“See5 Algorithm versus Discriminant Analysis. An Application to the Prediction of Insolvency in Spanish Non-life Insurance Companies”

AUTHORS	Zuleyka Díaz-Martínez José Fernández-Menéndez María Jesús Segovia-Vargas Eva María del Pozo-García

ARTICLE INFO	Zuleyka Díaz-Martínez, José Fernández-Menéndez, María Jesús Segovia-Vargas and Eva María del Pozo-García (2004). See5 Algorithm versus Discriminant Analysis. An Application to the Prediction of Insolvency in Spanish Non-life Insurance Companies. *Investment Management and Financial Innovations*, *1*(4)
RELEASED ON	Wednesday, 19 January 2005
JOURNAL	"Investment Management and Financial Innovations"
FOUNDER	LLC “Consulting Publishing Company “Business Perspectives”

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
0	0	0

See5 Algorithm versus Discriminant Analysis. An Application to the Prediction of Insolvency in Spanish Non-life Insurance Companies

Zuleyka Díaz-Martínez[1], José Fernández-Menéndez[2], María Jesús Segovia-Vargas[3], Eva María del Pozo-García[4]

Abstract

Prediction of insurance companies insolvency has arised as an important problem in the field of financial research, in order to protect both Society and customers and minimize the costs associated with this issue. Most methods applied in the past to tackle this question are traditional statistical techniques which use financial ratios as explicative variables. However, these variables often do not satisfy statistical assumptions, what complicates the application of the mentioned methods.

In this paper, a comparative study of the performance of a well-known parametric statistical technique (Linear Discriminant Analysis) and a non-parametric machine learning technique (See5) is carried out. We have applied the two methods to the problem of the prediction of insolvency among Spanish non-life insurance companies upon the basis of a set of financial ratios. Results indicate a higher performance of the machine learning technique, which shows that this method can be a useful tool to evaluate insolvency of insurance firms.

Key words: Insolvency, Insurance Companies, Discriminant Analysis, See5.

1. Introduction.

Unlike other financial problems, a great number of agents face business failure, so research in this topic has gained growing interest in the last decades.

Insolvency, early detection of financial distress, or conditions leading to insolvency among insurance companies have been a concern for many insurance regulators, investors, management, financial analysts, banks, auditors, policy holders and consumers. This concern has arised from the necessity of protecting the Society from the consequences of insurer insolvencies, as well as minimizing the costs associated with this problem such as the effects on state insurance guaranty funds or the responsibilities for management and auditors.

It has been widely recognized that there should be some kind of supervision on such entities to attempt to minimize the risk of failure. Nowadays, *Solvency II* project is intended to lead to the reform of the existing solvency rules in the European Union.

Many insolvency cases appeared after the insurance cycles of the 1970s and 1980s in the United States and the European Union. Several surveys have been devoted to identify the main causes of insurers' insolvency, in particular, the Müller Group Report (1997) analyses the main identified causes of insurance insolvencies in the European Union. The main reasons can be summarized as follows: operational risks (operational failure related to inexperienced or incompetent management, fraud); underwriting risks (inadequate reinsurance programme and failure to recover from reinsurers, higher losses due to rapid growth, excessive operating costs, poor underwriting process); insufficient provisions and imprudent investments.

On the other hand, many insurance companies, specially large ones, have developed internal risk models for a number of purposes. In Spain, where there is no a formal requirement, many insurers use internal risk models, developed in further or shorter degrees. In general, models are limited in their own nature and do not cover the whole of the risks. When the Spanish insurance

[1] Department of Financial Economics and Accounting I, Universidad Complutense de Madrid, Spain.
[2] Department of Business Administration, Universidad Complutense de Madrid, Spain.
[3] Department of Financial Economics and Accounting I, Universidad Complutense de Madrid. Campus de Somosaguas, Spain.
[4] Department of Financial Economics and Accounting I, Universidad Complutense de Madrid, Spain.

supervisor analyses models, difficulties appear around the accuracy in the level of reliability, congruency with accounting data, lack of harmonization, and limited level of application as a management tool at business unit level. Nevertheless, the Spanish supervisor is working on the development of an early warning system based on insurers' internal models (KPMG, 2002).

Therefore, developing new methods to tackle prudential supervision in insurance companies is a highly topical question, especially for all countries that belong to the European Union, like Spain's case.

A large number of methods have been proposed to predict business failure; needless to say, the special characteristics of the insurance sector have made most of them unfeasible, and just a few have been eventually applied to this sector. Most approaches applied to prediction of failure in insurance companies are statistical methods such as discriminant or logit analysis (Ambrose and Carroll, 1994; Bar-Niv and Smith, 1987; Mora, 1994; Sanchís *et al.*, 2003), which use financial ratios as explicative variables. However, this kind of variables does not usually satisfy statistical assumptions. In order to avoid these problems, a number of non-parametric techniques have been developed, most of them belong to the field of Machine Learning, such as neural networks (Serrano and Martín, 1993; Tam, 1991), which have been successfully applied to solve this kind of problems. However, their black-box character make them difficult to interpret, and hence the obtained results cannot be clearly analysed and related to the economical variables for discussion.

Other machine learning methods such as the one tested in this paper (See5 algorithm) are more useful in economic analysis, because the models provided by them can be easily understood and interpreted by human analysts. We will compare the accuracy of the See5 algorithm and the Linear Discriminant Analysis (LDA) to predict insolvency of insurance companies. Some prior researches have focused on the comparison of machine learning methods with the traditional statistical approaches (Altman *et al.*, 1994; De Andrés, 2001; Dimitras *et al*., 1999; Dizdarevic *et al.*, 1999), but only in few cases they have focused on the insurance sector (Martínez de Lejarza, 1999; Segovia *et al*., 2003).

In this paper a sample of Spanish non-life insurance firms is used. General financial ratios and those that are specifically proposed for evaluating insolvency inside insurance sector are employed. The results of See5 are very encouraging in comparison with LDA and show that this technique can be a useful tool for parts interested in evaluating insolvency of an insurance firm.

The rest of the paper is structured as follows: section 2 introduces some concepts related to the tested techniques. In section 3 we describe the data and input variables. In section 4 the results of the two approaches are presented. The discussion and comparison of these results are also provided in this section. Finally, section 5 closes the paper with some concluding remarks.

2. A brief overview of the tested techniques

2.1. The See5 algoithm

Perhaps learning systems based on decision trees are the easiest to use and understand among all machine learning methods. Moreover, the condition and ramification structure of a decision tree is suitable for classification problems. Prediction of insolvency is a kind of classification problem, as we try to classify firms into solvent or insolvent.

The automatic construction of decision trees begins with the studies developed in the social sciences by Morgan and Sonquist (1963) and Morgan and Messenger (1973). In statistics, the CART (Classification and Regression Trees) algorithm to generate decision trees proposed by Breiman et al. (1984) is one of the most important contributions. At around the same time decision tree induction was beginning to be used in the field of machine learning, notably by Quinlan (1979, 1983, 1986, 1988, 1993 and 1997), and in engineering by Henrichon and Fu (1969) and Sethi and Sarvarayudu (1982).

The successive branches of a decision tree achieve a series of exhaustive and exclusive partitions among the set of objects that a decision maker wants to classify. The main difference among the various algorithms used, is the criterion followed to carry out the partitions previously mentioned.

The See5 algorithm (Quinlan, 1997) is the latest version of the ID3 and C4.5 algorithms developed by this author in the last two decades. The criterion employed in See5 algorithm to carry out the partitions is based on some concepts from Information Theory and has been improved significantly over time.

The main idea shared with similar algorithms is to choose the variable that provides more information to realize the appropriate partition in each branch in order to classify the training set.

The information that provides a message or the achievement of a random variable x is inversely proportional to its probability (Reza, 1994). This quantity is usually measured in *bits* obtained through the relation: $\log_2 \frac{1}{p_x}$. The average of this relation for all the possible cases of the random variable x is called *entropy* of x:

$$H(x) = \sum_x p(x) \log_2 \frac{1}{p(x)}. \quad (1)$$

The entropy is a measure of the randomness or uncertainty of x or a measure of the average amount of information that is supplied by the knowledge of x.

In the same way, we can define the *joint entropy* of two random variables x and y: $H(x,y) = \sum_{x,y} p(x,y) \log_2 \frac{1}{p(x,y)}$, which represents the average amount of information supplied by the knowledge of x and y.

The *conditional entropy* of x given the variable y, $H(x/y)$, is defined as $H(x/y) = \sum_{x,y} p(x,y) \log_2 \frac{1}{p(x/y)}$, and this relation is a measure of the uncertainty of x when we know the variable y. That is, the amount of information necessary to know completely x when we know the information provided by y-variable. Naturally, $H(x/y) \le H(x)$, because if y-variable is known we have more information that can help us to reduce the uncertainty about x-variable. This reduction in the uncertainty is called *mutual information* between x and y:

$I(x;y) = H(x) - H(x/y)$, which is the information provided for one of the variables about the other one. It is always verified that $I(x;y) = I(y;x)$. Consequently the amounts of information that each variable provides about the other one are equal.

The mutual information is similar to covariance but the first one verifies some properties that make it preferable.

We can consider that x is a random variable that represents the category to which an object belongs. On the other hand, y_i, $i = 1,2,...,n$, represents the set of attributes that describe the objects we want to classify.

In a first time, Quinlan chose to make each partition the y_i-variable that provided the maximum information about x-variable, that is, he maximized $I(x;y_i)$ (this is called *gain* by him). Though this procedure provided good results, at the same time, it introduces a bias in favour of y_i-variables with many outcomes. In order to avoid this inconveniency, the subsequent releases of the algorithm choose the y_i-variable that maximizes the following relation called *gain ratio*,

$\frac{I(x\,;y_i)}{H(y_i)}$. This ratio represents the percentage of information provided by y_i that is useful in order to characterize *x*.

Notice that $I(x\,;y_i)$ should be big enough to prevent that an attribute could be only chosen because it has a low value for entropy, what would increase the *gain ratio*.

A common problem for the majority of rules and tree induction systems is that the generated models can be quite adapted to the training set, so the classification obtained will be nearly perfect. Consequently, the model derived will be very specific and in the case we want to classify new objects the model will not provide good results, moreover if the training set has noise. In this last case the model would be influenced by errors (noise) what would lead to a lack of generalization. This problem is known as *overfitting*.

The most frequent way of limiting this problem in the context of decision trees and set of rules consists in eliminating some conditions of the tree branches or the rules, in order to achieve more general models with these modifications in the algorithms. In the case of the decision trees, this procedure can be considered as a *pruning* process. This way we will increase the misclassifications in the training set but, at the same time, we probably will decrease the misclassifications in the test set that has not been used to derive the decision tree.

Quinlan incorporates a *post-pruning* method for an original fitted tree. This method consists in replacing a branch of the tree by a leaf, conditional on a predicted error rate. Suppose that there is a leaf that covers *N* objects and misclassifies *E* of them. This could be considered as a binomial distribution in which the experiment is repeated *N* times obtaining *E* errors. From this issue, the probability of error p_e is estimated, and it will be taken as the aforementioned predicted error rate. So it is necessary to estimate a confidence interval for the probability of error of the binomial distribution. The upper limit of this interval will be p_e (this is a pessimistic estimate).

Then, in the case of a leaf that covers *N* objects, the number of predicted errors will be $N \cdot P_e$. If we consider a branch instead of a leaf, the number of predicted errors associated with a branch will be just the sum of the predicted errors for its leaves. Therefore, a branch will be replaced by a leaf when the number of predicted errors for the last one is lower than the one for the branch.

Furthermore, See5 algorithm includes additional functions such as a method to change the obtained tree into a set of classification rules that are generally easier to understand than the tree. For a more detailed description of the features and working of See5 algorithm see Quinlan (1993 and 1997).

2.2. Linear Discriminant Analysis (LDA)

Although the classical methods of multivariate analysis have been superseded by methods from pattern recognition (Duda *et al.*, 2001; Venables and Ripley, 2002) they still have a place. In this paper, we have used one of these classical methods, LDA, as a benchmark to compare the performance of the machine learning method aforementioned, i.e., the See5 algorithm.

LDA was introduced by Fisher in 1936. The aim of LDA is to classify a new object according to the value of an estimated linear function of some attributes of that object. Geometrically, the new object is mapped to the same class that those located in its neighbourhood. LDA is subject to certain restrictive assumptions: each group follows a multivariate normal distribution, the covariance matrices of each group are identical (homoscedasticity) and prior probabilities and misclassifications costs are known. If this theoretical assumptions are violated, the results obtained may be questionable. When this happens, LDA could be seen as a non-parametric classification method, not optimal, but quite good in many situations (Krzanowski, 1996).

3. Methodological aspects

In this section, we show the main characteristics of the data and variables that will be used to develop our models. We have used the sample of Spanish firms used by Sanchís *et al.* (2003). This data sample consists of non-life insurance firms data five years prior to failure. The firms were in operation or went bankrupt between 1983 and 1994. From this period, 72 firms (36 failed and 36 non-failed) are selected. As a control measure, a failed firm is matched with a non failed one in terms of industry and size (premiums volume).

We have developed three models using data of one, two and three years before the firms declared bankruptcy. Thus, it has to be noted that the prediction of the insolvency achieved by each of them will be one, two and three years in advance, respectively. We refer to these models as *Model 1*, *Model 2* and *Model 3*.

In order to test the predictive accuracy of the models, we have split the set of original data to form the training sets and the holdout samples to validate the obtained models, i.e., the test sets. For *Model 1*, the training set consisted of 54 firms (27 failed and 27 non-failed firms) randomly generated. Therefore we have left 18 firms (9 failed and 9 non-failed) for testing. Sample size is different each year from the others, because data didn't exist for all the firms. In the following table these sample sizes are shown as well as the sizes of the training sets (randomly generated) to develop the models and the test sets to validate them.

Table 1

Sample sizes

Model	Sample size (number of firms)	Training set (number of firms)	Test set (number of firms)
1	72 (36 failed and 36 non-failed)	54 (27 failed and 27 non-failed)	18 (9 failed and 9 non-failed)
2	68 (34 failed and 34 non-failed)	52 (26 failed and 26 non-failed)	16 (8 failed and 8 non-failed)
3	54 (27 failed and 27 non-failed)	40 (20 failed and 20 non-failed)	14 (7 failed and 7 non-failed)

Like variables, each firm is described by 21 financial ratios that have come from a detailed analysis of the variables and previous bankruptcy studies for insurance companies. Table 2 shows the 21 ratios which describe the firms. Note that special financial characteristics of insurance companies require general financial ratios as well as those that are specifically proposed for evaluating insolvency of insurance sector.

Table 2

List of Ratios

RATIO	**DEFINITION**
1	2
R1	Working capital/ Total Assets
R2	Earnings before Taxes (EBT)/ (Capital+ Reserves)
R3	Investment Income/ Investments
R4	EBT*/ Total Liabilities EBT* = EBT+ Reserves for Depreciation+ Provisions + (Extraordinary Income-Extraordinary Charges)
R5	Earned Premiums/ (Capital+ Reserves)
R6	Earned Premiums Net of Reinsurance/ (Capital+ Reserves)
R7	Earned Premiums/ (Capital+ Reserves+ Technical Provisions)
R8	Earned Premiums Net of Reinsurance/ (Capital+ Reserves+ Technical Provisions)
R9	(Capital +Reserves)/ Total Liabilities

Table 2 (continuous)

1	2
R10	Technical Provisions/ (Capital + Reserves)
R11	Claims Incurred/ (Capital+ Reserves)
R12	Claims Incurred Net of Reinsurance/ (Capital+ Reserves)
R13	Claims Incurred / (Capital+ Reserves + Technical Provisions)
R14	Claims Incurred Net of Reinsurance/ (Capital+ Reserves+ Technical provisions)
R15	Combined Ratio 1 = (Claims Incurred/ Earned Premiums)+ (Other Charges and Commissions/ Other Income)
R16	Combined Ratio 2 = (Claims Incurred Net of Reinsurance/ Earned Premiums Net of Reinsurance)+ (Other Charges and Commissions/ Other income)
R17	(Claims Incurred + Other Charges and Commissions)/ Earned Premiums
R18	(Claims Incurred Net of Reinsurance + Other Charges and Commissions)/ Earned Premiums Net of Reinsurance
R19	Technical Provisions of Assigned Reinsurance/ Technical Provisions
R20	Claims Incurred / Earned Premiums
R21	Claims Incurred Net of Reinsurance / Earned Premiums Net of Reinsurance

The ratios have been calculated from the financial statements (balance sheets and income statements) issued one, two and three years before the firms declared bankruptcy. Ratios 15 and 16 have been removed in our study due to most of the firms not having "other income" so there is no sense to use them for an economic analysis. This reduces the total number of ratios to 19.

We want to mention that Linear Discriminant Analysis has been performed by using *SPSS 11.0* software, and the software used to implement See5 algorithm is *See5* by *RULEQUEST RESEARCH*.

4. Results

4.1. See5 algorithm

We have developed three models (three decision trees). We refer to them as *Model 1*, *Model 2* and *Model 3*. They have been developed by using, respectively, the previously mentioned training sets 1, 2 and 3, and we have tested them with the test sets 1, 2 and 3, as we can see below:

Model 1

```
R13 > 0.68:
:...R9 <= 0.59: failed (14)
: R9 > 0.59:
: :...R17 <= 0.99: failed (3)
: R17 > 0.99: healthy (3)
R13 <= 0.68:
:...R1 > 0.29: healthy (20/2)
 R1 <= 0.29:
 :...R2 > 0.04: failed (3)
 R2 <= 0.04:
 :...R6 > 0.64: healthy (3)
 R6 <= 0.64:
 :...R9 <= 0.85: failed (4)
 R9 > 0.85: healthy (4/1)


Evaluation on training data (54 cases):
```

```
  Decision Tree
  ----------------
  Size Errors

  8 3(5.6%)


  (a) (b) <-classified as
  ---- ----
  27 (a): class healthy
  3 24 (b): class failed


Evaluation on test data (18 cases):

  Decision Tree
  ----------------
  Size Errors

  8 5(27.8%)


  (a) (b) <-classified as
  ---- ----
  7 2 (a): class healthy
  3 6 (b): class failed
```

As we can see, only 6 ratios appear in the tree instead of the 19 initial ones. This indicates that these 6 variables are the most relevant ones for discrimination between solvent and insolvent firms in our sample and, consequently, it shows the strong support of this approach in feature selection. Our tree would be read in the following way:

- If the ratio R13 is greater than 0.68 and the ratio R9 is less than or equal to 0.59, then the company will be classified as "failed". This fact is verified by 14 firms in our sample.
- If the ratio R13 is greater than 0,68 and the ratio R9 is greater than 0,59 and the ratio R17 is less than or equal to 0,99, then the company will be classified as "failed", completing these conditions 3 companies.
- If...

and so on.

Every leaf of the tree is followed by a number *n* or *n*/*m*. The value of *n* is the number of cases in the sample that are mapped to this leaf, and *m* (if it appears) is the number of them that are classified incorrectly by the leaf.

The section under the tree concerns the evaluation of the decision tree, first on the cases of the training set from which it was constructed, and then on the new cases of the test set. The size of the tree is its number of leaves and the column headed “Errors” shows the number and percentage of cases misclassified. The tree, with 8 leaves, misclassifies 3 out of the 54 given cases, what implies an error rate of 5.6%, that is, 94.4% of correctly classified firms. Performance on the training cases is furher analyzed in a *confusion matrix* that pinpoints the kinds of errors made. A similar report of performance is given for the test cases, that shows the model’s accuracy on unseen test cases: an error rate of 27.8%, that is, 72.2% of correctly classified firms.

Though the tree we have derived is quite easy to understand, sometimes the trees developed are difficult to interpret. An important feature of See5 is its ability to generate unordered col-

lections of *if-then* rules, which are simpler and easier to understand than decision trees. The rules that are obtained starting from the previous tree are as follows:

```
Rules:

Rule 1: (20/2, lift 1.7)
 R1 > 0.29
 R13 <= 0.68
 -> class healthy [0.864]

Rule 2: (12/1, lift 1.7)
 R2 <= 0.04
 R6 > 0.64
 R13 <= 0.68
 -> class healthy [0.857]

Rule 3: (7/1, lift 1.6)
 R9 > 0.85
 -> class healthy [0.778]

Rule 4: (14, lift 1.9)
 R9 <= 0.59
 R13 > 0.68
 -> class failed [0.938]

Rule 5: (7, lift 1.8)
 R13 > 0.68
 R17 <= 0.99
 -> class failed [0.889]

Rule 6: (26/6, lift 1.5)
 R1 <= 0.29
 -> class failed [0.750]

Default class: healthy
```

Each rule consists of:

- Statistics (*n*, lift *x* or *n/m* lift *x*) that summarize the performance of the rule. Similarly to a leaf, *n* is the number of training cases covered by the rule and *m*, if it appears, shows how many of them do not belong to the class predicted by the rule. The lift *x* is the result of dividing the estimated accuracy of the rule by the relative frequency of the predicted class in the training set. The accuracy of the rule is estimated by the Laplace ratio $(n-m+1)/(n+2)$ (Clark and Boswell, 1991; Niblett, 1987).
- One or more conditions that must all be satisfied if the rule is to be applicable.
- A class predicted by the rule.
- A value between 0 and 1 that indicates the confidence with which this prediction is made.

There is also a *default class*, here "healthy", that is used when an object does not match any rule.

In this model, performance on the training cases and on the test cases is the same with this ruleset that with the previous tree, but it won't always be in this way.

Although these results are satisfactory, they can improve appealing to the *boosting* option that See5 incorporates, based on the research of Freund and Schapire (1997). Boosting is a technique for generating and combining multiple classifiers to improve predictive accuracy. Very

briefly, the idea is to generate several classifiers (either decision trees or rulesets) rather than just one. As the first step, a single decision tree or ruleset is constructed as before from the training data. This classifier will usually make mistakes on some cases in the data. When the second classifier is constructed, more attention is paid to these cases in an attempt to get them right. As a consequence, the second classifier will generally be different from the first one. It also will make errors on some cases, and these become the focus of attention during construction of the third classifier. This process continues for a pre-determined number of iterations or *trials*. Finally, when a new case is to be classified, each classifier votes for its predicted class and the votes are counted to determine the final class. The results obtained with this method are frequently very good.

In this way, starting from the previous tree, the results that are reached by means of the boosting option with 18 trials are shown in the following table, in percent of correctly classified firms.

Table 3

Boosting results for Model 1

Correct classifications	Training set	Test set
"Healthy" firms	100%	77.78%
"Failed" firms	100%	88.89%
Total	100%	83.33%

The sets of variables in the trees that constitute the rest of the models are shown in the next table. This table also displays performance in the training cases and in the test cases, in percent of correctly classified firms. The trees 2 and 3 have been pruned, because previously we observed that the error rates were quite smaller in the training sets than in the test sets, and this could be due to an overfitting problem. However, pruning doesn't improve performance on the first tree.

Table 4

See5 results

Model	Set of variables	Size of the tree	Correct classifications			
			Training set		Test set	
			"Healthy" firms	"Failed" firms	"Healthy" firms	"Failed" firms
1	R13, R9, R17, R1, R2, R6	8	100%	88.89%	77.78%	66.77%
			Total: 94.44%		Total: 72.22%	
2	R1, R13, R20, R7, R3	6	96.15%	84.62%	87.5%	75%
			Total: 90.39%		Total: 81.25%	
3	R4, R19, R1	5	100%	70%	100%	57.14%
			Total: 85%		Total: 78.57%	

As we have previously mentioned, in many occasions the classifications accuracy can be improved by means of boosting. For example, for the model 2, the results obtained by means of boosting with 11 trials are shown in the following table, in percent of correctly classified firms.

Table 5

Boosting results for Model 2

Correct classifications	Training set	Test set
"Healthy" firms	100%	87.5%
"Failed" firms	100%	87.5%
Total	100%	87.5%

4.2. Linear Discriminant Analysis

As a previous step, we have detected the univariate outliers and due to the shortage of data they have been substituted by the median of the correspondent attribute instead of being eliminated. We have verified that the great majority of variables follows a normal univariate distribution and that these variables are not discriminatory, in other words, their means are not significantly different between groups.

Next, the discriminant analysis has been carried out using the *stepwise* method for the selection of the variables to introduce in the models. The variables have been always chosen from those that present the most significant difference of means between groups. Furthermore, we have checked using M Box estimator that homoscedasticity assumption is not satisfied.

The obtained results are shown in the following table.

Table 6

LDA results

Model	Set of variables	Correct classifications			
		Training set		Test set	
		"Healthy" firms	"Failed" firms	"Healthy" firms	"Failed" firms
1	R1, R7	77.78%	59.26%	77.78%	44.44%
		Total: 68.52%		Total: 61.11%	
2	R12, R17	73.08%	65.38%	25%	75%
		Total: 69.23%		Total: 50%	
3	R4	90%	60%	57.14%	42.86%
		Total: 75%		Total: 50%	

4.3. Results comparison

In order to make easier the comparison between the two approaches, Table 7 shows the results for the test samples, in percent of correctly classified firms.

Roughly speaking See5 outperforms clearly LDA. In fact, the last one works as a random classificator in models 2 and 3.

Machine learning technique selects many more ratios than LDA, so it makes a better use of the available information which leads to a higher correct classification rate. Probably the structure of data space is too much complex to achieve a good classification with a linear hypersurface as LDA does it. The more complex rules generated by machine learning technique adapt better to data structure. It is a very powerful tool to capture the peculiarities of data in detail.

Moreover, as we could see previously, results of See5 for some models can be clearly improved by means of boosting.

Naturally, the ratios which appear in the solutions are not the same for each year, because the prediction of the insolvency achieved by each model will be one, two and three years in advance, respectively. We can consider that the ratios which appear in the two solutions achieved by See5 and LDA are highly discriminatory variables between solvent and insolvent firms. Consequently, those parts interested in evaluating the solvency of non-life insurance companies should take into account the following questions:

a) R1 – One of the most important questions in order to assure the proper functioning of any firm is the need of having sufficient liquidity. But in the case of an insurance firm, the lack of liquidity should not arise due to "productive activity inversion" which implies that premiums are paid in before claims occur. If an insurance firm cannot pay the incurred claims, the clients and public in general could lose faith in that company. On the other hand, this ratio is a measure of financial equilibrium if it is positive as it implies that the working capital is also positive.

b) R4 – This ratio is a general measure of profitability. The variable that appears in the numerator is the cashflow (cashflow plus extraordinary results) because sometimes it would be better to use this variable than profits because the former is less manipulated than the latter. In any case, it is necessary to generate sufficient profitability to follow a right self-financing.

c) R7 – This ratio is considered as a "solvency ratio in strictu sense". The numerator shows the risk exposure through earned premiums and the denominator shows the real financial support because technical provisions are considered along with capital and reserves. This demonstrates the need of having sufficient shareholder' funds and the need of complying correctly with the technical provisions to guarantee the financial viability of the insurance company. This ratio belongs to IRIS (Insurance Regulatory Information System) ratios. IRIS ratios are tests developed by the National Association of Insurance Commissioners (USA) as an early warning system.

d) R17 – Combined ratio. It is a traditional measuring of underwriting profitability and it indicates if the firm is following a correct rating in order to calculate right premiums that take into account the whole costs.

Table 7

Results comparison

Model	Technique	Set of variables	Correct classifications	
			"Healthy" firms	"Failed" firms
1	See5	R13, R9, R17, R1, R2, R6	77.78%	66.77%
			Total: 72.22%	
	LDA	R1, R7	77.78%	44.44%
			Total: 61.11%	
2	See5	R1, R13, R20, R7, R3	87.5%	75%
			Total: 81.25%	
	LDA	R12, R17	25%	75%
			Total: 50%	
3	See5	R4, R19, R1	100%	57.14%
			Total: 78.57%	
	LDA	R4	57.14%	42.86%
			Total: 50%	

4. Conclusions

In this paper we have applied the See5 algorithm and the Linear Discriminant Analysis to a real problem of classification of non-life insurance companies into healthy or failed. We have used a sample of Spanish companies described by a set of 19 financial ratios and we have compared the obtained results for each model.

In the light of the experiments carried out, the machine learning approach (See5) is a competitive alternative to existing bankruptcy prediction models in insurance sector and has great potential capacities that undoubtedly make it attractive for application to the field of business classification.

Our empirical results show that this method offers better predictive accuracy than the Linear Discriminant Analysis we have developed. Moreover, the See5 algorithm doesn't require the adoption of restrictive assumptions about the characteristics of statistical distributions of the variables and errors of the models. Furthermore, the decision models provided by this technique are easy to understand and interpret.

In practical terms, the trees and decision rules generated could be used to preselect companies to examine more thoroughly, quickly and inexpensively, thereby, managing the financial user's time efficiently. They can also be used to check and monitor insurance firms as a "warning

system" for insurance regulators, investors, managers, financial analysts, banks, auditors, policy holders and consumers.

However, our work has some limitations, such as the few available cases and the uncertain quality of some information. Furthermore, if we want to use these models for predicting insolvency, we should take into account that they have been developed without including some aspects which could be relevant for this issue, such as size and industry.

But despite these problems, our focus is to show the suitability of this machine learning technique as a support decision method for insurance sector. In short, we believe that this method, without replacing analyst's opinion and in combination with another ones, will play a bright role in the decision making process inside insurance sector.

References

1. Altman, E.I., Marco, G. and Varetto, F. (1994): "Corporate distress diagnosis: comparisions using linear discriminant analysis and neural networks (the Italian experience)", *Journal of Banking and Finance*, 18, 505-529.
2. Ambrose, J.M. and Carroll, A.M. (1994): "Using Best's Ratings in Life Insurer Insolvency Prediction", *The Journal of Risk and Insurance*, 61 (2), 317-327.
3. Bar-niv, R. and Smith, M.L. (1987): "Underwriting, Investment and Solvency", *Journal of Insurance Regulation*, 5, 409-428.
4. Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984): *Classification and regression trees*, Wadsworth, Belmont.
5. Clark, P. and Boswell, R. (1991): "Rule Induction with CN2: Some Recent Improvements", in Kodratoff, Y. (Ed.): *Machine Learning - Proceedings of the Fifth European Conference (EWSL-91)*, Springer-Verlag, Berlin, 151-163.
6. De Andrés, J. (2001): "Statistical Techniques vs. SEE5 Algorithm. An Application to a Small Business Environment", *International Journal of Digital Accounting Research*, 1 (2), 153-179.
7. Dimitras, A.I., Slowinski, R., Susmaga, R. and Zopounidis, C. (1999): "Business failure prediction using Rough Sets", *European Journal of Operational Research,* 114, 263-280.
8. Dizdarevic, S., Larrañaga, P., Peña, J.M., Sierra, B., Gallego, M.J. and Lozano, J.A. (1999): "Predicción del fracaso empresarial mediante la combinación de clasificadores provenientes de la estadística y el aprendizaje automático", in Bonsón, E. (Ed.): *Tecnologías Inteligentes para la Gestión Empresarial*, RA-MA Editorial, Madrid, 71-113.
9. Duda, R.O., Hart, P.E. and Stork, D.G. (2001): *Pattern Classification*, John Wiley & Sons, Inc., New York.
10. Freund, Y. and Schapire, R.E. (1997): "A decision-theoretic generalization of on-line learning and an application to boosting", *Journal of Computer and System Sciences*, 55(1), 119-139.
11. Henrichon, Jr., E.G. and Fu, K.S. (1969): "A nonparametric partitioning procedure for pattern classification", *IEEE Transactions on Computers*, 18, 614-624.
12. KPMG (2002): "Study into the methodologies to assess the overall financial position of an insurance undertaking from the perspective of prudential supervision", (in http://europa.eu.int/comm/internal_market/en/finances/insur/index.htm).
13. Krzanowski, W.J. (1996): *Principles of Multivariate Analysis. A User's Perspective*, Oxford University Press, Oxford.
14. Martínez de Lejarza, I. (1999): "Previsión del fracaso empresarial mediante redes neuronales: un estudio comparativo con el análisis discriminante", in Bonsón, E. (Ed.): *Tecnologías Inteligentes para la Gestión Empresarial*, RA-MA Editorial, Madrid, 53-70.
15. Mora, A. (1994): "Los modelos de predicción del fracaso empresarial: una aplicación empírica del logit", *Revista Española de Financiación y Contabilidad*, 78, enero-marzo, 203-233.

16. Morgan, J.N. and MessengeR, R.C. (1973): *THAID: a Sequential Search Program for the Analysis of Nominal Scale Dependent Variables*, Survey Research Center, Institute for Social Research, University of Michigan.
17. Morgan, J.N. and Sonquist, J.A. (1963): "Problems in the analysis of survey data, and a proposal", *Journal of the American Statistical Association*, 58, 415-434.
18. Müller Group (1997): *Müller Group Report. 1997. Solvency of insurance undertakings*, Conference of Insurance Supervisory Authorities of The Member States of The European Union.
19. Niblett, T. (1987): "Constructing decision trees in noisy domains", in Bratko, I. and Lavrač, N. (Eds.): *Progress in Machine Learning (proceedings of the 2nd European Working Session on Learning)*, Sigma, Wilmslow, UK, 67-78.
20. Quinlan, J.R. (1979): "Discovering rules by induction from large collections of examples", in Michie, D. (Ed.): *Expert systems in the microelectronic age*, Edimburgh University Press, Edimburgh.
21. Quinlan, J.R. (1983): "Learning efficient classification procedures", in *Machine learning: an Artificial Intelligence approach*, Tioga Press, Palo Alto.
22. Quinlan, J.R. (1986): "Induction of decision trees", *Machine Learning*, 1 (1), 81-106.
23. Quinlan, J.R. (1988): "Decision trees and multivalued attributes", *Machine Intelligence*, 11, 305-318.
24. Quinlan, J.R. (1993): C4.5: Programs for machine learning, Morgan Kaufmann Publishers, Inc., California.
25. Quinlan, J.R. (1997) : *See5* (available from http://www.rulequest.com/see5-info.html).
26. Reza, F.M. (1994) : *An introduction to Information Theory*, Dover Publications, Inc., New York.
27. Sanchís, A., Gil, J.A. and Heras, A. (2003): "El análisis discriminante en la previsión de la insolvencia en las empresas de seguros no vida", *Revista Española de Financiación y Contabilidad*, 116, enero-marzo, 183-233.
28. Segovia, M.J., Gil, J.A., Heras, A. and Vilar, J.L. (2003): "La metodología Rough Set frente al Análisis Discriminante en los problemas de clasificación multiatributo", XI Jornadas ASEPUMA, Oviedo, Spain.
29. Serrano, C. and Martín, B. (1993): "Predicción de la crisis bancaria mediante el empleo de redes neuronales artificiales", *Revista Española de Financiación y Contabilidad*, 74, enero-marzo, 153-176.
30. Sethi, I.K. and Sarvarayudu, G.P.R. (1982): "Hierarchical classifier design using mutual information", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4, 441-445.
31. Tam, K.Y. (1991): "Neural network models and the prediction of bankruptcy", *Omega*, 19 (5), 429-445.
32. Venables, W.N. and Ripley, B.D. (2002): *Modern Applied Statistics with S*, Springer-Verlag, New York.