Predicting bankruptcy
with discriminant analysis and decision tree
using financial ratios

Lonneke Mous (274799)

Bachelor Thesis Informatics & Economics

Faculty of Economics at Erasmus University Rotterdam

July 2005

Table of contents

1. INTRODUCTION

1.1 Background

When you hand a loan out to someone, one of the most important things you would like to know is whether or not that person can pay you back. For similar reasons banks want to know how credible a business is, before it will loan the business money. In other words: banks need to know what the chances are of the business going bankrupt.

However, banks aren't the only stakeholders of a business. Shareholders have invested their money in the business and they would prefer to get their money back at a certain point in time. When the business goes bankrupt however, the shareholders are often last in line of getting what's left of the business. So they too would want to know if the business will still be around in a couple of years time. But just knowing what you want, does not mean you know how to get it.

1.2 General goal of this thesis

Financial ratios will be used for predicting bankruptcy. These ratios give information about a certain aspect of a business' condition. However it is then still be unclear what the meaning of a ratio is concerning bankruptcy. Because of this you must use a method, which will predict the bankruptcy of a business based on its financial ratios. But there are almost as many different methods available as there are financial ratios. The question is therefore which ratios and which methods to use.

The general goal of this thesis then becomes to compare the performance of two methods on predicting bankruptcy using financial ratios.

1.3 Methodology

Literature is used to find out which methods were used in the past and what their (dis)advantages were. Literature also shows how well each method performed. Based on this information, two methods were chosen: one statistical model and one from the field of Artificial Intelligence. The first was chosen, because it has been used a lot in the past and like that it has become a benchmark for other methods. The second method was chosen mainly out

of personal preferences for its results, which are easy to interpret and which provide information about each financial ratio's specific significance.

Literature will also be used to find out which financial ratios were used by the two methods in the past. These ratios were then used in this thesis as well.

1.4   Structure of this thesis

The following section will handle the literature about predicting bankruptcy. What methods have been used so far and were they successful? Based on that literature a problem definition and a hypothesis will be formed.

Then in section three the theory behind the chosen methods will be examined more closely. The basics of each method will be explained and the two theoretical workings will be compared on flexibility and their assumptions.

In the fourth section the experiments set-up will be discussed. It will show which samples and financial ratios were used and how it will be determined that one method is better than the other.

After that the results of the tests are shown. These results show whether certain assumptions are true, what variables the methods used and how well each method performed. These results are discussed and finally the thesis ends with a conclusion.

2.   *PREDICTING BANKRUPTCY IN THE PAST*

2.1   When is a business in trouble?

Predicting bankruptcy is about determining which business you can trust to meet its obligations and which one you cannot. But when a business cannot meet its obligations, it does not mean it goes bankrupt as well. Failure, insolvency, default and bankruptcy are four different terms and they all mean that a business is in distress (Altman, 1993). Bankruptcy may be the worst case scenario, but default for example poses problems to the business' stakeholders as well. Some studies therefore do not try to predict bankruptcy, but failure instead.

Failure in economic terms means that a business has a rate of return on invested capital that is significantly and continuously lower than that of similar investments. This does not mean that the business will cease to exist nor that it cannot pay its debts on time. It simply means that the business is not making as much money as it should. So the only obligation the business cannot meet is that to its shareholders: it cannot give enough return on its shares. However this is not a hard obligation the business is forced to meet. It is simply a very unfortunate event for the shareholders.

Insolvency can be split into technical insolvency and insolvency in a bankrupt sense. In a way this is simply a difference of short-term and long-term. Technical insolvency means that a business cannot meet its short-term obligations. This can be temporary, for example when a business has invested a lot of money in a project and so it does not have any money to spare at the moment.

Insolvency in a bankrupt sense is worse than technical insolvency, because it means that something is chronically wrong with the business. This happens when the total liabilities are bigger than the total assets.

Default happens when a business cannot pay an obligation to its creditors, for example not paying the periodically interest or paying off the debt at the promised time. This is a serious sign that a business is in trouble, though it may not lead to bankruptcy. A business may negotiate with its creditors to find a way to solve the situation.

Finally bankruptcy occurs simply when a business files for bankruptcy. This is normally done at a public institution in a formal way. After that the business may cease to exist and its assets are sold in order to pay off as much debt as possible. Or the business will try to start over.

Clearly there are different criteria one can use when trying to determine if a business is a distressed one or not. This study focuses on the ones that filed for bankruptcy. The filing often occurs publicly, so it is easy to know when a business goes bankrupt. Also there may be a blurry line between a business suffering from insolvency and a healthy one. When is a business troubled enough? But filing for bankruptcy is a clear-cut distinction: either a business has done it, or it has not. This makes it easier to split the samples in bankrupt and alive businesses.

2.2   History of predicting bankruptcy

So it is desired to predict bankruptcy. But how can you do that?

One way is by using the information stated in a business' annual report. Especially businesses that are listed on stock exchanges have the obligation to give various information about their condition via the reports. With this information you can calculate different financial ratios, which are indicators of a business' condition. Beaver (1966) was the first one to use these financial ratios for predicting bankruptcy. His study however was limited to looking at only one ratio at a time. Altman (1968) changed this by using a multiple discriminant analysis (MDA). This analysis combined the information from several financial ratios in a single prediction. Altman's Z-model was the result of this multiple discriminant analysis and has been popular for a number of decades as it was easy to use and highly accurate.

But there was critique on the MDA-model as well. An important point was that Altman treated businesses from different sectors as the same. Some scientists have argued that different sectors have different kinds of businesses and thus different values for a 'healthy' indication by the financial ratios. It was argued that this difference needed to be corrected. One suggested way of doing so was by only focusing on one sector at a time. An example of this is when Altman himself made a Z-model using only railroad companies (Altman, 1973). A different way of correcting for different sectors has been to divide financial ratios by their sector's average.

Although MDA is the method, which is used the most in this field, it is not the only one. Of the several other statistical methods, logit has been used the most (Aziz, 2004). Logit uses the logarithm of the chance that a business will go bankrupt or not.

But no matter which method you use, it is important to have enough data to be able to predict bankruptcy. Especially information from failed firms may be difficult to obtain. This is one of the main reasons why studies have used several sectors: there were not enough samples from

one sector alone. However more data is available these days. The digital COMPUSTAT file from 1994 for example, contained information from 14.000 firms and it was still growing (McGurr, 1998). Unfortunately this file is not publicly available. Still for those with access to it is now much easier to get the needed samples than a couple decades ago.

But computers did not only have an impact on the availability of data. New methods became available from the area of Artificial Intelligence and its expert systems. Frydman (1985) introduced recursive partitioning or decision tree for processing the financial ratios. This is an inductive learning algorithm: it learns from examples and then tries to learn general rules from them. Humans do the same thing: each time I see a dog, it has a tail. Thus I assume that all dogs have a tail. But unless I have seen every dog in existence, I cannot be certain that the assumption is true. In a way the statistical methods were used for that purpose as well. Though the result from MDA only says something about the samples used, it was then used by the researchers to make general claims and to predict the bankruptcy of non-sample businesses.

Decision tree is not the only AI-method used: neural networks were first used to predict bankruptcy around 1990. Like a tree they take the financial ratios as input and then give a 'bankrupt' or 'alive' classification. But the problem with neural networks is that it is much less clear how the network has reached a classification. The inner workings are difficult to interpret.

Apart from statistical and AI-methods there are also the theoretical methods. Instead of looking at the financial ratios or 'symptoms' of trouble, they try to look at the causes of failure. One of these methods is the "Gambler's Ruin Theory" which was introduced in this field by J. W. Wilcox. However, these methods have not been used a lot and therefore it is difficult to say something about their past performance.

2.3   Problem definition

The purpose of this thesis is to compare two methods based on their performance on predicting if a business will go bankrupt within two years or not using its financial ratios. The two different methods are multiple discriminant analysis and decision tree.

MDA has been chosen, because it is one of the first successful methods for this problem and has been used the most for it. That makes the method a benchmark for other methods. Decision tree has been chosen, because I wanted to use a method from the field of Artificial Intelligence. In that field neural networks and decision tree have been used the most. I decided

not to go with neural networks, because they work like a black box and deliver results that are difficult to interpret. I wanted a method of which I could really understand how it achieved its results. Decision tree is such a method. An added bonus is that not only does decision tree use all the variables as a whole to make predictions, but it also says something about how significant each single variable is.

Past studies about these methods are often a few decades ago. This study however will use recent examples of bankruptcy, to see if the methods are still relevant today.

These methods will (at least try to) predict bankruptcy based on a business' financial ratios. These ratios are indicators of a business' wellbeing. They can be calculated based on information from balance sheets, which many businesses are forced by law to make public. The availability of this information therefore makes it practical to use them.

But before determining which business will go bankrupt, the two methods must first 'learn' the difference between bankrupt and alive businesses. It will do this with a data-driven-approach. The data consists out of samples, where one sample consists of a business with its financial ratios (its attributes) and the knowledge of whether or not it went bankrupt (its classification). The methods then use these samples to determine the difference. This group of samples is called the training set.

2.4 Hypothesis

Previous studies can be examined not only to see which methods were used, but also to predict how well each method will perform. Aziz (2004) has gathered the results from several studies (using different methods) done in this field. On average MDA was 86% accurate. Decision tree (or RPA as it was called in Aziz' study) scored on average an accuracy of 87%. So both methods have performed well and decision tree is slightly better. The closeness of the accuracy-results is not unusual: "Almost all the models are capable of successfully predicting firm's financial health achieving a collective average of more than 85% predictive accuracy rate." (Aziz, 2004).

The goal of this thesis is to determine which method will perform better. The main hypothesis therefore is that the decision tree will perform better than MDA. But because the difference between the two methods in the past studies has only been 1%, this hypothesis is a very weak one.

A stronger, second hypothesis is that MDA and decision tree will both be able to correctly classify whether or not a business will go bankrupt in 80% - 90% of the cases. It is also quite

possible that both methods will be in the lower end of this range. Because when Aziz compared the results of different methods for predicting bankruptcy, he noted that: "It is still not common, to almost half the researchers discussed in present study, to use a holdout sample for validation of their results." (Aziz 2004). Not using a holdout sample for validation causes an upward bias of the performance, as will be discussed in section 4.3. This study however will use a holdout sample and thus the methods may perform worse than they have in the past.

3. DESCRIPTION METHODS USED

3.1 Multiple discriminant analysis

Some studies talk about multivariate discriminant analysis instead of multiple. However, their analysis basically works in the same way.

Multiple discriminant analysis (MDA) tries, as the name implies, to discriminate between different groups. In this case it discriminates between the group of bankrupt businesses and the group of alive businesses.

In order to do this MDA will take into account various samples from both groups. MDA tries to separate these two groups based on the financial ratios of each sample. Each ratio becomes a variable X and gets its own coefficient V. This leads to the following formula, which calculates a sample's Z-score:

$$Z = \sum_{i=1}^{n} V_i X_i$$

Where there are n different independent variables, of which X_i is an example, and V_i is its coefficient.

MDA then works in the following way. Suppose there are G different groups (in this case 2) and N_g is then the number of samples in group g. Then you can determine the average score ($\overline{Z}_g$) of group g in the following way:

$$\overline{Z}_g = \frac{1}{N_g} \sum_{p=1}^{Ng} Z_{pg}$$

Where $p = 1, 2, ..., Ng$ and represents a sample of group g. Z_{pg} is the Z-score of sample p in group g.

In this case there will be two group averages: the average Z-score of the bankrupt businesses and the average Z-score of the alive businesses.

With that the goal of MDA is to maximize the sum of squares among groups divided by the sum of squares within groups by changing V_i.

The sum of squares among groups is determined in the following fashion:

$$SS(Among) = \sum_{g=1}^{G} N_g \left[\overline{Z}_g - \overline{Z} \right]^2$$

SS stands for sum of squares. The sum of squares among groups is therefore measured by using the distance between the different group averages and the total average. MDA tries to maximize this distance.

The sum of squares within groups is determined in the following fashion:

$$SS(Within) = \sum_{g=1}^{G} \sum_{p=1}^{Ng} \left(Z_{pg} - \bar{Z}_g \right)^2$$

The sum of squares within groups is therefore measured using the distance between a sample of a group and its group average. MDA tries to minimize this distance.

Combined this means that MDA maximizes *?*, where it is defined as:

$$\boldsymbol{?} = \frac{SS(Among)}{SS(Within)}$$

As noted before the *Z*-score of a sample depends on X_i, the value of a financial ratio, and its V_i, the coefficient. MDA then searches for the precise value of V_i that maximizes *?*.

The result of the MDA is therefore a formula with the coefficients filled in and a cut-off score for *Z*. To determine whether or not a business will go bankrupt, you should look to whether its *Z*-score lies beneath or above the cut-off score.

3.1.1 Stepwise selection method

Many financial ratios will initially be available for MDA to use, but it isn't necessary to use them all. In fact a model with few variables relative to the sample size yield relatively more accurate results (Huberty, 1994). It is therefore important to reduce the size of the variables used. One way to do that would be to perform MDA on every possible subset of variables and then see which one does best. But a more efficient way is stepwise selection.

Stepwise selection is an iterative process. The set of variables used starts out empty and then step-by-step it is filled with other variables. Of course the 'best' variable is added every time. But how can you determine which one is the best? There are different criteria for this. The one that is used the most is Wilks' lambda. First it is computed how much each variable lowers Wilks' lambda. The one which lowers it the most is chosen. This only occurs however if that variable also makes a significant difference. This is determined via its F statistic: if the probability of F is lower than a (usually 5%), then the variable is considered to be significant.

In the next iteration, stepwise selection chooses the next variable to be added. Again this is the one with the lowest value for Wilks' lambda on the condition that it is significant. However, due to correlations it is possible that the first variable no longer matters. So after

adding the second variable, stepwise selection checks if all variables are still significant. When they are not, they are removed.

The process continues adding another variable and checking for removal until there are no longer significant variables that can be added.

3.2 Decision tree

Decision tree is a well-known and widely used method for solving classification problems. A tree consists out of several nodes. The first node is the tree's root. Apart from that it also has internal nodes and leafs. An example of this can be found in figure 1.
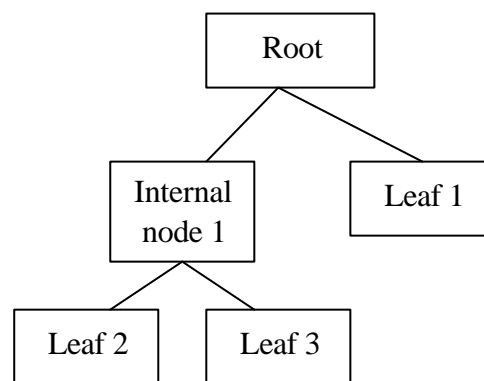

Figure 1. A basic binary tree

3.2.1 Growing the tree

A tree starts with the same data as multiple discriminant analysis, namely a group of samples with various attributes (the financial ratios) and a classification of whether or not it went bankrupt. At each internal node, the tree splits the group of samples based on an attribute value.

Entropy is almost always used to determine on which attribute the tree should split. The concept of entropy originates from physics and it "characterizes the (im)purity of an arbitrary collection of examples" (Mitchell, 1997). The decision tree has to classify whether or not a business will go bankrupt. So if you have a group of samples, who are all classified as bankrupt, you have a very 'pure' group. And if the group is fifty-fifty when it comes to bankruptcy, you have a very impure group. The higher the entropy value of a group, the more impure it is. You can determine the entropy of a group of samples *S* via the following formula:

$$Entropy(S) = -p_1 \log_2 p_1 - p_2 \log_2 p_2$$

Here p_1 is the proportion of the group that went bankrupt and p_2 is the proportion that did not. If p_1 and p_2 are 0.5, then the entropy is at its maximum of 1. If either p_1 or p_2 is 1, then the group is completely pure and the entropy is at its minimum of 0.

Ideally you want leaf nodes in which there are only samples of businesses that went bankrupt or of those that stayed alive. In other words you want completely pure groups. So when you split a group of samples, you want the remaining groups to be 'purer' and so the entropy must be reduced. To determine on which attribute you will split the samples, you therefore use the attribute which reduces the entropy the most. This reduction of entropy is called the information gain. If you split group S based on attribute A, then the information gain is:

$$InformationGain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

The formula aggregates over the different values attribute A can have. As you can see the information gain depends not only on the entropy of a new node, but also on how many samples there are in that new node.

After you have split the group, you split the new remaining groups and so on until you cannot reduce the entropy any further.

3.2.2   Splitting criteria on a numerical attribute

Because financial ratios are used as input for the tree, the tree must split based on numerical attributes. This generally happens via a binary split (meaning two nodes will be the result of the split), while using a threshold value (Berzal, 2004). Splitting the tree into more than two nodes is much more complex and therefore it is not done. But it is not a problem if the tree should be divided into three groups based on a numerical attribute. That is still possible with binary splits, though in a more complicated way. Say there is an attribute A and the samples should be divided into three groups: those whose value of A is lower than x, those with a value between x and y and those with a value higher than y. Then the tree will simply first split the group on whether its value of A is higher or lower than y (or whether its value is higher or lower than x, that depends on the information gain of the two splits). After that the samples with a value for A lower than y will be split based on whether their value is lower than x as well. So instead of splitting the tree once on the attribute in three ways, the tree now splits the samples twice based on the attribute and both times in two ways.

The threshold value used in the splitting (the x and y in the sample above) is a continuous number like the ratios, which implies an infinite number of possible thresholds. Fortunately

the group of possible thresholds can be drastically reduced. Say you wish to split on attribute V and there are n samples being used. Then there are n different values for V, namely $\{v_1, v_2, ... v_{n-1}, v_n\}$. If the tree will split on a value between v_1 and v_2, the result will be the same as that of a split on a different value between v_1 and v_2. In other words only $n-1$ splits have actual different results, so you should only examine $n-1$ different threshold values.

Once it is determined that there should be a split between v_i and v_{i+1}, then the precise value of the threshold should be determined. A simple way of doing this is by taking the average of v_i and v_{i+1}.

3.2.3   Pruning the tree

When you have grown a complete tree and every leaf node contains either bankrupt or alive samples, you do not necessarily have a good tree. There is the problem of overfitting. Because the tree is too much tailored to the specific training set, it can no longer make general statements about other businesses. Because of this there is often a pruning step after having grown the tree. There are various ways to prune a tree. You can simply start cutting off leaf nodes or use a pruning rule. A pruning rule can be very complex and containing a formula, which rewards accuracy, but penalizes complexity.

In the end it does not really matter how you prune the tree. What matters is determining whether the pruned tree is better or not.

An obvious way of doing this is using a separate test set or holdout set. The samples in the test set were not used for training. So if a tree does well on the test set, it does well in making general statements and that is what we want. Pruning in this way is called reduced-error pruning.

However it is difficult to get a lot of bankrupt samples for this study. So I want to use all the samples I can get for training the tree. But then there will be no samples left for a test set.

The algorithm used for making the decision tree in this study is C4.5, a popular and non-commercial algorithm. It handles this problem by estimating the error rate made in general based on the training data alone. Of course there are theoretical objections to this heuristic, especially because the statistical underpinning is rather weak, but it seems to work well in practice (Witten, 2000). Say there are N samples being used for training. Of those N samples the tree classifies E samples incorrect; therefore E is the number of errors. Then assume that the N samples were generated by a Bernoulli process of which E turn out to be errors. The process uses q as a parameter for this. Basically the Bernoulli process would randomly pick out some of the N samples and tells them they are errors. The parameter q it uses represents

the chance that the process will call a sample an error. This means that q represents the true error rate. It is therefore q that we need to know in order to estimate how accurate a (pruned) tree is.

E/N is the error rate on the training test, so it gives an estimation of q, but how confident can one be of this estimate? That is where confidence intervals are used. Given a desired confidence (C4.5 uses 25% by default) we can find the confidence limits z of that interval. But because *E* and *N* are not from an independent test set, *E/N* may give an optimistic picture of the true error rate. To compensate this C4.5 looks only at the upper confidence limit instead of the entire interval, which is a pessimistic thing to do (Witten, 2000).

$$\Pr\left[\frac{f - q}{\sqrt{q(1-q)/N}} > z \right] = c$$

Here c is the desired confidence (25% by default) and f is *E/N* or the observed error rate.

With this formula we can get the upper confidence limit z. This is then used to estimate the true error rate e in the following way:

$$e = \frac{f + \dfrac{z^2}{2N} + z\sqrt{\dfrac{f}{N} - \dfrac{f^2}{N} + \dfrac{z^2}{4N^2}}}{1 + \dfrac{z^2}{N}}$$

In order to decide which (pruned) tree is the best, the one with the lowest value for e is chosen.

3.3 Flexibility comparison

Even before actually testing how well each method performs, you can already say something about how useful each method is.

MDA is a parametric method. This means that MDA makes an assumption about the distribution of the variables. In this case it assumes that the financial ratios have a normal distribution. Research has shown that this is probably not the case. MDA also assumes that the covariance matrices across the different groups are equal. Again it is questionable whether this is true. Both assumptions will be tested in this study. The normal distribution assumption via a chi square goodness-of-fit test. Using the data, the mean and standard deviation of each financial ratio can be calculated. If a ratio would follow the normal distribution, 68% of all the observations must lie within one standard deviation away from the mean. Also 95% of the observation would have to lie within two standard deviations away from the mean. A

goodness-of-fit test therefore constructs different intervals and knows how many observations should lie in the interval. It then compares that number with the actual number of observations in the interval. If O_i is the number of actual observations in interval i (and there is a total of k intervals) and assuming E_i is the number that should lie in the interval if a normal distribution is followed, then the chi-square is calculated like so (Aczel, 2002):

$$c^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

It is then calculated how likely it is that chi-square has this value. That probability becomes the p-value with which the null hypothesis of a normal distribution may be rejected.

Whether the covariance matrices are equal or not will be tested using the well-known and often used Box's M test. This test is more complicated. Explaining it would therefore take a lot of time, even though the test is not the main scope of this thesis. That is why I simply assume that the Box's M test is able to do the required job.

Decision tree is a non-parametric method. So it does not assume anything about the distribution nor about the covariance matrices. Because these assumptions may be violated, this is an advantage of the decision trees over MDA.

In fact a decision tree only assumes that the different groups are discrete, non-overlapping and identifiable. MDA shares this assumption and in this case the assumption is true. There are only two different groups and a business either goes bankrupt or not; it cannot do both.

A different kind of flexibility is in the variables that are being used. Decision trees can handle non-numerical attributes and MDA cannot. Because of this, decision trees can use more indicators for determining if a business goes bankrupt. However, this thesis focuses on financial ratios, which are all numerical attributes. Still, the extra flexibility of the decision tree can be an advantage.

4.   EXPERIMENT SET-UP

In this experiment the program SPSS was used to conduct the MDA. The program Weka was used to perform the decision tree analysis, which uses the C4.5 Revision 8 algorithm for this. C4.5 is a popular non-commercial algorithm.

But first this section looks at exactly which samples were used for this experiment. Then it is examined which financial ratios from these examples were used. These samples and their ratios will be used by MDA and the decision tree for training.

Finally the matter of the performance measure is addressed: how can it be determined which method is better?

4.1   Samples used

The samples came from Thomson One Banker Analysis (TOBA), which received the balance sheet information from ThomsonFinancial. TOBA was used simply because it was available to me via the Erasmus University Rotterdam.

The samples used are businesses in the non-cyclical consumer market. Other businesses like financial ones were not included, because their financial ratios for a healthy business may be different. No further focus on a specific sector was taken, because that would reduce the number of samples even more and there should still be enough samples left.

The resulting samples containing missing values for one (or more) of the needed financial ratios were removed. Note that this is necessary for MDA, because it cannot handle missing values. A decision tree, however, can handle them and thus would have been able to use a bigger training set, even though that was not done here.

TOBA makes a distinction between inactive and active businesses and has information of about 320 inactive businesses in this specific market (including ones with missing values). Inactive here means that TOBA no longer gets new data from the business. Not all these 320 businesses were used, because not every inactive business really did go bankrupt. Some businesses simply do not exist anymore because they merged with another business. Others were no longer listed at the stock exchange and became private businesses. To gather information about which business really did go bankrupt, the websites www.bankrupt.com (its

news archive of the 'troubled company reporter') and www.bankruptcydata.com (its alphabetical list of bankruptcies) were used.

Eventually data of 43 bankrupt businesses remained. Originally the goal was to predict bankruptcy a year before it occurred. This required reports concerning a business' wellbeing from no more than a year before the business filed for bankruptcy. But a lot of the most recent reports were from over a year before bankruptcy. Therefore the goal was changed into predicting bankruptcy two years before it occurred. Because there are more bad indicators a year before bankruptcy than two years before it, it is likely that this will have a negative effect on the performance of the methods.

Then from TOBA 54 non-bankrupt businesses from the same sector were randomly drawn. So in the end 97 samples were used. These samples contain mostly American businesses, but there are businesses from other Western countries as well. Naturally the data must have the same dimension in order to compare the samples, so all the data has been transformed into US dollars. However the exchange rate may blur the information: a hard or soft dollar may make the figures bigger or smaller. Luckily all but one of the financial ratios are (as the name says) a ratio. So a number of US dollars is divided by another number of US dollars. Due to the division, the result is dimensionless and the exchange rate loses its effect.

All the samples are from around the year 2000, so somewhat recent samples were used in this study.

4.2   Financial ratios used

There truly are as many different financial ratios as you can imagine. Simply divide one financial number by another and voila: you have a financial ratio. However, not all the resulting ratios will be relevant or meaningful. So to determine which financial ratios to use, I observed previous studies to see which financial ratios they used. I gathered those and shorted it to 28 ratios, which were used for this thesis. When shorting it to 28, I looked at whether two financial ratios looked a lot like each other and whether I had enough data for the ratio. I did not invent extra financial ratios on my own. The exact set can be found in Appendix A. Care was taken to make sure that different kinds of ratios were used. "Different kinds" here means representing different aspects of the business.

One important aspect is profitability. When a business is not profitable, there obviously is a chance of it going bankrupt. So some financial ratios will say something about its profitability like "earned income before interest and taxes divided by total assets". Not only does that say

something about how much a business has earned, but it also indicates how efficient the business' assets have been put to use in earning that amount of money.

However, profitability is not the only aspect worth looking at. The liquidity or solvability of a business is also important. Liquidity looks at how easily a business' assets can be exchanged for money. If a business is in trouble and it needs money to pay off debts, then it is easier to do so when its assets are only cash than when its assets are big specialized machines. A financial ratio that says something about liquidity is for example the current ratio: the current assets of a business divided by its current liabilities. It shows how easily short-term debt can be paid off with the current assets (which are very liquid assets).

Associated with the previous concept is leverage. A business needs funds to invest. Part of those funds comes from its owners or shareholders. Another part comes from debt, which is often provided by banks. A company is said to have high leverage, when there is a lot of debt and little money from the shareholders. Therefore a company with high leverage isn't very solvent, because it has a lot of debts it needs to pay. An indication of leverage is "total common equity divided by total liabilities". So you have the shareholders' part divided by the debt part.

Another interesting aspect is a business' turnover. If there is a high turnover, then the products made by the business do not stay in the inventory long, but are sold quickly. In a way this too says something about liquidity, namely how fast the products can be exchanged for money. It may also indicate how well the products are doing. Inventory divided by sales is a financial ratio concerning turnover.

Finally the size of a business can be important. New, small businesses go bankrupt more easily than big, mature businesses. Big businesses often have better management and if it is in trouble, there are more stakeholders willing to help it. Governments for example have granted loans to big, troubled companies, because they do not want many people to become unemployed. A financial ratio used for size is log(total assets).

4.3 Performance measure

In order to determine which method is better, which is the goal of this study, a way of measuring the performance of each method must exist.

Eventually MDA and the decision tree both classify a business as bankrupt or alive. MDA does this based on the Z-score and the tree based on which leaf the business ends up in. An

obvious way would be to determine the percentage of the accurately classified samples. In other words: which percentage of the training set is classified accurately.

However, when you use this measure, there is the problem of overfitting. You don't want a tree or a formula that fits the training set perfectly. Instead you want to predict bankruptcy for businesses, that were not in the training set and of whom you do not yet know whether they will go bankrupt. So it is important to look at how well the tree or formula generalizes, or how well it does on non-training set samples.

The easiest way to do this is to split the data into two distinct, non-overlapping groups: a training set and a test set (or holdout samples). The samples in the training set will be used to determine the coefficients in the MDA-formula and will be used to build the tree. Then you can see how accurately they both classify the samples from the test set. This would indicate how well they can make general statements.

There is however one problem with this: when data is limited as is the case in this study, you will want to use all the data you've got for training. Bootstrap is a method which can solve this problem. This thesis however uses k-fold cross validation instead, because this method was much easier to use in both programs and I have experience with this method. Cross validation is also recommended for a decision tree, as Frydman (1985) says that "for classification trees the V-fold cross validation procedure is preferred to the bootstrap procedure, especially when, as is the case in our study, the effective size of the sample is relatively large". This is because bootstrap causes a downward bias for decision tree.

K-fold cross validation is, like bootstrap, a method of generating a test set from the training set. With k-fold cross validation the sample-group is split into k groups. In each iteration one of the k groups will serve as a test set and the remaining samples as a training set. So if you perform a 10-fold cross validation, you will have 10 combinations of a training set and a test set. The MDA and the tree use every combination, which results in 10 different trees and MDA's based on the 10 different training sets. But it also results into 10 different error rates based on an independent test set. The average of those 10 error rates is then taken to estimate the true error rate of MDA and decision trees. In the statistic world it is common to use the "Leave-One-Out hit rate" (L-O-O). Here hit rate means the percentage of samples that is correctly classified. L-O-O is basically a k-fold cross validation, where k is the total number of samples. In other words: every time the entire group of samples is used for training except for one sample, which will be the test set. There are 97 samples used in this study, so for this study L-O-O is equal to a 97-fold cross validation. It may not always be wise to use L-O-O for judging an Artificial Intelligence method. Neural networks for example take a long while

to train. And if the data set contains a lot of samples (meaning that a lot of networks need to be trained), then using L-O-O will take quite a while. However, the number of samples in this study is limited and decision trees are fast to make (it took the program a mere 0.05 seconds to make a tree). So using L-O-O in this study is acceptable and it will be the prime measure for determining which method is better.

A final note on different kinds of misclassification. In essence there are two types of errors: the tree will classify a bankrupt business as alive (Type I error) or it will classify an alive business as bankrupt (Type II error). Up till now the two errors were considered to be equally important. But it is "estimated that, for commercial bank lending officers, classification of a bankrupt firm in the non-bankrupt group is 32 to 62 times more costly than the reverse misclassification." (Frydman, 1985). Because of this difference less Type I errors are preferred over less Type II errors.

The C4.5 algorithm can keep these different costs in mind by simply adding a cost matrix. And in a more complicated way MDA can cope with the different costs as well. MDA can assume that the prior probabilities of belonging to a group are equal (so without any other knowledge there is a 50% chance the business will go bankrupt). Of course this is not true in real life: there are many more businesses that stay alive than those that do not. Therefore MDA will be more inclined to classify a sample as 'alive'. However, a Type I error costs more than a Type II, so in that respect MDA should be more inclined to classify a sample as 'bankrupt'. And this is done by increasing the prior probability of being a bankrupt business. After determining the prior probabilities, the cut-off score for Z will lie further away from the bankrupt group average. That makes it more likely that a business will fall in the bankrupt-side of the Z score.

But because it is unknown how much the prior probabilities really are and because it is unknown how much more a Type I error costs relative to a Type II error, no difference has been made between the two. The C4.5 algorithm did not receive different costs and the prior probabilities for MDA were equal.

5. *RESULTS AND DISCUSSION*

First the results from testing the assumptions of MDA will be shown. Then the resulting formula and tree from both methods are shown and with it a discussion about the specific chosen variables. Finally the actual tests of how well each method performed are discussed.

5.1 Testing MDA assumptions

Two main assumptions of MDA are tested here. The first is that the financial ratios follow a normal distribution. The second is that the covariance matrices are equal.

Instead of checking to see if the assumption holds for each financial ratio, the test was only performed on the three financial ratios that MDA actually used to predict bankruptcy. The test is a goodness-of-fit test calculated with a chi square. The null hypothesis is that the ratios follow a normal distribution; the alternative hypothesis is that they are not. The *p*-values as a result of the tests are:

Financial ratio	*P*-value
V4	0.000
V22	0.000
V26	0.0272

Table 1: Goodness-of-fit

Clearly with an a of 5% (which is a common choice of a), the chances are all below *p*-value. This means that there is only an incredibly small chance that the null hypothesis is true and there is significant enough evidence to reject the null hypothesis. So the null hypothesis has to be rejected for all the ratios and the alternative hypothesis is accepted. None of the ratios follow a normal distribution and the assumption is rejected. If they had followed the normal distribution, then the resulting *Z*-score would have followed a normal distribution as well. In this case the *Z*-score did not follow a normal distribution either: it scored a *p*-value of 0.000 on the same test as the financial ratios.

Box's M test is used to see if another assumption of MDA holds: are the covariance matrices equal or not? The null hypothesis is that they are, the alternative hypothesis is that they are not.

Box's M	208.133
(F) Approx	33.477
df1	6
df2	57010.589
Sig.	0.000

Table 2: Box's M test

The *p*-value is 0.000. So no matter which a one may use, the chance is below it. Therefore there is significant enough evidence to reject the null hypothesis and to accept the alternative hypothesis. This means that the covariance matrices are not equal. So another important assumption of MDA does not hold. When the covariance matrices are not equal, a quadratic discriminant analysis becomes possible. This analysis does not force a linear formula and does not assume equal covariance matrices. However it did not perform better than the linear discriminant analysis and it is more complicated. So this study continued working with a linear analysis.

5.2   Resulting formula and tree

In this section we look at the readable results from MDA and decision tree: its formula and its tree. Special attention is made to exactly which variables are used by the two methods and whether the use of them makes sense.

5.2.1 Resulting formula

First a look at the resulting formula from the MDA:

	Function 1
V26	.740
V4	.522
V22	-.569

Table 3: Discriminant Function Coefficients

The table shows which variables are used and what their coefficients are. This means that the formula is:

$$Z = 0.740 * V26 + 0.522 * V4 - 0.569 * V22$$

So of the 28 different ratios the MDA only used 3 in the end. When using equal prior probabilities (as was done in this experiment) the cut-off score for Z is calculated by taking the average of the group averages Z-score. Here that is $(0.776 + -0.618)/2 = 0.079$. The positive number is the Z-score of the bankrupt businesses average; the negative one of the surviving businesses. In order to qualify a sample, its Z-score is calculated using the aforementioned formula. If the score is below 0.079, it is classified as alive. If the score is above 0.079, the sample is classified as bankrupt.

MDA used three variables: V26 (total liabilities / total assets), V4 (Cash / Sales) and V22 (Retained earnings / total assets). Interestingly enough when Frydman first used decision trees for solving this problem, the tree used five variables including V26 and V4 (Frydman 1985). The third variable V22 was used by Altman's MDA when he first used MDA to predict bankruptcy (Altman, 1993). He used it for his study in 1968, but apparently it still is an important factor with businesses today.

Now a look at what each variable represents. V26 is about leverage: how big is the business' debt and is it possible to pay it off with its assets? Higher leverage means more debt and so it means that there is a bigger chance that the business cannot pay its obligations and therefore goes bankrupt. This can be seen in the formula: V26 has a positive coefficient, so a higher value of V26 (and therefore a higher leverage) means a bigger chance of being a bankrupt business.

V22 is a clear variable as well. Retained earnings are basically a cumulation of past profit minus the dividends that were paid. So V22 says something about the past, continuous profitability from a business. A high V22 means high past profitability and is therefore a good thing. This shows the formula as well: V22 has a negative coefficient. So a higher value of V22 causes a lower Z-score and therefore a bigger chance that it will be an alive business.

V4 however isn't so obvious. When Frydman used it in one of its lower layers, a high V4 indicated a healthy business and a lower value indicated a bankrupt business. So the higher V4 is, the better. Here however V4 has a positive coefficient and the higher the Z-score, the higher the chance of bankruptcy. So here a high V4 is a negative thing. This is clearly a strange situation. A better look at V4 may explain this difference: a high V4 means the business has low sales (which is a bad thing due to profitability) or it has a lot of cash at hand (which is a good thing, because cash is very liquid, giving the business a good solvency position) or naturally a combination of both. Perhaps this dual nature of V4 causes a high value to be a good thing at one time and a bad thing at another time. However that has still not solved the mystery, because there are other financial ratios that say something about

profitability or liquidity. If V4 is so dual, then why isn't a clearer ratio preferred by MDA? Shouldn't a clearer ratio provide a better distinction between bankrupt and alive businesses?

5.2.2   Resulting tree

The resulting tree of the decision tree analysis is shown below. Each node shows the numbers $x;y$. Here the x is the number of bankrupt businesses in that node and y is the number of businesses that stayed alive.
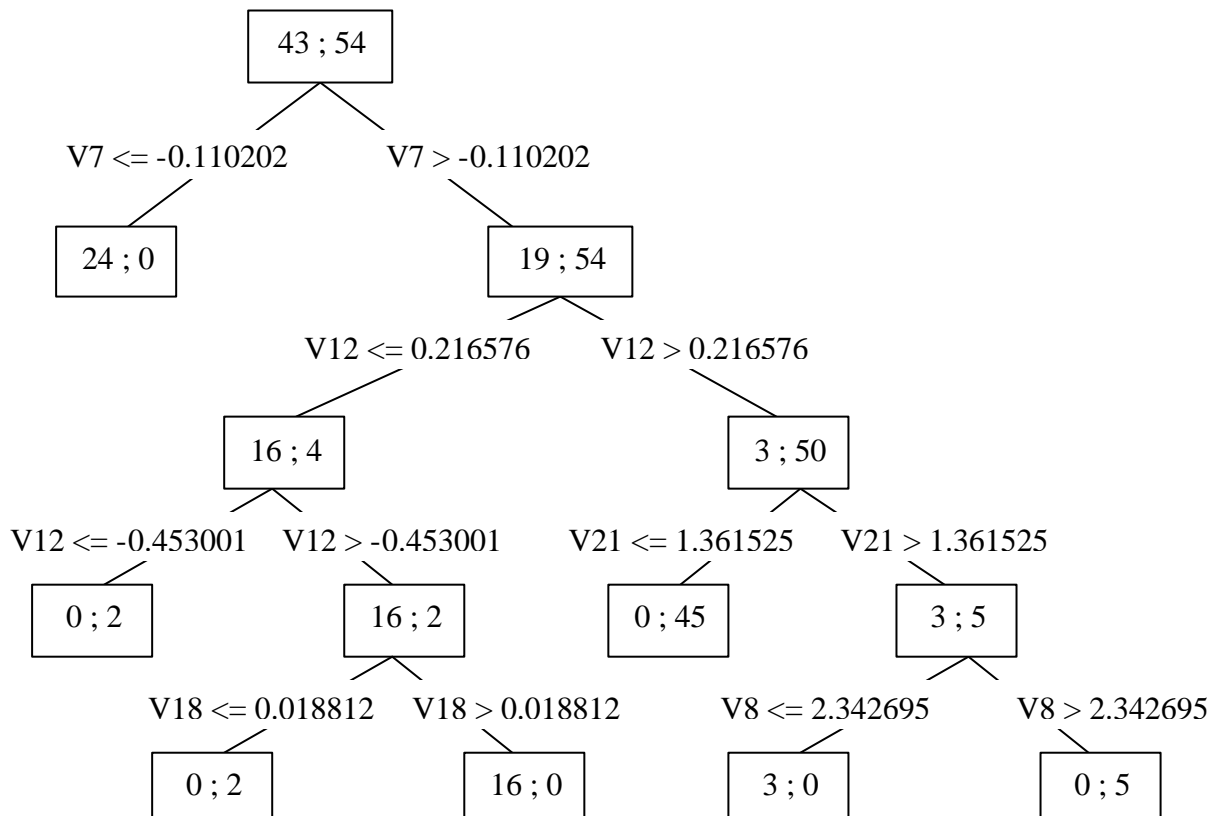

Figure 2: resulting tree

The tree used 5 financial ratios of the 28; two more than MDA. What is more striking however, is that MDA and decision tree do not have a single variable in common. Apparently some variables have more impact with one method than with the other. It is therefore wise to have several financial ratios to choose from: what works for one method may not work for the other. However it was shown that while MDA does not use the same variables as this tree, it did use a variable which was used by Frydmans tree. And this tree uses a variable, which was also used in Altman's discriminant analyses. So while it may seem at first that different methods require different ratios, that thought is now discarded.

Another thing that jumps into mind when looking at the tree is possible overfitting. This can be seen in the leaf nodes. Instead of being satisfied with 2 misclassifications, the tree

introduces the variable V18 to correctly classify them after all. And it also introduces V8 in the end just to separate 8 samples (or 8.24% of the total). Introducing variables in order to correctly classify only a small part of the samples indicates overfitting.

The decision tree used five variables: V7 (Earned income before interest and taxes / total assets), V12 (Total common equity / total liabilities), V21 (Rate of annual growth in net sales), V8 (Earned income before interest and taxes / (interest expenses on debt – interest income) and V18 (Inventory / sales).

But of all these V7 and V12 are clearly the most important. V7 alone for example results in a leaf node with bankrupt businesses only, which has the size of about a quarter of the total samples. V7 was also used in Altman's original MDA of this problem (Altman, 1993) and represents profitability, but also efficiency. What matters is how much you earn and how much you had to spend (the total assets) to earn that amount. V12 is more about leverage: how much debt does a business have. If a business has a lot of debt, then there is a bigger chance of not being able to make its obligations to pay all the debt and thus causing bankruptcy.

So far the variables in the tree make sense. But then things get a bit odd. As said before the tree may suffer from overfitting. V12 for example is used a second time. Apparently high leverage (and therefore a low value of V12) does indicate bankruptcy, but now a really high leverage (a negative value of V12 even) indicates that the business will stay alive. This clearly does not make a lot of sense from an economic point of view. And not being able to logically explain the splitting may again indicate overfitting: in general a negative value of V12 means bankruptcy, but now it means staying alive because of two samples in the data.

V21 partly suffers from the same problem: in general a good growth in sales means that the business is growing in a positive way and will probably stay alive. However here a growth lower than 1.3 means that the business will stay alive and if it is above it, then all of the sudden there is a chance the business will go bankrupt. The tree therefore indicates that businesses can grow too fast. There may be some truth in that: when a business grows fast, it needs to expand fast as well. That means that extra money needs to be borrowed to make those investments. However the samples in this part of the tree have already proven that they are not high leveraged businesses, because they have passed V12. So it is still doubtful whether the split on V21 does not go against common sense.

V8 was also used in Frydman's original tree (Frydman, 1985). V8 handles interest coverage: is the business able to pay its mandatory interest expenses (minus its interest income) from its own income? The values are high for the healthy businesses, because then the income is much

bigger than the interest expenses and thus it is easy for the business to meet its obligations. So the businesses with a higher value stay alive and those with a lower value have the insolvency risk and may go bankrupt. Therefore this split is logical as well.

Finally V18 and fortunately this one makes sense as well: a high value of V18 means that products stay in the inventory a long time and it takes a while to sell them. It therefore means a low turnover, which is a bad thing. And here the businesses with a high value of V18 go bankrupt and those with a lower value stay alive. So V18 makes sense as well.

In general the decision tree does look at the various different aspects of a business: its profitability and efficiency (V7), its leverage (V12), its ability to pay its interest expenses (V8), its growth (V21) and its turnover (V18). On one hand I consider this diversity to be a good thing. On the other hand all the variables may have caused overfitting, which is a bad thing.

5.3 Performance results

Now a look at the actual performance

			Predic. bankrupt	Predic. alive	Total
Original	Count	Bankrupt	28	15	43
		Alive	2	52	54
	%	Bankrupt	65.1	34.9	100
		Alive	3.7	96.3	100
Cross validated	Count	Bankrupt	28	15	43
		Alive	3	51	54
	%	Bankrupt	65.1	34.9	100
		Alive	5.6	94.4	100

Table 4: Classification results MDA

			Predic. Bankrupt	Predic. alive	Total
Original	Count	Bankrupt	43	*0*	43
		Alive	*0*	54	54
	%	Bankrupt	100	*0*	100
		Alive	*0*	100	100
Cross validated	Count	Bankrupt	34	*9*	43
		Alive	*8*	46	54
	%	Bankrupt	79.1	*20.9*	100
		Alive	*14.8*	85.2	100

Table 5: Classification results Decision Tree

The italic numbers represent misclassifications. The upper part of the table ("original") shows the performance of the methods on the entire training set. Here all the samples were used for the training set and no test set was used. "Count" shows the precise number of samples with that classification. For example with MDA with the "original" test, 28 of the 43 bankrupt samples were classified as such and the other 15 were classified as alive. These same results are shown below "count", only now in percentages.

The lower part of the table ("cross-validated") shows the performance of the methods with the L-O-O criteria. The split in "count" and "%" is the same as in the upper part. So with the cross validation, 14.8% alive businesses were wrongly classified as bankrupt by the decision tree and 85.2% were rightly classified.

To calculate the overall performance you add the number of correctly classified bankrupt businesses plus the number of correctly classified alive businesses. You then divide this by the total number of samples, which is 97, and you have the overall performance. MDA with cross validation for example classified 28 bankrupt businesses and 51 alive business correctly. The overall performance is then (28+51)/97 = 81.4%

Now that it is known how to read the tables, you can look at the actual performance of the two methods. Both methods have clearly performed well. When all the samples were used for training (the "original" section of the tables), MDA classified 82.5% of the samples correctly and the decision tree even 100%! The second hypothesis was that the methods will correctly classify about 80%-90% of the samples. It may at first seem that the decision tree performed better than expected. However, it was decided to use cross validation to measure performance,

because it is a more reliable estimate. With the Leave-One-Out method or (as it is in this case) 97-fold cross validation MDA classified 81.4% correctly and decision tree 82.47%. So both are in the range of 80%-90% and the second hypothesis is confirmed. Still they are in the lower end of the range. This may be caused by the data: an extra couple of hard-to-classify samples already lowers the percentage a lot due to a small set of samples.

Both with and without the cross-validation did the decision tree perform better than MDA, so the main hypothesis is confirmed. However the difference between the two is small as expected.

What is very noticeable though is that with cross validation MDA only got 1.1% worse; it misclassified one extra sample. This is a very small difference and the percentage is still high, so it seems that MDA is robust and generalizes very well.

The performance of the decision tree unfortunately makes a bigger drop. This again may indicate that the tree overfits the data too much. And because the shape of the tree and the variables it used also indicated overfitting, it is worth to see if the tree can become more robust against overfitting.

5.3.1 More robust decision tree

As seen in section 3.2.3 there is a certain amount of confidence c that has to be chosen when it comes to pruning. A lower c punishes bigger trees and as such overfitting. However lowering the c did not have any effect.

A second method of making the tree more robust is stating a minimum. This minimum tells how many samples there should at least be in each node. By default the minimum is two. Raising the minimum to three for example would make the split on variable V18 impossible, since it leads to a leaf node with only two (alive) businesses. Thus the tree can no longer introduce new variables to correctly classify only a small group of samples. While increasing the minimum, the performance on the cross validation kept improving. The best performance comes with a minimum of six (and higher). The resulting tree is then cut back to only two splits:
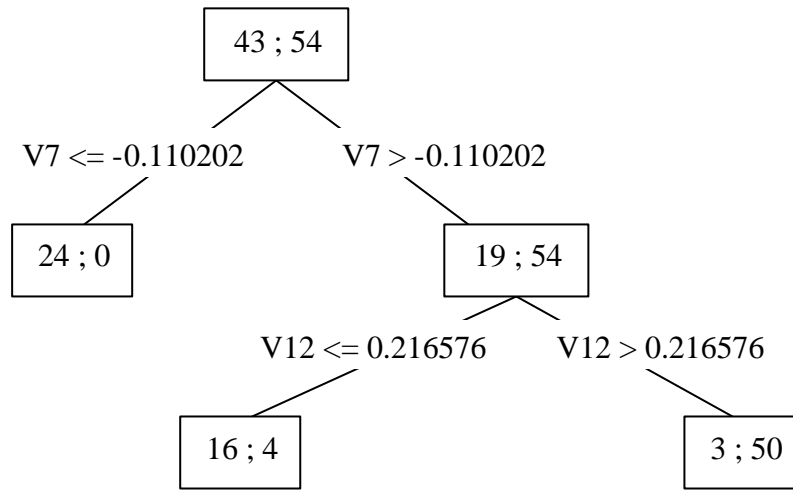
```
                    ┌─────────┐
                    │ 43 ; 54 │
                    └─────────┘
          V7 <= -0.110202    V7 > -0.110202
        ┌─────────┐              ┌─────────┐
        │ 24 ; 0  │              │ 19 ; 54 │
        └─────────┘              └─────────┘
                      V12 <= 0.216576    V12 > 0.216576
                    ┌─────────┐          ┌─────────┐
                    │ 16 ; 4  │          │ 3 ; 50  │
                    └─────────┘          └─────────┘
```

Figure 3: more robust tree

The result is simply the first two splits of the original tree. Its performance table is:

			Predic. bankrupt	Predic. alive	Total
Original	Count	Bankrupt	40	*3*	43
		Alive	*4*	50	54
	%	Bankrupt	93.02	*6.98*	100
		Alive	*7.4*	92.6	100
Cross validated	Count	Bankrupt	38	*5*	43
		Alive	*6*	48	54
	%	Bankrupt	88.4	*11.6*	100
		Alive	*11.1*	88.8	100

Table 6: Classification results Robust Decision Tree

So the tree's performance on the entire training set is 92.8% and with the cross validation it is 88.7%. The tree itself is smaller and it makes a smaller drop in performance. Both indicate that the new tree is more robust and suffers less from overfitting. And although the performance in case of the "original" has decreased compared to the old tree, its performance on the cross validation has increased and that is what matters.

Both the hypothesis still hold: the performance of the tree is in the range of 80%-90% with its 88.7% and it still performed better than MDA.

6.   CONCLUSION

Multiple discriminant analysis and decision tree are both able to predict the bankruptcy of a business within two years pretty well with MDA correctly classifying 81% and the more robust tree classifying 89% correctly. Decision tree thus performed better than MDA, though that difference is a bit small with 7.3%. Still this difference is bigger than what Aziz (2004) found, because here MDA performed worse than in the past studies Aziz examined and the decision tree performed better.

Still both methods are able to predict bankruptcy well with only a small difference, so when choosing between the two methods other factors are worth looking at.

The decision of which method to use, may then become a personal taste. How well do you know each method and how easy can you use one with the available programs?

However apart from the quantitative measure of performance and personal taste, you can also look at the general qualities of each method. And then the decision tree comes out the winner yet again. A decision tree is able to use samples with missing values. For MDA the entire information of such a sample would be lost.

Decision tree can handle qualitative variables as well. If you wish to predict bankruptcy using other indicators than figures like financial ratios (for example your personal opinion of the business' management), you can include this variable in the decision tree, but not in the MDA. Decision tree relies on less strict assumptions than MDA. This again makes the trees more flexible. And in this study it was shown that the assumptions of MDA were violated. Though MDA still performed well, it is in general not a good thing when the underlying assumptions do not hold.

Both methods have an easily readable result. Everybody can fill in a basic formula. And everybody can follow the simple rules of a tree. So no method has a clear advantage there.

A downside of the decision tree is that in this study it suffered more from overfitting than MDA. There were several indications for this overfitting (like using extra variables for only a small amount of the sample), but it became very clear with the difference between the hit rate on the complete training set and the hit rate with cross validation. MDA only performed slightly worse with the cross validation, but the performance of the decision tree made a big drop. This overfitting was solved by making the tree more robust and thus improving its performance. But this means that the danger of overfitting does exist with a decision tree and you should be aware of it. This means that a decision tree may be unreliable and you have to

take the extra effort and time to protect the tree against overfitting. The problem of overfitting is therefore solvable, but it is still a problem.

All in all the decision tree not only performed better, but proved to be a more flexible method with more possibilities.

When it comes to the kind of variables used, both methods used different variables. But they both used (past) profitability and a ratio for leverage or liquidity. So it seems that these are the two most important aspects of a business when it comes to possible bankruptcy.

There was not a single financial ratio, which was used by both methods. However the MDA of this study used a ratio, which was previously used for a decision tree (Frydman, 1985) and the decision tree used a ratio, which was previously used for a MDA (Altman, 1993). So there are no different ratios for different methods after all.

It may be clear that profitability should be used, but there are different ways to measure profitability. Each financial ratio covers a different aspect of profitability. Maybe you should look at profitability in a different way for some sectors or countries and thus use a different financial ratio. So it is still wise to have several financial ratios to choose from.

Further research is possible regarding a more specific approach. This study used businesses from different sectors and even different countries. It is possible, however, that other sectors have other criteria of 'healthy' businesses. Further research is therefore possible in looking at only one specific kind of businesses.

Further research is also possible concerning variable V4. In Frydman's study (1985) a high value of V4 was a good thing; in this study a high value was a bad thing. It may be interesting to research how this difference is possible.

Finally further research is also possible in making a difference between Type I and Type II errors, which were mentioned at the end of section 4.3. It first needs to be researched exactly how much more a Type I error costs compared to Type II error. Then you can let the methods keep the different costs in mind and see how the cost-difference affects their performance.

BIBLIOGRAPHY/REFERENCES

Aczel, Amir D. and Jayavel Sounderpandian (2002) *Complete Business Statistics – 5ᵗʰ edition.* (International edition) McGraw-Hill Higer Education, New York

Altman, Edward. (September 1968). Financial ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy. In *Journal of Finance,* pp 589-609

Altman, Edward I. (1973). Predicting Railroad Bankruptcies in America. In *The Bell Journal of Economics and Management Science,* Vol. 2, No. 1 pp 184 - 211

Altman, Edward I. (1993) *Corporate Financial Distress and Bankruptcy: a Complete Guide to Predicting & Avoiding Distress and Profiting from Bankruptcy.* John Wiley & Sons Inc., Canada

Aziz, M. Adnan and Humayon A. Dar (2004) *Predicting Corporate Bankruptcy: Whither do We Stand?*
http://www.lboro.ac.uk/departments/ec/Reasearchpapers/2004/Departmental%20Paper%20_Aziz%20and%20Dar_.pdf

Beaver, William H. (1966) Financial ratios as Predictors of Failure. In *Empirical Research in Accounting: Selected Studies,* Supplement to Vol. IV, pp 71-111

Berzal, F. et al. (2004) Building multi-way decision trees with numerical attributes. In *Information Sciences* 165: pp 73 – 90

Frydman, Halina, Edward I. Altman and Duen-Li Kao. (1985) Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress. In *Journal of Finance*, 40: pp 269-291

Huberty, Carl J. (1994) *Applied discriminant analysis.* John Wiley & Sons Inc., Canada

McGurr, Paul T. and Sharon A. DeVany (1998) Predicting Business Failure of Retail Firms: An Analysis Using Mixed Industry Models. In *Journal of Business Research* nr. 43, pp 169-176

Mitchell, Tom M. (1997) *Machine Learning*. (International edition) McGraw-Hill, Singapore

Witten, Ian H. and Eibe Frank (2000) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Academic Press, San Diego.

APPENDIX A

Financial ratios used

V1: Cash flow / total assets

V2: Cash flow / total liabilities

V3: Cash flow / sales

V4: Cash / sales

V5: Current assets / current liabilities

V6: Current assets / total assets

V7: Earned income before interest and taxes / total assets

V8: Earned income before interest and taxes / (interest expenses on debt – interest income)

V9: log(total assets)

V10: Long-term debt / working capital*

V11: Long-term debt / total assets

V12: Total common equity / total liabilities

V13: Net income / total assets

V14: Net income / total liabilities

V15: Net income / sales

V16: Net income / working capital*

V17: Sales / working capital

V18: Inventory / sales

V19: Quick assets / current liabilities

V20: Quick assets / total assets

V21: Rate of annual growth in net sales = sales / sales of the previous year

V22: Retained earnings / total assets

V23: Sales / current assets

V24: Sales / total assets

V25: Stockholders capital / total capital

V26: Total liabilities / total assets

V27: Working capital / total assets

V28: Working capital / sales

*Working capital is defined as "current assets minus current liabilities".