

Efficiency of Machine Learning Techniques in Bankruptcy Prediction

S. Kotsiantis¹, D. Tzelepis¹, E. Koumanakos² and V. Tampakas¹

¹Department of Accounting, Technological Educational Institute of Patras, Greece

²National Bank of Greece, Credit Division, Greece

sotos@math.upatras.gr, tzelepis@upatras.gr, koumanakos.e@nbg.gr, tampakas@teipat.gr

Abstract

Prediction of corporate bankruptcy is a phenomenon of increasing interest to investors/creditors, borrowing firms, and governments alike. Timely identification of firms' impending failure is indeed desirable. The scope of the research reported here is to investigate the efficiency of machine learning techniques in such an environment. To this end, a number of experiments have been conducted using representative learning algorithms, which were trained using a data set of 150 failed and solvent Greek firms in the recent period 2003-2004*. It was found that learning algorithms could enable users to predict bankruptcies with satisfying accuracy long before the final bankruptcy.

Keywords: supervised machine learning algorithms, prediction of bankruptcy.

1 INTRODUCTION

The problem of Bankruptcy prediction is a classical one in the financial literature (see e.g. Altman (1993) for a review). The main impact of Bankruptcy prediction is in bank lending. Banks need to predict the possibility of default of a potential counterparty before they extend a loan. This can lead to sounder lending decisions, and therefore result in significant savings.

There are two main approaches to loan default/bankruptcy prediction. The first approach, the structural approach, is based on modelling the underlying dynamics of interest rates and firm characteristics and deriving the default probability based on these dynamics. The second approach is the empirical or the statistical approach. Instead of modelling the relationship of default with the characteristics of a firm, this relationship is learned from the data. Overall, Morris (1998) provides the most comprehensive review of to date bankruptcy prediction models. The book offers a very useful discussion on many important prediction techniques and their empirical use. It, however, lacks a deserved discussion on some important machine learning models. The focus of this article is on the empirical approach, especially the use of machine learning in bankruptcy prediction (Pompe and Feelders, 1997), (Kotsiantis et al., 2004).

* The Project is co-funded by the European Social Fund and National Resources

The availability of vast amounts of historical data in recent years, coupled with the enormous processing power of desktop computers, has enabled the use of automated systems to assist in complex decision making environments. The automated system examines financial ratios as predictors of performance, and assesses posterior probabilities of financial health (alternatively, financial distress).

Many studies have been conducted for bankruptcy prediction using models such as neural networks (Zhang et al., 1999), instance based learners (Park and Han, 2002), Bayesian models (Sarkar and Sriram, 2001), rule learners (Thomaidis et al., 1999), decision trees algorithms (Mckee and Greenstein, 2000) and Support Vector Machines (Shin et al., 2005). These methods of corporate bankruptcy prediction have their own strengths and weaknesses and, hence, choosing a particular model may not be straightforward.

Searching for best distress prediction models is still in progress. This study provides a critical analysis of most commonly used corporate bankruptcy learning models. Thus, we use a representative algorithm for each one of the most common machine learning techniques so as to investigate the efficiency of ML techniques in such an environment. Indeed, it is proved that learning algorithms can predict bankruptcy with satisfying accuracy long before the final bankruptcy.

The following section describes the data set of our study and the feature selection process. Some elementary Machine Learning definitions and a more detailed description of the used techniques and algorithms are given in section 3. It is hoped that the brief introduction to methodological details of these models would be of great use to those with recent interest in this field. Section 4 presents the experimental results for the representative compared algorithms. Finally, section 5 discusses the conclusions and some future research directions.

2 DATA DESCRIPTION

Logit analysis, probit analysis, and the linear probability model are the most commonly used techniques applied in the Greek context in several studies over the past two decades (Papoulias and Theodossiou, 1992), (Vranas, 1992) and (Negakis, 1995). Recently the performance of alternative non-parametric approaches has been explored in the Greek context to overcome the aforementioned shortcomings of the statistical and econometric techniques such as rough sets (Dimitras et al., 1999) and multicriteria discrimination method (Doumpos and Zopounidis, 1999). As we have already mentioned, in this paper we analyzed the performance of several machine learning models on the problem of Bankruptcy prediction in the Greek context.

Bankruptcy filings in the years 2003 and 2004 were provided directly from the National Bank of Greece directories and the business database of the financial information services company called ICAP, in Greece. Financial statement data for the fiscal years prior to bankruptcy were obtained from ICAP financial directories. The financial statements of these firms were collected for a period of three years. The critical year of failure denoted as year 0, three years before as year -3 and year -1 is the final year prior to bankruptcy filing.

As the control sample, each selected bankrupt firm was matched with two non-bankrupt (healthy) firms of exactly the same industry, by carefully comparing the year of the reported data (year -1) assets size and the number of employees. The selected non-bankrupt corporations were within 20% of the selection criteria.

Following the prior literature, we examine the probability of a firm's initial filing for bankruptcy and eliminate any observations for a firm after it has filed for bankruptcy during our sample period.

Our final bankruptcy sample consists of 50 initial bankruptcies in the year period 2003-2004 and is similar in size but more complete and recent compared to previous studies. The final pooled sample of failed and solvent firms is composed of 150 individual firms with financial data for a three-year period, which attributes 450 firm-year observations.

Table 1 presents the sample distribution of the bankrupted firms across industries and calendar years. The 50 failed firms are distributed across 24 industries.

Industry	Year		Total
	2003	2004	
Wholesale Trade	2	4	6
Publishing & Printing	1		1
Information Technology		1	1
Advertisement	1	2	3
Plastic and Rubber		1	1
Clothing	2	2	4
Other Services		1	1
Industrial Minerals		1	1
Machinery		2	2
Private Education	1		1
Textiles	3	1	4
Food		2	2
Supermarkets		1	1
Retail Trade	3	7	10
Freight Forwarding	1		1
Constructions	2		2
Telecommunications		2	2
Logistics		1	1
Electronics Equipment		1	1
Agriculture and Farming	1		1
Motor Vehicle Trade & Maintenance	1		1
Metal Products	1		1
Health Services		1	1
Restaurants		1	1
Total	19	31	50

Table 1. Distribution of Bankruptcies by Year and Industry

Through extensive literature review on bankruptcy prediction about 50 financial ratios were recorded. The final set of the calculated input variables is 21 due to missing financial data and financial ratio duplication. Table 2 provides a brief description of the financial variables used in the present study classified in 5 groups.

Category	Independent variables	Variable Description
<i>Profitability variables</i>	<i>OPIMAR</i>	Operating income divided by net sales
	<i>NIMAR</i>	Net income divided by sales
	<i>GIMAR</i>	Gross income divided by sales
	<i>ROE</i>	Net income pre tax divided by Shareholder's equity capital
	<i>ROCE</i>	Net income pre tax divided by capital employed
<i>Liquidity-Leverage variables</i>	<i>EQ/CE</i>	Shareholder's equity to capital employed
	<i>CE/NFA</i>	Capital employed to net fixed assets
	<i>TD/EQ</i>	Total debt to shareholder's equity capital
	<i>CA/CL</i>	Current assets to current liabilities
	<i>QA/CL</i>	Quick assets to current liabilities
	<i>WC/TA</i>	Working capital divided by total assets
<i>Efficiency variables</i>	<i>COLPER</i>	Average collection period for receivables
	<i>INVTURN</i>	Average turnover period for inventories
	<i>PAYPER</i>	Average payment period to creditors
	<i>S/EQ</i>	Sales divided by Shareholder's equity capital
	<i>S/CE</i>	Sales divided by capital employed
	<i>S/TA</i>	Sales divided by Total Assets
<i>Growth variables</i>	<i>GRTA</i>	Growth rate of total assets $(TA_t - TA_{t-1}) / (ABS(TA_t) + ABS(TA_{t-1}))$
	<i>GRNI</i>	Growth rate of net income
	<i>GRNS</i>	Growth rate of net sales
Size variable	<i>SIZE</i>	Size of firm is the $\ln(\text{Total Assets}/\text{GDP price index})$

Table 2. Research Variables description

Several previous studies had used a brute empirical approach of initially choosing variables followed by Stepwise procedure to select the variables in the final discriminant function. However, these studies are limited in their ability to provide generalized results as to what financial variables can consistently predict financial distress. Therefore, in this study, a statistical approach is taken to select the best set of variables for predicting bankruptcy.

2.1 Attribute Selection

In an attempt to show how much each attribute influences the induction, we rank the influence of each one according to a statistical measure – ReliefF (Sikonja and Kononenko, 1997). In general, ReliefF assign relevance to features based on their ability to disambiguate similar samples, where similarity is defined by proximity in feature space. Relevant features accumulate high positive weights, while irrelevant features retain near-zero weights.

The average ReliefF score of each attribute according to our dataset are presented in Table 3. The larger the value of the ReliefF scores is, the more influence of the attribute in the induction.

Attribute	Average Score
WC/TA	0.035 ± 0.003
EQ/CE	0.011 ± 0.002
GRNI	0.012 ± 0.005
SIZE	0.006 ± 0.003
GRTA	0.004 ± 0.001
TD/EQ	0.003 ± 0.001
S/CE	0.003 ± 0.001
COLPER	0.002 ± 0.001
S/EQ	0.002 ± 0.001
CE/NFA	0.002 ± 0.001
PAYPER	0.001 ± 0
INVTURN	0.001 ± 0.001
GIMAR	0.001 ± 0.001
CA/CL	0
NIMAR	0
ROCE	0
GRNS	0
ROE	0
QA/CL	0
S/TA	0
OPIMAR	0

Table 3. Average ReliefF score of each attribute

Thus, the attributes that mostly influence the induction are: WC/TA, EQ/CE and GRNI. It seems that the attributes: CA/CL, NIMAR, ROCE, GRNS, ROE, QA/CL, S/TA and OPIMAR do not influence the induction at all. For this reason, the previous attributes are not included in the training set of the learning algorithms.

In order to examine the usage of the learning techniques in this domain, the application of six most common machine learning techniques namely Decision Trees (Murthy, 1998), Artificial Neural Networks (Mitchell, 1997), Bayesian Nets (Jensen, 1996), Instance-Based Learning (Aha, 1997), Rule-learning (Furnkranz, 1999) and Support Vector Machines (Burges, 1998) are used. In the next section we give some elementary Machine Learning definitions and we briefly describe these supervised machine-learning techniques.

3 MACHINE LEARNING ISSUES

Each company is described by a set of variables (predictors) x , such as financial ratios, and its class y that can be either $y = -1$ ('successful') or $y = 1$ ('bankrupt'). Initially, an unknown classifier function $f : x \rightarrow y$ is estimated on a training set of companies (x_i, y_i) , $i = 1, \dots, n$. The training set represents the data for companies which are known to have survived or gone bankrupt. Finally, f is applied to other companies.

A recent overview of existing work in decision trees is provided by (Murthy, 1998). Decision trees are trees that classify examples by sorting them based on attribute

values. Each node in a decision tree represents an attribute in an example to be classified, and each branch represents a value that the node can take. Examples are classified starting at the root node and sorting them based on their attribute values. The attribute that best divides the training data would be the root node of the tree. The same process is then repeated on each partition of the divided data, creating sub trees until the training data sets are divided into subsets of the same class. However, a decision tree is said to overfit training data if there exists another hypothesis h' that has a larger error than h when tested on the training data, but a smaller error than h when tested on the entire data set. For this reason, there are two common approaches that decision tree algorithms can use to avoid overfitting training data: 1) Stop the training algorithm before it reaches a point in which it perfectly fits the training data, 2) Prune the induced decision tree.

Artificial Neural Networks (ANNs) are another method of inductive learning based on computational models of biological neurons and networks of neurons as found in the central nervous system of humans (Mitchell, 1997). A multi layer neural network consists of large number of units (neurons) joined together in a pattern of connections. Units in a net are usually segregated into three classes: input units, which receive information to be processed, output units where the results of the processing are found, and units in between called hidden units. Classification with a neural network takes place in two distinct phases. First, the network is trained on a set of paired data to determine the input-output mapping. The weights of the connections between neurons are then fixed and the network is used to determine the classifications of a new set of data.

An excellent book about the Bayesian networks is provided by (Jensen, 1996). A Bayesian network is a graphical model for probabilistic relationships among a set of attributes. The Bayesian network structure S is a directed acyclic graph (DAG) and the nodes in S are in one-to-one correspondence with the attributes. The arcs represent casual influences among the variables while the *lack* of possible arcs in S encodes conditional independencies. Moreover, an attribute (node) is conditionally independent of its non-descendants given its parents. Using a suitable training method, one can induce the structure of the Bayesian Network from a given training set. In spite of the remarkable power of the Bayesian Networks, there is an inherent limitation. This is the computational difficulty of exploring a previously unknown network. Given a problem described by n attributes, the number of possible structure hypotheses is more than exponential in n . In the case that the structure is unknown but we can assume that the data is complete, the most common approach is to introduce a scoring function (or a score) that evaluates the "fitness" of networks with respect to the training data, and then to search for the best network (according to this score). The classifier based on this network and on the given set of attributes X_1, X_2, \dots, X_n , returns the label c that maximizes the posterior probability $p(c | X_1, X_2, \dots, X_n)$.

Instance-based learning algorithms belong in the category of lazy-learning algorithms (Aha, 1997), as they defer in the induction or generalization process until classification is performed. One of the most straightforward instance-based learning algorithms is the nearest neighbour algorithm. K-Nearest Neighbour (kNN) is based on the principal that the examples within a data set will generally exist in close proximity with other examples that have similar properties. If the examples are tagged with a classification label, then the value of the label of an unclassified

example can be determined by observing the class of its nearest neighbours. The absolute position of the examples within this space is not as significant as the relative distance between examples. This relative distance is determined using a distance metric. Ideally, the distance metric must minimize the distance between two similarly classified examples, while maximizing the distance between examples of different classes.

In rule induction systems, a decision rule is defined as a sequence of Boolean clauses linked by logical AND operators that together imply membership in a particular class (Furnkranz, 1999). The general goal is to construct the smallest rule-set that is consistent with the training data. A large number of learned rules is usually a sign that the learning algorithm tries to “remember” the training set, instead of discovering the assumptions that govern it. During classification, the left hand sides of the rules are applied sequentially until one of them evaluates to true, and then the implied class label from the right hand side of the rule is offered as the class prediction.

The SVM technique revolves around the notion of a ‘margin’, either side of a hyperplane that separates two data classes. Maximizing the margin, and thereby creating the largest possible distance between the separating hyperplane and the instances on either side of it, is proven to reduce an upper bound on the expected generalization error (Burges, 1998). Nevertheless, most real-world problems involve non-separable data for which no hyperplane exists that successfully separates the positive from negative instances in the training set. One solution to the inseparability problem is to map the data into a higher-dimensional space and define a separating hyperplane there. This higher-dimensional space is called the feature space, as opposed to the input space occupied by the training instances. Generally, with an appropriately chosen feature space of sufficient dimensionality, any consistent training set can be made separable.

For the purpose of the present study, a representative algorithm for each described machine learning technique was selected.

3.1 Brief description of the used machine learning algorithms

The most commonly used C4.5 algorithm (Quinlan, 1993) was the representative of the decision trees in our study. At each level in the partitioning process a statistical property known as *information gain* is used by C4.5 algorithm to determine which attribute best divides the training examples. The approach that C4.5 algorithm uses to avoid overfitting is by converting the decision tree into a set of rules (one for each path from the root node to a leaf) and then each rule is generalized by removing any of its conditions that will improve the estimated accuracy of the rule.

Naive Bayes algorithm was the representative of the Bayesian networks (Domingos and Pazzani 1997). It is a simple learning that captures the assumption that every attribute is independent from the rest of the attributes, given the state of the class attribute.

We also used the Local Decision Stump (DS) algorithm, with Euclidean distance as distance metric as instance based learner (Aha, 1997). Attributes with missing values are given imputed values so that comparisons can be made between every pair of examples on all attributes.

The RIPPER (Cohen, 1995) algorithm was the representative of the rule-learning techniques because it is one of the most usually used methods that produce

classification rules. RIPPER forms rules through a process of repeated growing and pruning. During the growing phase the rules are made more restrictive in order to fit the training data as closely as possible. During the pruning phase, the rules are made less restrictive in order to avoid overfitting, which can cause poor performance on unseen examples. The grow heuristic used in RIPPER is the information gain function.

A well-known learning algorithm for building a neural network - RBF algorithm (Mitchell, 1997) - was the representative of the ANNs. Finally, the Sequential Minimal Optimization (or SMO) algorithm was the representative of the SVMs in our study because it is one of the fastest methods to train SVMs (Platt, 1999).

4 EXPERIMENTS AND RESULTS

To facilitate the presentation and discussion of the results, each year prior to financial distress is denoted as year -1, year -2, year -3, Year -1 refers to the first year prior to financial distress (e.g., for the firms that faced financial distress in 2004, year -1 refers to 2003); year -2 refers to the second year prior to financial distress (e.g., for the firms that faced financial distress in 2004, year -2 refers to 2002), etc.

The objective of learning classifications from sample data is to classify and predict successfully on new data. The most commonly used measure of success or failure is a classifier's error rate. The error rate is statistically defined as the error rate of the classifier on an asymptotically large number of new cases that converge in the limit to the actual population distribution. The classification accuracy is defined as the ratio of the number of the correct classifications to the number of cases examined.

In Table 4, there is the classification accuracy for each representative learning algorithm for each examined year. In order to calculate the classifiers accuracy for our experiments, the whole training sets (one for each year - *i*) was divided into ten mutually exclusive and equal-sized subsets and for each sub-set the model was trained on the union of all of the other subsets. Then, cross validation was run 10 times for each algorithm and the average value of the 10-cross validations was calculated.

	Naive Bayes	Local DS	RIPPER	C4.5	SMO	RBF
Year (-3)	68.75	68.31	65.19	63.81	67.23	68.68
Year (-2)	71.08	68.37	66.94	63.89	68.17	70.14
Year (-1)	71.83	72.90	74.32	72.35	68.61	71.17

Table 4. Accuracy of the algorithms in each testing step

It was found that learning algorithms could enable users to predict bankruptcies with satisfying accuracy long before the final bankruptcy. The experts are in the position to know 3 years before, which of the industries will bankrupt or not with sufficient precision, which reaches the 68% in the initial forecasts (3 years before the examined year) and exceeds the 74% the last year. However, because of the small size of training set it is not useful to run a statistical test (t-test) in order to compare

these algorithms. The resulting differences between algorithms are not statistically significant. In a following study, we will include more data for this proposes.

It should be emphasized that a machine learning tool cannot, and should not, entirely replace professional judgment. It should be used to provide auditors with objective information to help them with their reporting decision on a client in the presence of financial distress and going-concern contingencies. It must be noted that many important qualitative variables such as management's ability and future plans (that could potentially mitigate the stress faced by a firm) are not formally incorporated into the models.

5 CONCLUSION AND FUTURE WORK

With the help of machine learning techniques, the experts are in the position to know 3 years before, which of the industries will bankrupt or not with sufficient precision, which reaches the 68% in the initial forecasts (3 years before the examined year) and exceeds the 74% the last year. For this reason, a prototype version of a software support tool can be constructed implementing learning algorithms. Tracking progress is a time-consuming job that can be handled automatically by such a tool. While the experts will still have an essential role in monitoring and evaluating progress, the tool can compile the data required for reasonable and efficient monitoring. The prediction model derived from the present study suggests the importance of liquidity defined by the ratio working capital to total assets, capital structure defined as equity to capital employed and profitability growth defined as net income growth.

However, there were a number of limitations in this study that must be noted. First, the sample size was relatively small. Thus, the generalization of the research results is somewhat limited. The second limitation was that only financial ratio variables were included in this study. There may be other important key quantitative variables (i.e., stock data, market value, age) as well as qualitative variables (leadership, reputation, type of ownership, etc.) and there is rich literature in organization theory that reports the importance of these variables. Above all, the main limitation of this paper and of the great majority of the preceding studies that should be underlined is the absence of a robust theoretical framework for the selection of the potential explanatory variables of corporate failure. These limitations open up a wide opportunity for future research.

Moreover, the algorithms discussed here aim at achieving high classification accuracy that is lower error rate in the prediction of unseen examples. However, these algorithms do not distinguish the types of errors. That is for these algorithms classifying a bankrupt case as a non-bankrupt has the same error as classifying a non-bankrupt as a bankrupt. However, in real life, these costs may not be the same for the decision maker. For example, the cost of predicting a case as non-bankrupt that is actually bankrupt may be higher than vice versa. Therefore, as a future work we plan to extend this comparisons with cost sensitive learning (Japkowicz and Stephen, 2002).

The concept of combining classifiers is proposed as a new direction for the improvement of the performance of individual classifiers. The goal of classification result integration algorithms is to generate more certain, precise and accurate

system results. In a following study we will use a method for constructing ensembles for obtaining better classification accuracy in bankruptcy prediction.

Finally, all the techniques employed in the problem of predicting bankruptcy can be straight forwardly used in other financial classification problems such as bond rating or credit scoring.

References

- Aha, D. (1997), *Lazy Learning*, Dordrecht: Kluwer Academic Publishers.
- Altman, E.L. (1993). *Corporate Financial Distress and Bankruptcy*. John Wiley and Sons.
- Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*. 2(2):1-47.
- Cohen, W. (1995), "Fast Effective Rule Induction", *Proceeding of International Conference on Machine Learning 1995*, pp. 115-123.
- 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.
- Domingos, P. and Pazzani, M. (1997), "On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*", Vol. 29, pp. 103-130.
- Doumpos, M., Zopounidis, C. (1999), A Multicriteria Discrimination Method for the Prediction of Financial Distress: The Case of Greece, *Multinational Finance Journal*, 1999, vol. 3, no. 2, pp. 71–101.
- Furnkranz, J. (1999), "Separate-and-Conquer Rule Learning", *Artificial Intelligence Review*, Vol. 13, pp. 3-54.
- Japkowicz N. and Stephen, S. (2002), The Class Imbalance Problem: A Systematic Study *Intelligent Data Analysis*, Volume 6, Number 5.
- Jensen, F. (1996), *An Introduction to Bayesian Networks*, Springer.
- Kotsiantis, S., Zaharakis, I., Tampakas, V., Pintelas, P. (2004), On Constructing a Financial Decision Support System, *Proceedings of International Conference on Enterprise Systems and Accounting 2004*, September 3-4, Thessaloniki, Greece, pp 319-331.
- Mckee, T., Greenstein, M. (2000), Predicting Bankruptcy Using Recursive Partitioning and a Realistically Proportioned Data Set, *Journal of Forecasting*, 19, 219-230.
- Mitchell, T. (1997), *Machine Learning*, McGraw Hill.
- Morris, R., 1998. Early Warning Indicators of Corporate Failure: A Critical Review of Previous Research and Further Empirical Evidence. Ashgate Publishing Company.
- Murthy, S. (1998), "Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey", *Data Mining and Knowledge Discovery*, Vol. 2, pp. 345–389.
- Negakis, C. (1995), "Robustness of Greek business failure prediction models", *International Review of Economics and Business*, Vol. 42 No. 3, pp. 203-215.
- Papoulias, C., and Theodossiou, P. (1992), Analysis and modeling of recent business failures in Greece. *Managerial and Decision Economics* 13(2): 163-169.
- Park, C., Han, I. (2002), A Case-Based Reasoning with the Feature Weights Derived by Analytic Hierarchy Process for Bankruptcy Prediction. *Expert Systems With Applications* 23 (3), 225-264.

- Platt, J. (1999), Using sparseness and analytic QP to speed training of support vector machines. In M. S. Kearns, S. A. Solla, & D. A. Cohn (Eds.), *Advances in neural information processing systems* 11. MA: MIT Press.
- Pompe, P., Feelders, A. (1997), Using Machine Learning, Neural Networks, and Statistics to Predict Corporate Bankruptcy. *Microcomputers in Civil Engineering* 12, 267-276.
- Quinlan, J. R. (1993), *C4.5: Programs for machine learning*, Morgan Kaufmann, San Francisco
- Sarkar, S., and R. S. Sriram, (2001), Bayesian Models for Early Warning of Bank Failures. *Management Science*, 47 (11): 1457–1475.
- Shin, K., Lee, T., Kim, H., (2005), An application of support vector machines in bankruptcy prediction model, *Expert Systems with Applications*, Volume 28, pp. 127-135.
- Sikonja M. and Kononenko I. (1997), An adaptation of Relief for attribute estimation in regression, *Proceedings of the Fourteenth International Conference (ICML'97)*, ed., Dough Fisher, pp. 296-304. Morgan Kaufmann Publishers.
- Thomaidis, N., Gounias, G., Zopounidis, C. (1999). A fuzzy rule based learning method for corporate bankruptcy prediction, *ACAI 99*, Chania, Greece.
- Vranas, A. S. (1992), The significance of financial characteristics in predicting business failure: An analysis in the Greek context. *Foundations of Computing and Decision Sciences* 17(4): 257-275.
- Witten, I. and Frank E. (2000), *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, San Mateo.
- Zhang, G., Hu, M. Y., Patuwo, B. E., Indro, D. C. (1999), Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-Validation Analysis. *European Journal of Operational Research* 116, 16-32.