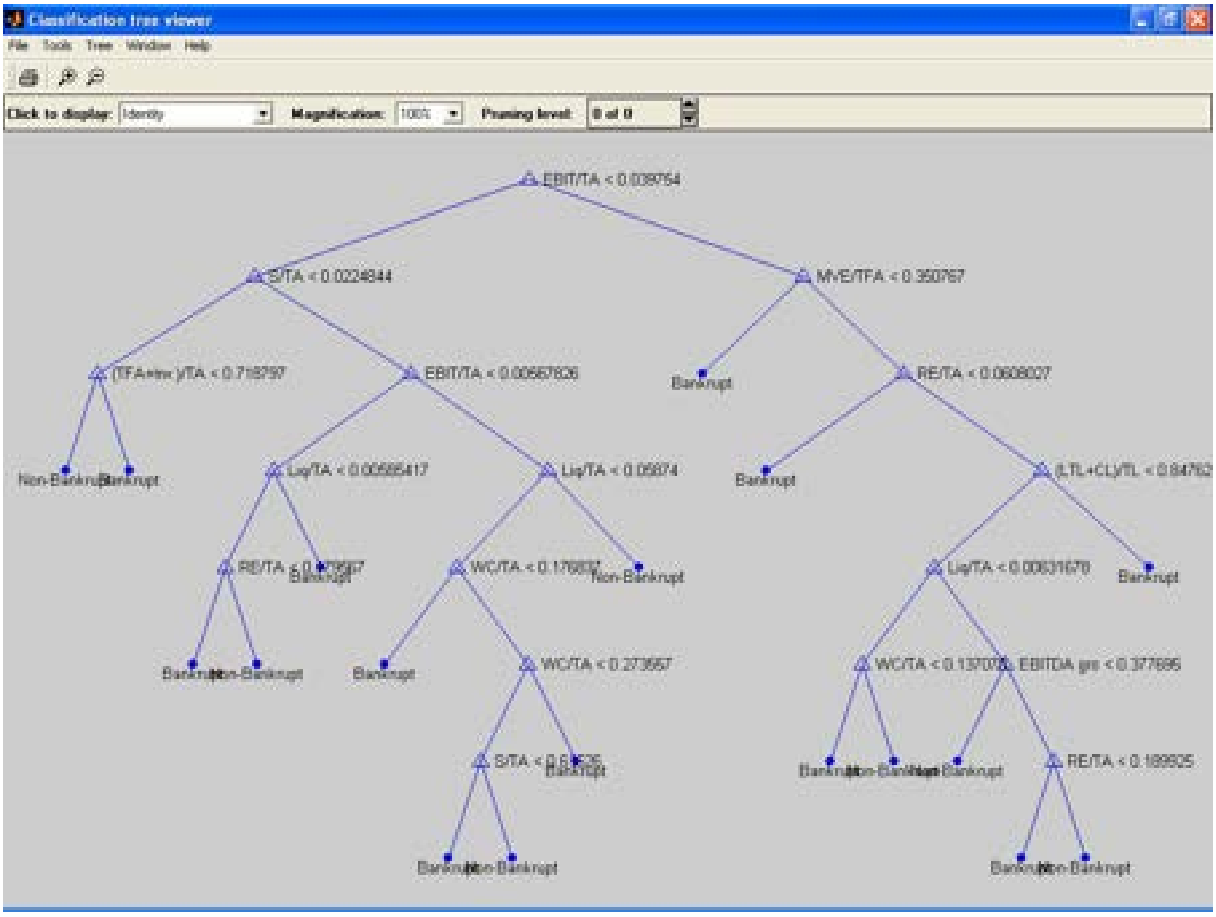

Figure E.2: Maximum tree for the complete data set using 11 ratios.

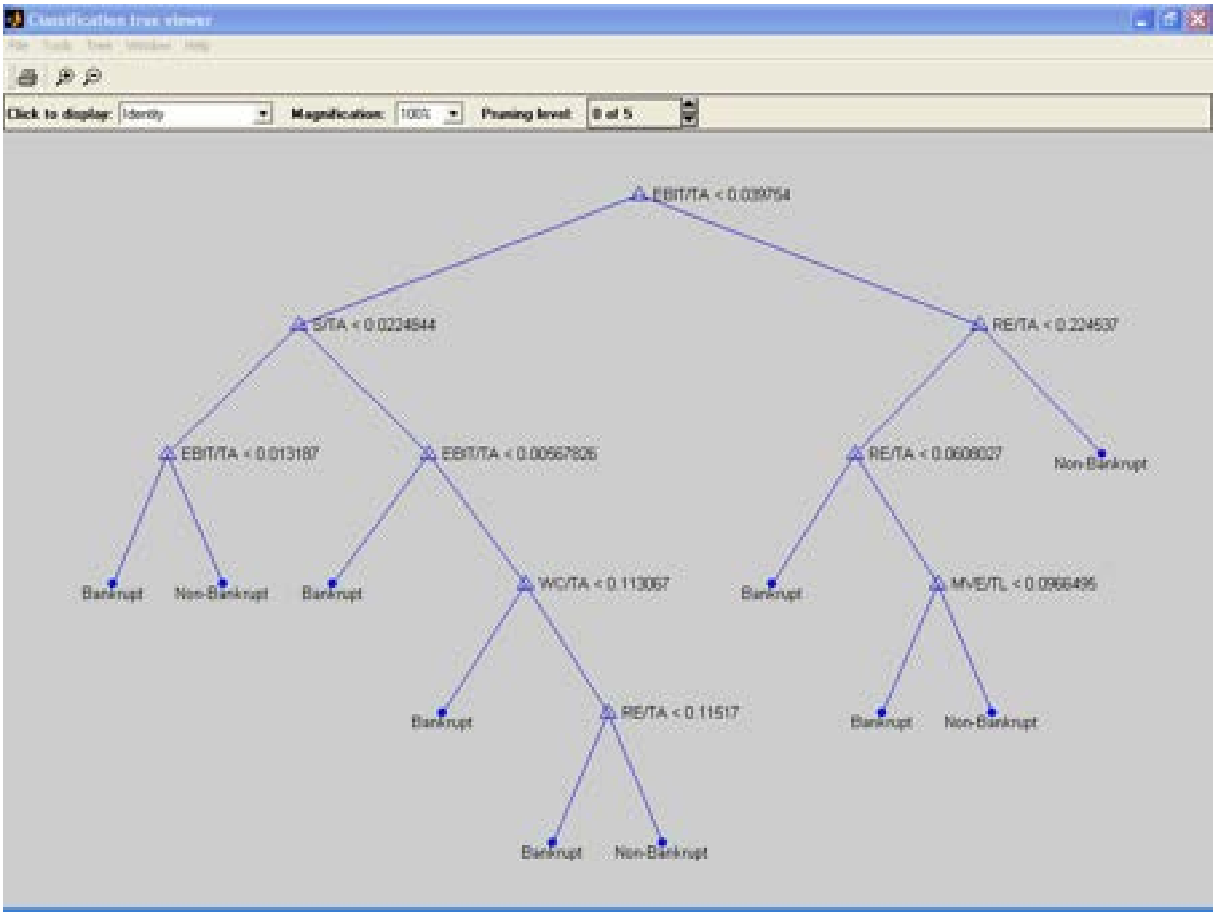

Figure E.3: Tree for the complete data set using 5 ratios. For this tree the impure nodes must have 10 or more observations to be split

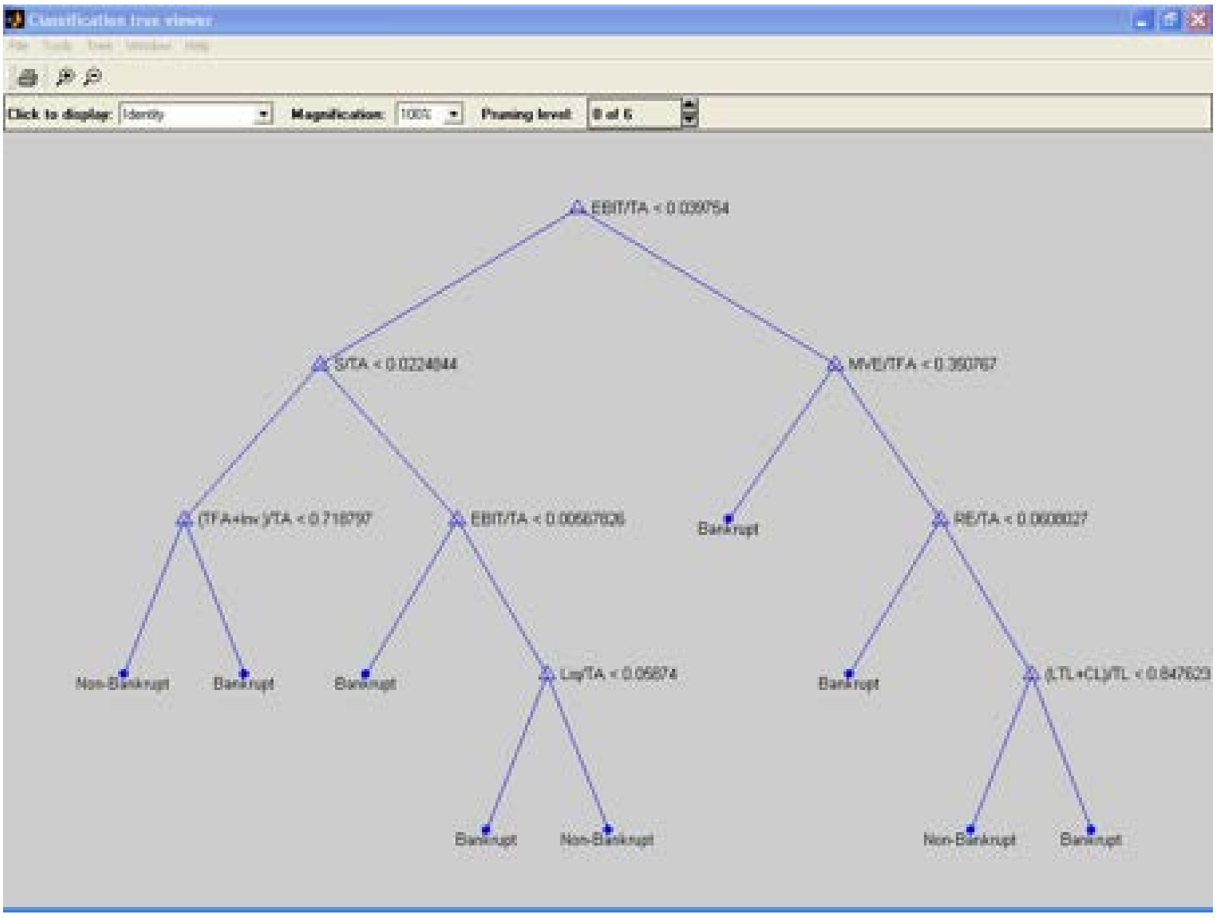

Figure E.4: Tree for the complete data set using 11 ratios. For this tree the impure nodes must have 10 or more observations to be split

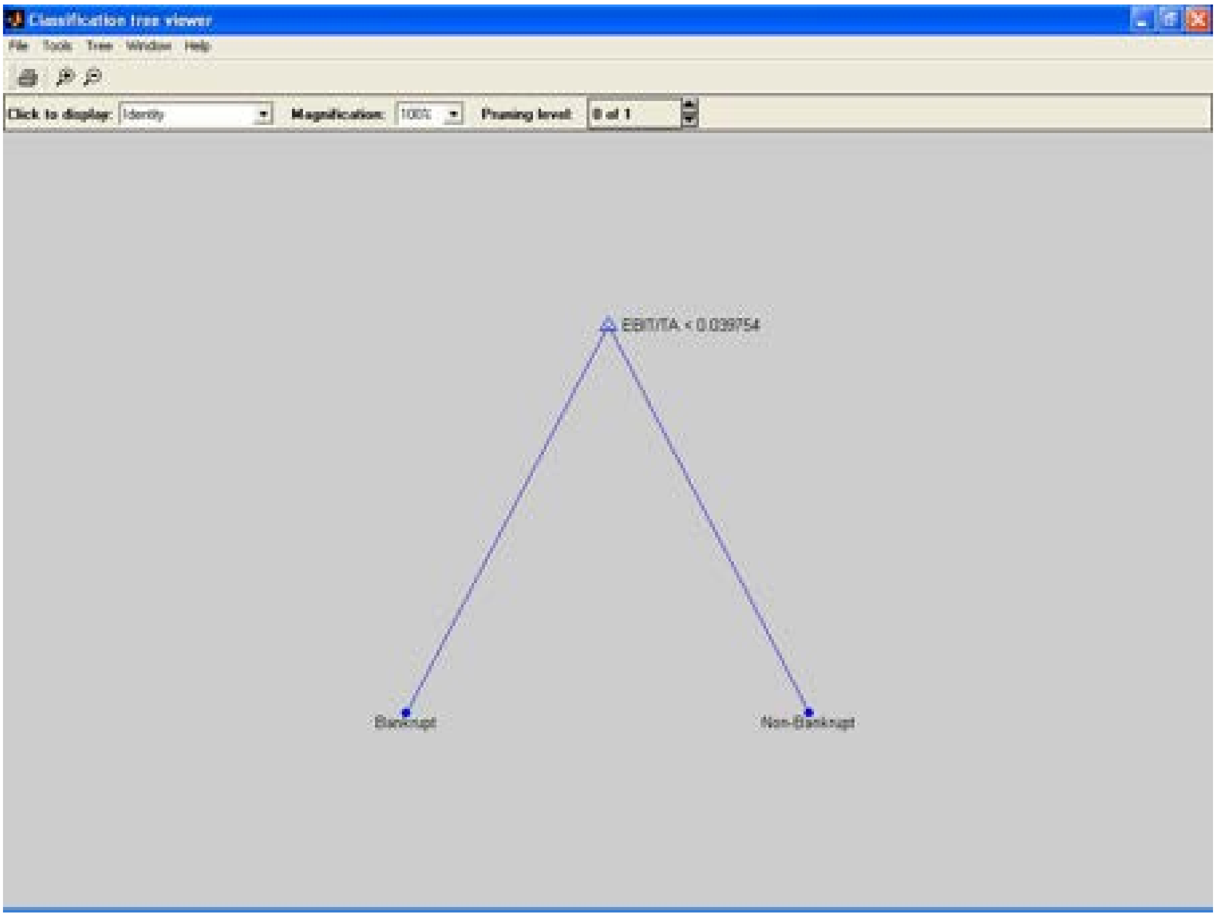

Figure E.5: Pruned tree for the complete data set using 5 ratios

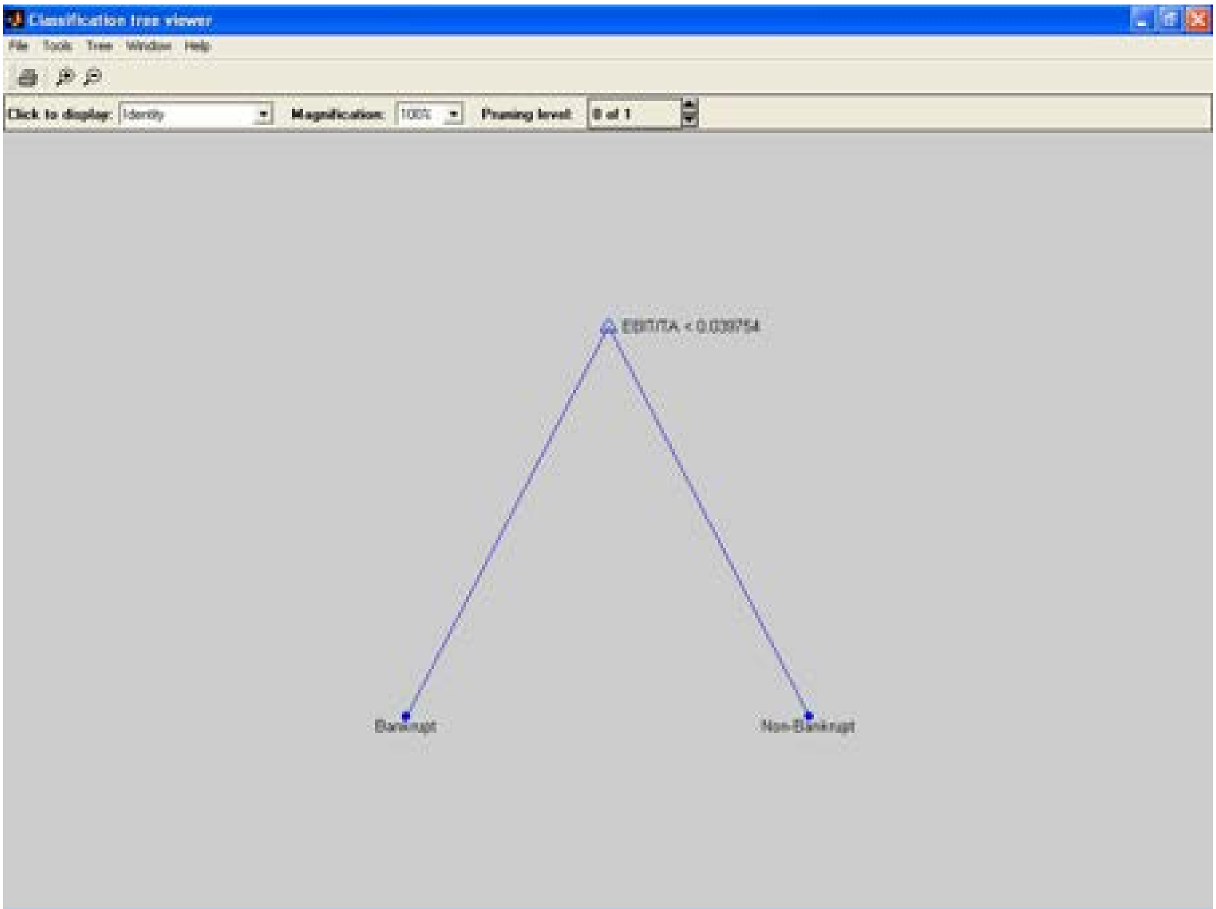

Figure E.6: Pruned tree for the complete data set using 11 ratios

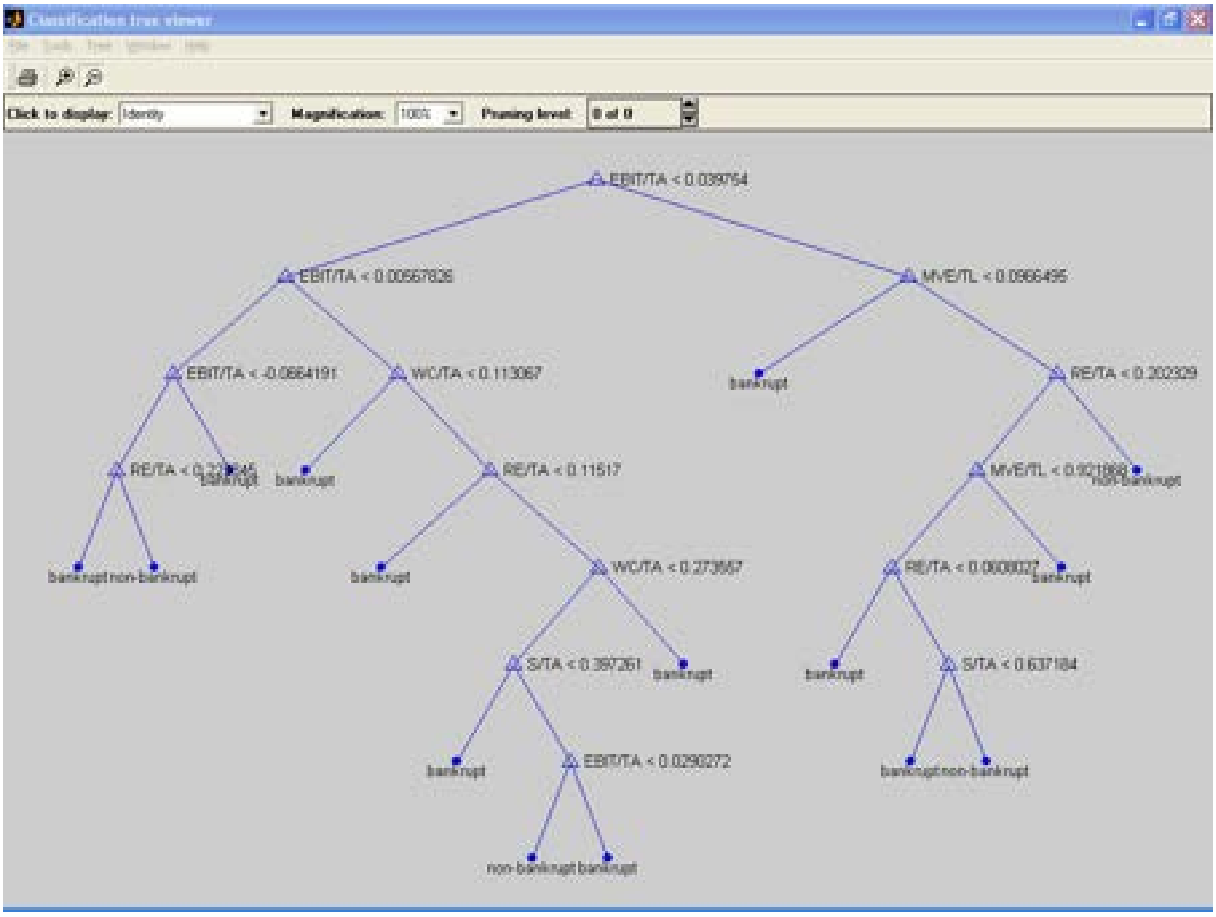

Figure E.7: Maximum tree for the adjusted data set using 5 ratios.

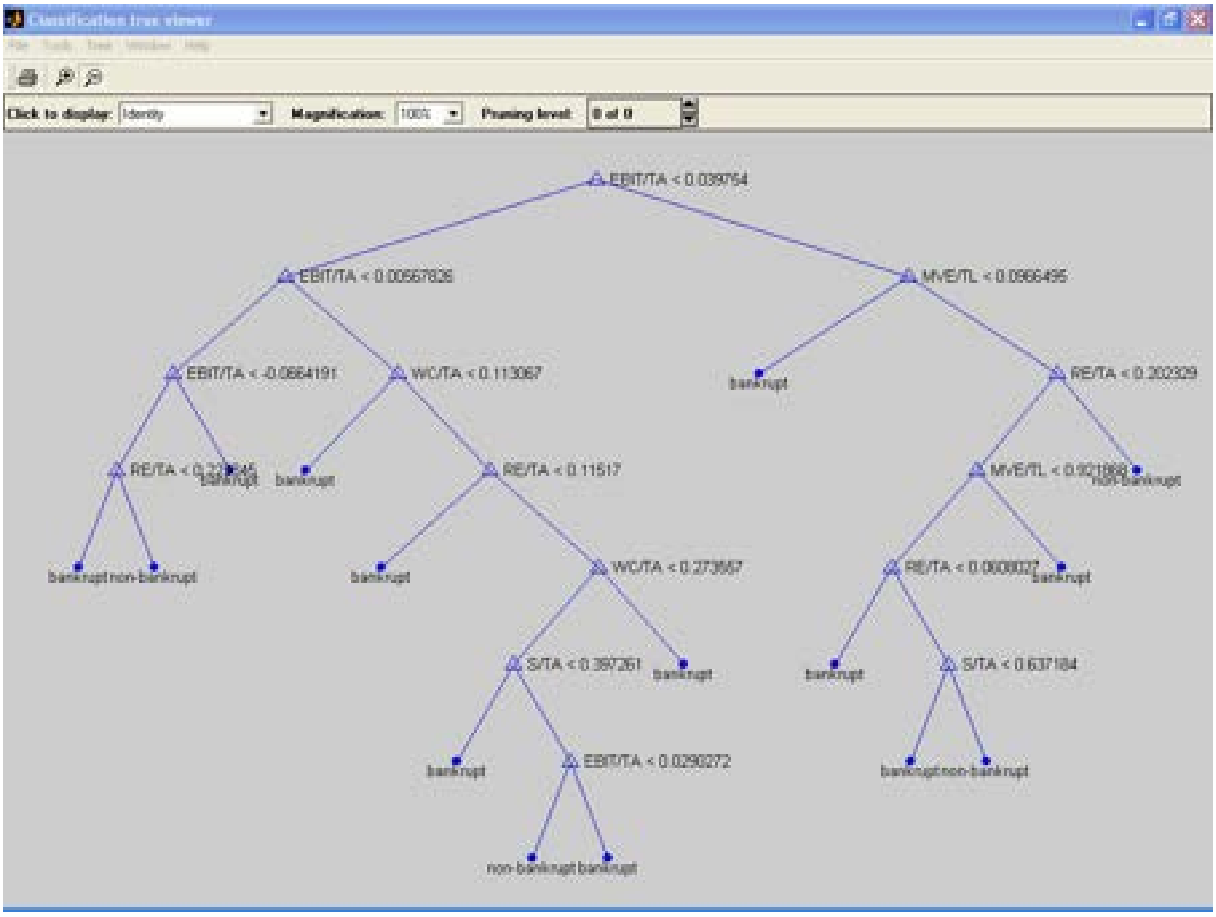

Figure E.8: Maximum tree for the adjusted data set using 11 ratios.

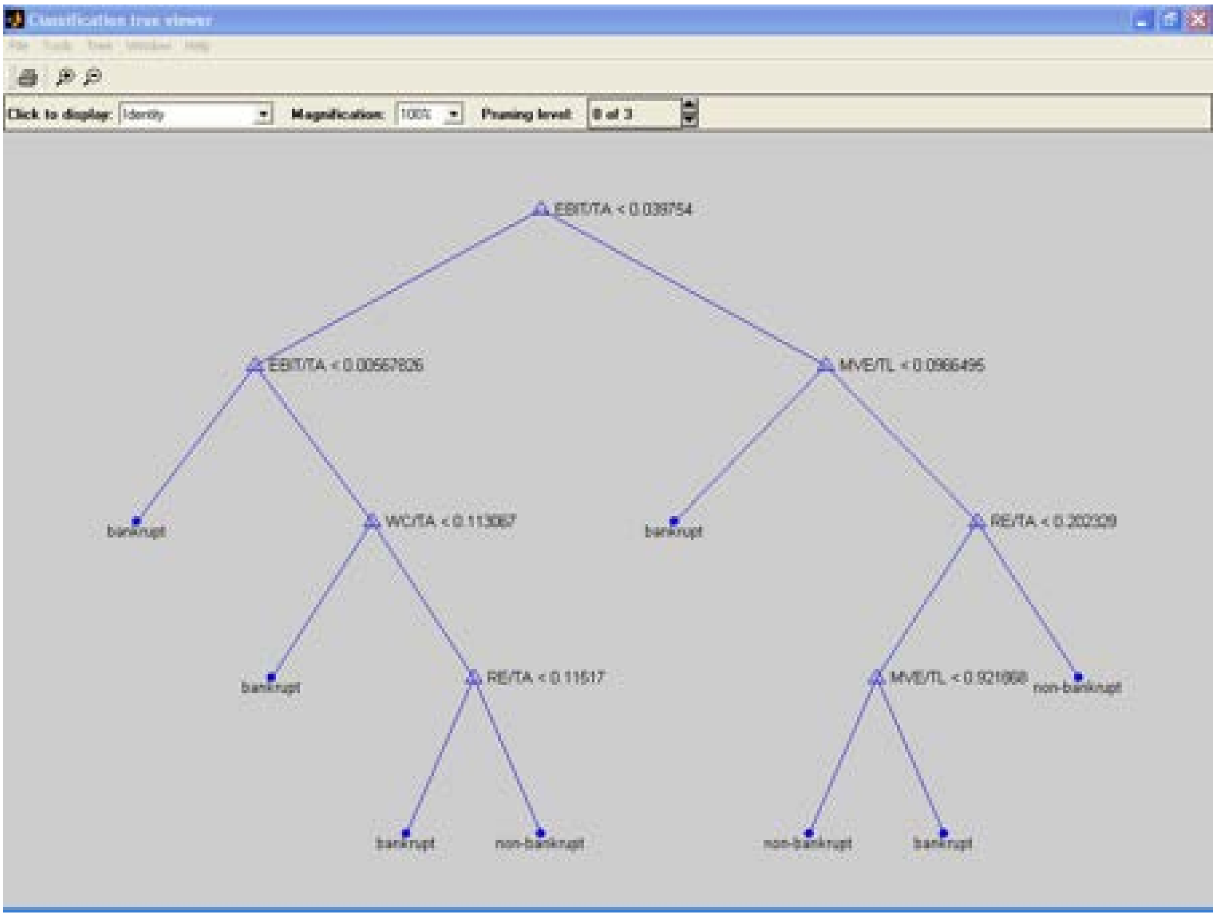

Figure E.9: Tree for the adjusted data set using 5 ratios. For this tree the impure nodes must have 10 or more observations to be split.

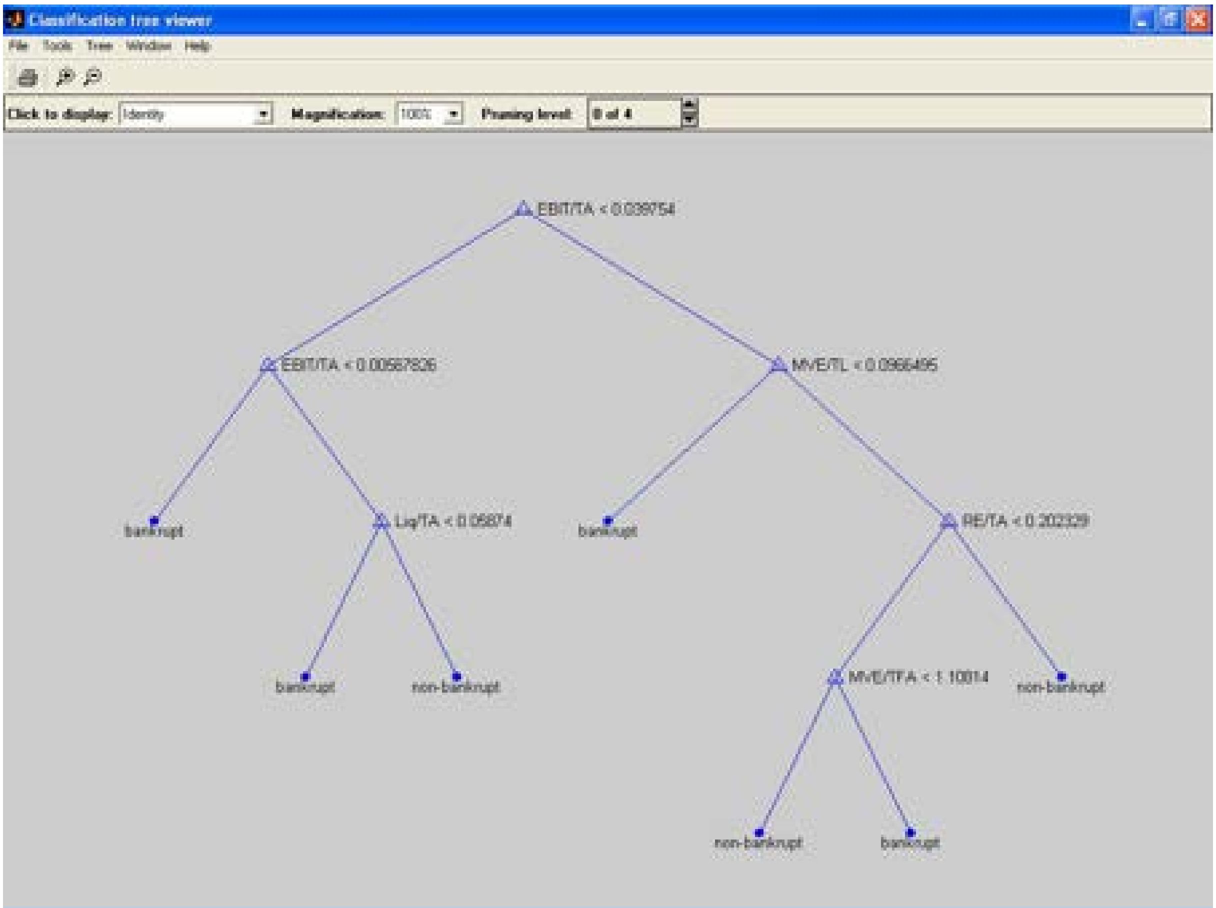

Figure E.10: Tree for the adjusted data set using 11 ratios. For this tree the impure nodes must have 10 or more observations to be split.

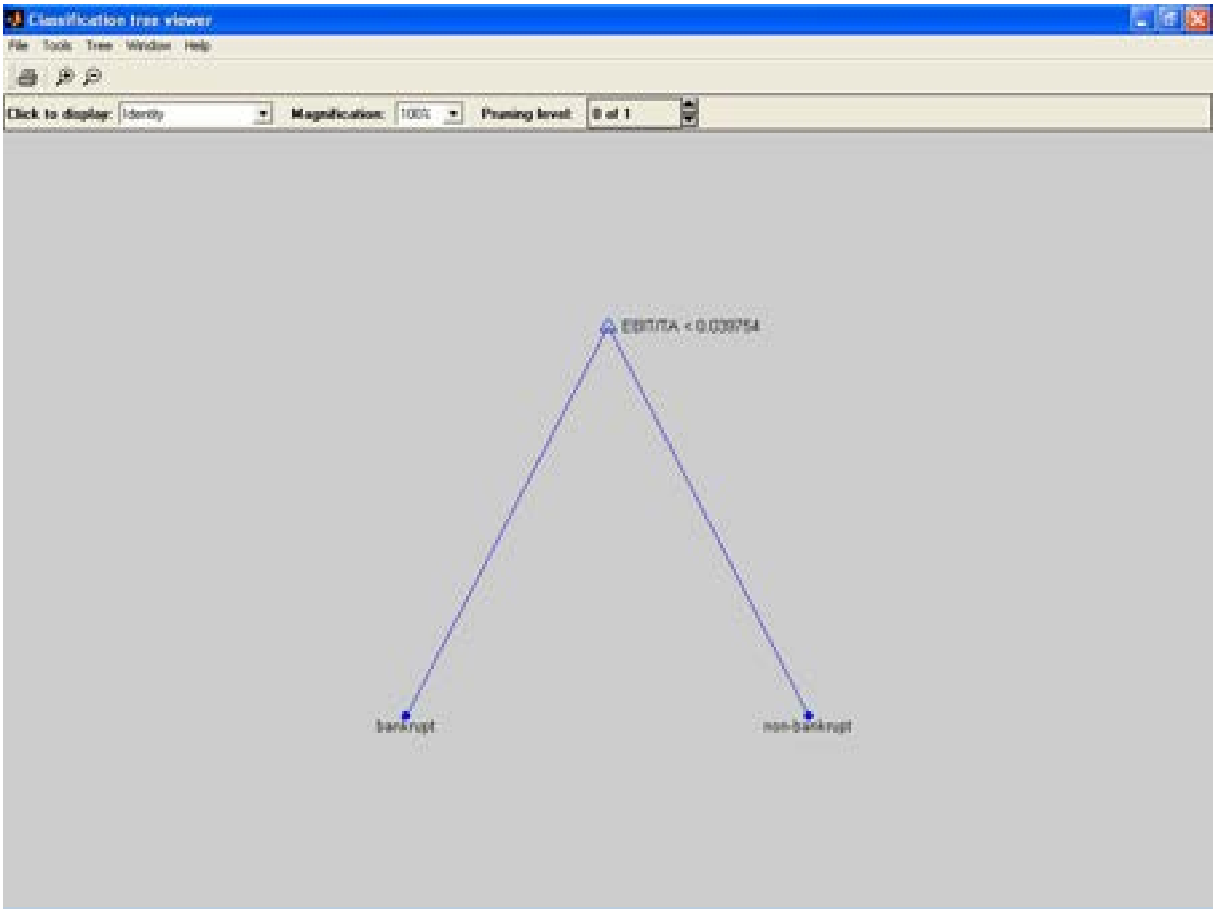

Figure E.11: Pruned tree for the adjusted data set using 5 ratios

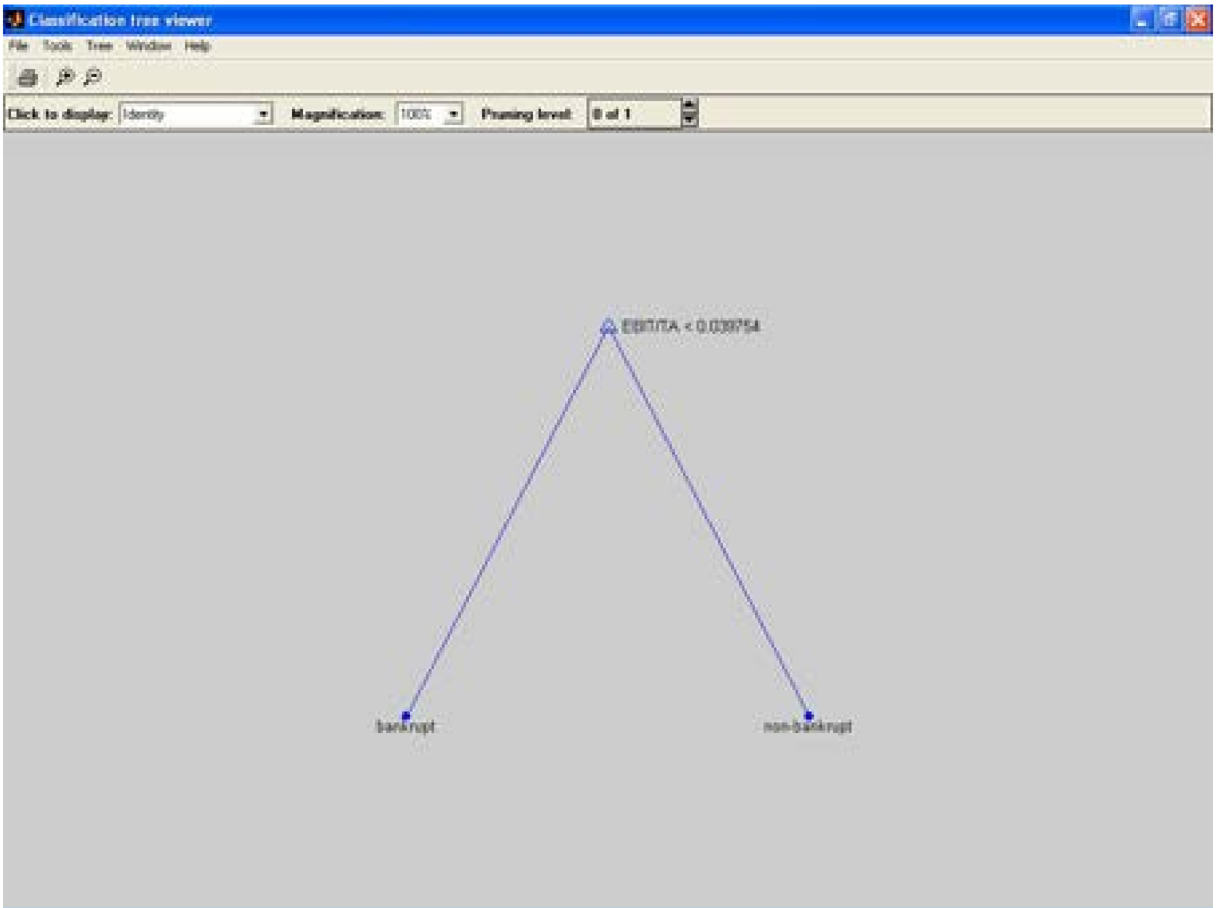

Figure E.12: Pruned tree for the adjusted data set using 11 ratios