Forecasting Fraudulent Financial Statements using Data Mining

S. Kotsiantis, E. Koumanakos, D. Tzelepis, and V. Tampakas

***Abstract*—**This paper explores the effectiveness of machine learning techniques in detecting firms that issue fraudulent financial statements (FFS) and deals with the identification of factors associated to FFS. To this end, a number of experiments have been conducted using representative learning algorithms, which were trained using a data set of 164 fraud and non-fraud Greek firms in the recent period 2001-2002. The decision of which particular method to choose is a complicated problem. A good alternative to choosing only one method is to create a hybrid forecasting system incorporating a number of possible solution methods as components (an ensemble of classifiers). For this purpose, we have implemented a hybrid decision support system that combines the representative algorithms using a stacking variant methodology and achieves better performance than any examined simple and ensemble method. To sum up, this study indicates that the investigation of financial information can be used in the identification of FFS and underline the importance of financial ratios.

***Keywords*—**Machine learning, stacking, classifier.

I. INTRODUCTION

ALTHOUGH it is not a new phenomenon, the number of corporate earnings restatements due to aggressive accounting practices, accounting irregularities, or accounting fraud has increased significantly during the past few years, and it has drawn much attention from investors, analysts, and regulators.

In 1998, the Chairman of the Securities and Exchange Commission (SEC), Arthur Levitt argued that extreme earnings management practice masked the underlying performance of the firm and advocated increasing the quality of the reported earnings. In December 1999, the New York Stock Exchange (NYSE) strengthened the rules for audit committee by requiring all listed firms to have an audit committee consisting of at least three independent directors, among whom at least one commit-tee member has accounting or financial management expertise. After many high profile accounting frauds and corporate scandals (Enron, WorldCom, Adelphia etc.) Arthur Levitt's speech seems more like a prophecy. These fraudulent events have been followed by increased governmental intervention and regulation. In 2002, the U.S. congress passed the Sarbanes-Oxley Act to improve the accuracy and reliability of corporate financial reporting and disclosures. Europe also had financial scandals over this same period (with the Parmalat scandal being the most notorious) even if most of which were characteristically different from the US style. In this context, Bollen et. al [5] in an attempt to identify the true causes of Europe's biggest business failures over the past 25 years discovered that high leveraging and management fraud were the only two characteristics common in more than half the cases investi-gated. However, the authors conclude that although accounting issues found to play a role in a number of business failures in their study, it is less significant compared with large US business failures.

Accounting frauds can be classified as either fraudulent financial reporting or misappropriation of assets, or both. Fraudulent financial reporting is commonly known as "cooking the books." The Treadway Commission defined fraudulent financial reporting as the intentional or reckless conduct, whether by act or omission, that results in materially misleading financial statements. In presenting inaccurate financial statements, fraudulent financial reporting will have significant consequences for both the organization and for the public's confidence in the capital markets. Misappropriation of assets is simply using assets and resources for unintended purposes. Such fraud includes thievery, embezzlement, and cash skimming.

Researchers have used various techniques and models to detect accounting fraud in circumstances in which, a priori, is likely to exist. However few studies have tested the predictive ability of different types of models and methods used by means of a common data set. In this study, we carry out an in-depth examination of publicly available data from the financial statements of various firms in order to detect FFS by using alternative supervised machine learning methods. The goal of this research is to identify the financial factors to be used by auditors in assessing the likelihood of FFS.

The detection of fraudulent financial statements, along with the qualification of financial statements, have recently been in the limelight in Greece because of the increase in the number

Manuscript received Feb 19, 2006. Any views or opinions presented are solely those of the authors and do not necessarily represent those of National Bank of Greece Group.The Project is Co-Funded by the European Social Fund & National Resources - EPEAEK II.

S. B. Kotsiantis is with Department of Accounting, Technological Educational Institute of Patras, Greece (phone: +302610997833; fax: +302610997313; e-mail: sotos@ math.upatras.gr).

E. Koumanakos is with National Bank of Greece, Credit Division (e-mail: koumanakos.e@nbg.gr).

D. Tzelepis is with Department of Accounting, Technological Educational Institute of Patras, Greece (e-mail: tzelepis@upatras.gr).

V. Tampakas is with Department of Accounting, Technological Educational Institute of Patras, Greece (e-mail: tampakas@teipat.gr).

of companies listed on the Athens Stock Exchange (and raising capital through public offerings) and the attempts to reduce taxation on profits. In Greece, the public has been consistent in its demand for fraudulent financial statements and qualified opinions as warning signs of business failure. There is an increasing demand for greater transparency, consistency and more information to be incorporated within financial statements [25].

The decision of which particular learning algorithm to choose for the specific problem, is a complicated problem. A good alternative to choosing only one method is to create a hybrid forecasting system incorporating a number of possible solution methods as components (an ensemble of classifiers). For this purpose, we have implemented a hybrid decision support system that combines the representative algorithms using a stacking variant methodology and achieves better performance than any simple method.

The following section attempts a brief literature review. Section 3 describes the data set of our study and the feature selection process. Section 4 presents some elementary Machine Learning definitions. Section 5 presents the presented method and the experimental results for the representative compared algorithms. Finally, section 6 discusses the conclusions and some future research directions.

II. Literature Review

As Watts and Zimmerman [28] argue the financial statement audit is a monitoring mechanism that helps reduce information asymmetry and protect the interests of the principals, specifically, stockholders and potential stockholders, by providing reasonable assurance that management's financial statements are free from material misstatements. However, in real life, detecting management fraud is a difficult task when using normal audit procedures [8] since there is a shortage of knowledge concerning the characteristics of management fraud. Additionally, given its infrequency, most auditors lack the experience necessary to detect it. Last but not least, managers deliberately try to deceive auditors [11].

Green and Choi [13] developed a Neural Network fraud classification model. The model used five ratios and three accounts as input. The results showed that Neural Networks have significant capabilities when used as a fraud detection tool.

Nieschwietz et al. [18] provide a comprehensive review of empirical studies related to external auditors' detection of fraudulent financial reporting while Albrecht et al. [2] review the fraud detection aspects of current auditing standards and the empirical research conducted on fraud detection.

Bell and Carcello [4] developed and tested a logistic regression to estimate the likelihood of fraudulent financial reporting using a sample of 77 fraud and 305 non-fraud engagements, based on the incidence of red flags as explanatory variables. They found that the significant red flags that effectively discriminated between fraud and non fraud engagements were: management lied to the auditor; a weak internal control environment; an unduly aggressive management attitude; undue manage-ment emphasis on meeting earning projections; and significant difficult-to-audit transactions.

Beasley et al. (2000) compare the company governance mechanisms of known fraud cases with "no-fraud" industry benchmarks; they found that companies who exhibited fraud had fewer audit committees, fewer independent audit committees, fewer audit committee meetings, less frequent internal audit support and fewer independent board members. Church et al. (2001) provide further evidence that internal auditors are sensitive to factors that affect the possibility of fraudulent financial re-porting. Specifically, they show that in a situation where operating income is greater than expected, an earnings-based bonus plan is used, and debt covenants are restrictive, internal auditors assigned a higher likelihood of fraud. Ansah et al. [3] investigate the relative influence of the size of audit firms, auditor's position tenure and auditor's year of experience in auditing on the likelihood of detecting fraud in the stock and warehouse cycle. They conclude that such factors are statistically significant predictors of the likelihood of detecting fraud, and increase the likelihood of fraud detection.

For Greek data, Spathis [24] constructed a model to detect falsified financial statements. He employed the statistical method of logistic regression. Two alternative input vectors containing financial ratios were used. The reported accuracy rate exceeded 84%. Kirkos et al [15] investigate the usefulness of Decision Trees, Neural Networks and Bayesian Belief Networks in the identification of fraudulent financial statements. In terms of performance, the Bayesian Belief Network model achieved the best performance managing to correctly classify 90.3% of the validation sample in a 10-fold cross validation procedure. For both studies [24] and [15] a balanced sample of a total of 76 manufacturing firms was used; 38 firms with FFS were matched with 38 with non-FFS (the sample did not include financial companies).

The application of machine learning techniques for financial classification is a fertile research area. Yet, their application for the purpose of management fraud detection has been rather minimal [7], [15]. As a consequence, our main objective for this study is to evaluate the predictive ability of machine learning techniques by conducting a number of experiments using representative learning algorithms, trained in a data set of 164 fraud and non-fraud Greek quoted firms. We also propose a stacking variant method that achieves better classification accuracy.

III. Data Description

Our sample contained data from 164 Greek listed on the Athens Stock Exchange (ASE) manufacturing firms (no financial companies were included). Auditors checked all the firms in the sample. For 41 of these firms, there was published indication or proof of involvement in issuing FFS. The classification of a financial statement as false was based on the following parameters: inclusion in the auditors' re-port of serious doubts as to the accuracy of the accounts, observations by the tax authorities regarding serious taxation intransigencies which significantly altered the company's annual balance sheet and income statement, the application of Greek legislation regarding negative net worth, the inclusion of the company in the Athens Stock Exchange categories of "under observation and "negotiation suspended" for reasons associated with the falsification of the company's financial data and, the existence of court proceedings pending with respect to FFS or serious taxation contraventions

The 41 FFS firms were matched with 123 non-FFS firms. All the variables used in the sample were extracted from formal financial statements, such as balance sheets and income statements. This implies that the usefulness of this study is not restricted by the fact that only Greek company data was used.

The selection of variables to be used as candidates for participation in the input vector was based upon prior research work, linked to the topic of FFS. Such work carried out by [11], [24], [25]. Additional variables were also added in an attempt to catch as many as possible predictors not previously identified. Table I provides a brief description of the financial variables used in the present study.

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 [23]. 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 II. The larger the value of the ReliefF scores is, the more influence of the attribute in the induction.

Thus, the attributes that mostly influence the induction are: RLTC/RCR02, AR/TA01, TL/TA02, AR/TA02, WC/TA02, DC/CA02, NFA/TA02, NDAP02. In general, the identification of the aforementioned variables as crucial factors agrees with the results of previous studies in this field. Specifically, financial leverage ratio is likely to be associated with accounting fraud given that the earnings figures used to determine its value have been shown by a plethora of studies to require subjective judgment and as a matter of fact to be manipulated by management.

With regard to the remaining variables, it seems that the other attributes do not influence the induction at all. For this reason, the previous attributes are not included in the training set of the learning algorithms.

TABLE I
RESEARCH VARIABLES DESCRIPTION

Category	Independent variables	Variable Description
Profitability Variables	*EBIT/TA02*	Earnings before interest and tax/total assets 2002
	RCF/TA02	Results carried forward/total assets 2002
	EBT02/EBIT02	Earnings before tax 2002/Earnings before interest and tax 2002
Leverage Variables	*RLTC/RCR02*	Return on Long -term capital / Return on Capital and Reserves 2002
	TL/TA02	Total liabilities/Total assets 2002
	TA/CR02	Total Assets/Capital and Reserves 2002
	LTD/TCR02	Long term debt/total capital and reserves 2002
	NFA/TA	Net Fixed Assets/Total Assets
Liquidity Variables	*DC/CA02*	Deposits and cash/current assets 2002
	WCL02	Working capital leveraged 2002
	CR02	Current assets to current liabilities 2002
	CR/TL02	Capital and Reserves/total liabilities 2002
	WC/TA 02	Working capital/total assets 2002
Efficiency Variables	*AR/TA 01*	Accounts Receivable/Total Assets 2001
	AR/TA02	Accounts Receivable/Total Assets 2002
	NDAR02	Number of days accounts receivable 2002
	NDAP02	Number of days accounts payable 2002
	CAR/TA	Change Accounts Receivable/Total Assets
	ITURN02	Inventory turnover 2002
	CAR/NS	Change Accounts Receivable/Net Sales
	S/TA02	Sales/total assets 2002
Cash Flow Variables	*GOCF*	Growth of Operational Cash Flow
	CFO02	Cash flows from operations 2002
	CFO01	Cash flows from operations 2001
Financial distress	*Z-SCORE02*	Altman z-score 2002

TABLE II
AVERAGE RELIEFF SCORE OF EACH ATTRIBUTE

Variables	ReliefF score
RLTC/RCR02	0.02603371
AR/TA 01	0.02587121
TL/TA02	0.02577709
AR/TA02	0.02257509
WC/TA 02	0.02118785
DC/CA02	0.01364156
NFA/TA	0.0133596
NDAP02	0.01085013
LTD/TCR02	0.00798901
S/TA02	0.00395956
RCF/TA02	0.00384807
NDAR02	0.00327257
CAR/TA	0.00320415
WCL02	0.00254562
ITURN02	0.00215535
TA/CR02	0.00208717
EBIT/TA02	0.00206301
CFO02	0.00169573
CFO01	0.0009421
CR02	0.00082761
GOCF	0.00073566
CAR/NS	0.00071853
EBT02/EBIT02	0.00049986
Z-SCORE02	0.00047192
CR/TL02	0.00041943

IV. MACHINE LEARNING TECHNIQUES AND FRAUD DETECTION

Supervised machine learning is the exploration for algorithms that reason from externally supplied instances to produce general hypotheses, which will make predictions about future instances. In other words, the goal of supervised learning is to build a concise model of the distribution of the class label in terms of the predictor features. The resulting classifier is then used to assign class labels to the testing instances where the values of the predictor features are known but the value of the class label is unknown.

Decision trees are trees that classify instances by sorting them based on attribute values. Each node in a decision tree represents an attribute in an instance to be classified, and each branch represents a value that the node can take. A recent overview of existing work in decision trees is provided in [17]. 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 [12]. The general goal is to construct the smallest rule-set that is consistent with the training data.

Artificial Neural Networks (ANNs) are another method of inductive learning and they all based on computational models of biological neurons [16]. A multi layer neural network consists of large number of units (neurons) joined together in a pat-tern of connections. 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.

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 [14].

Logistic regression analysis [14] extends the techniques of multiple regression analysis to research situations in which the outcome variable (class) is categorical. The relationship between the classifier and attributes is not a linear function; instead, the logistic regression function is used, which is the logit transformation of pi:

.

$$\operatorname{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} = \ln\left(\frac{\text{Prob}(y_i = 1)}{\text{Prob}(y_i = 0)}\right)$$

The dependent variable (class) in logistic regression is binary, that is, the dependent variable can take the value 1 with a probability of success pi, or the value 0 with probability of failure 1- pi. Comparing these two probabilities, the larger probability indicates the class label value that is more likely to be the actual label.

Instance-based learning algorithms belong in the category of lazy-learning algorithms, as they delay the induction process until classification is performed. One of the most straightforward instance-based learning algorithms is the nearest neighbour algorithm [1]. K-Nearest Neighbour (kNN) assumes that the instances within a data set will generally exist in close proximity with other instances of the similar class.

The SVM technique revolves around the notion of a ‘margin’ that separates two data classes. Maximizing the margin, and thereby creating the largest possible dis-tance between the separating hyperplanes can reduce the upper bound on the expected generalization error [6]. However, 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. The solution is then to map the data into a higher-dimensional space and define a separating hyperplane there.

V. EXPERIMENTAL RESULTS

For the purpose of this study, a representative algorithm for each described learning technique was used. The most commonly used C4.5 algorithm [20] was the representative of the decision trees in our study. RBF algorithm [16] - was the representative of the ANNs. The K2 algorithm [14] was the representative of the Bayesian networks in our study. The 3-NN algorithm that combines robustness to noise and less time for classification than using a larger k for kNN was also used

[1]. Ripper [10] was the representative of the rule-learners. Finally, the Sequential Minimal Optimization (or SMO) algorithm was the representative of the SVMs as one of the fastest methods to train SVMs [19].

All accuracy estimates were obtained by averaging the results from stratified 10-fold cross-validation in our dataset. It must be mentioned that we used the free available source code for our experiments by the book [29]. The results are presented in Table III.

TABLE III
ACCURACY OF SIMPLE MODELS IN OUR DATASET

ALGORITHMS	Accuracy	Fraud	Non-Fraud
K2	74.1	51.2	82.1
C4.5	91.2	85.2	93.3
3NN	79.7	56.1	88.0
RBF	73.4	36.6	86.3
RIPPER	86.8	65.7	94.1
LR	75.3	36.6	88.9
SMO	78.66	48.8	88.6

The K2 algorithm correctly classifies 74.1% of the total sample, 51.2% of the fraud cases and 82.1% of the non-fraud cases. The RBF algorithm manages to correctly classify 73.4% of the total validation sample, 36.6% of the fraud cases and 86.3% of the non-fraud cases. Moreover, C4.5 algorithm succeeds in correctly classifying 85.2% of the fraud cases, 93.3% of the non-fraud cases and 91.2% of the total validation sets. Furthermore, 3NN algorithm succeeds in correctly classifying 56.1% of the fraud cases, 88.0% of the non-fraud cases and 79.7% of the total validation sets. SMO algorithm correctly classifies 78.66% of the total sample, 48.8% of the fraud cases and 88.6% of the non-fraud cases. Ripper algorithm succeeds in correctly classifying 65.7% of the fraud cases, 94.1% of the non-fraud cases and 86.8% of the total validation set. What is more, logistic regression algorithm manages to correctly classify 75.3% of the total validation sample, 36.6% of the fraud cases and 88.9% of the non-fraud cases.

Finally, we combined the previous algorithms using a stacking variant methodology. Its basic idea may be derived as a generalization of voting as follows. Let us consider the voting step as a separate classification problem, whose input is the vector of the responses of the base classifiers. Simple voting uses a predetermined algorithm for this, namely to count the number of predictions for each class in the input and to predict the most frequently predicted class. Stacking replaces this with a trainable classifier. This is possible, since for the training set, we have both the predictions of the base learners and the true class. The matrix containing the predictions of the base learners as predictors and the true class for each training case will be called the meta-data set. The classifier trained on this matrix will be called the meta-classifier or the classifier at the meta-level. While stacking [26] uses all class probabilities for all models, our method uses only the class probabilities associated with the true class. The dimensionality of the meta-data set is reduced by a factor equal to the number of classes, which leads to faster learning. Concerning the choice of the algorithm for learning at the meta-level, we have explored the use of model trees instead of MLR [21] since model trees naturally extend MLR to construct piecewise linear approximations. Model trees have the same structure as decision trees, with one difference: they employ a linear regression function at each leaf node to make a prediction. The most well known model tree inducer - M5´ [27] – is used by our system. In the following, we briefly describe the learning algorithms that are used as base learners.

Subsequently, we compare the proposed stacking methodology (Stacking´) (Table IV) with:

- The methodology of selecting the best classifier according to 10-cross validation (BestCV) [29].
- Grading methodology using the instance based classifier IBk with ten nearest neighbors as the meta level classifier [22]. In grading, the meta-level classifier predicts whether the base-level classifier is to be trusted (i.e., whether its prediction will be correct). The base-level attributes are used also as meta-level attributes, while the meta-level class values are + (correct) and − (incorrect). Only the base-level classifiers that are predicted to be correct are taken and their predictions combined by summing up the probability distributions predicted.
- Simple Voting methodology using the same base classifiers [29].
- Stacking methodology that constructs the meta-data set by adding the entire predicted class probability distribution instead of only the most likely class using MLR as meta-level classifier [26].

TABLE IV
ACCURACY OF ENSEMBLES IN OUR DATASET

ALGORITHMS	Accuracy	Fraud	Non-Fraud
Stacking´	95.1	90.2	96.7
Voting	92.1	80.5	95.9
BestCV	91.2	85.2	93.3
Grading	93.3	85.4	95.9
Stacking	93.9	85.4	96.7

The presented algorithm correctly classifies 95.1% of the total sample, 90.2% of the fraud cases and 96.7% of the non-fraud cases. Voting algorithm correctly classifies 92.1% of the total sample, 80.5% of the fraud cases and 95.9% of the non-fraud cases. BestCV algorithm manages to correctly classify 91.2% of the total validation sample, 85.2% of the fraud cases and 93.3% of the non-fraud cases. Moreover, Grading algorithm succeeds in correctly classifying 85.4% of the fraud cases, 95.9% of the non-fraud cases and 93.3% of the total validation sets. Furthermore, Stacking algorithm succeeds in correctly classifying 85.4% of the fraud cases, 96.7% of the non-fraud cases and 93.9% of the total validation sets. As a conclusion, our approach performs better than selecting the best classifier from the ensemble by cross validation and other examined ensemble methods.

VI. Conclusion

Auditing practices nowadays have to cope with an increasing number of management fraud cases. Supervised machine learning techniques can facilitate auditors in accomplishing the task of management fraud detection. The aim of this study has been to investigate the usefulness and compare the performance of machine learning techniques in detecting fraudulent financial statements by using published financial data.

The results obtained from the experiments agree with prior research results indicating that published financial statement data contains falsification indicators. Furthermore, a relatively small list of financial ratios largely determines the classification results. This knowledge, coupled with machine learning algorithms, can provide models capable of achieving considerable classification accuracies.

In terms of performance, the proposed stacking variant methodology achieves better performance than any examined simple and ensemble method. Tracking progress is a time-consuming job that can be handled automatically by a learning 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.

It must be mentioned that our input vector solely consists of financial ratios. Enriching the input vector with qualitative information, such as previous auditors' qualifications or the composition of the administrative board, could increase the accuracy rate.

References

[1] Aha, D. (1997), Lazy Learning, Dordrecht: Kluwer Academic Publishers.
[2] Albrecht, C.C., Albrecht, W.S. and Dunn, J.G. (2001), "Can auditors detect fraud: a review of the research evidence", Journal of Forensic Accounting, Vol. 2 No. 1, pp. 1-12.
[3] Ansah, S.O., Moyes, G.D., Oyelere, P.B. and Hay, D. (2002), "An empirical analysis of the likelihood of detecting fraud in New Zealand", Managerial Auditing Journal, Vol. 17 No. 4, pp. 192-204.
[4] Bell T. and Carcello J. (2000) 'A decision aid for assessing the likelihood of fraudulent financial reporting', Auditing: A Journal of Practice & Theory, Vol. 9 (1), pp. 169- 178.
[5] Bollen L., Mertens G., Meuwissen R., VanRaak J., and Scelleman C. (2005), "Classification and Analysis of Major European Business Failures". Maastricht Accounting, Auditing and Information Management Research Center (MARC) of University Maastricht and RSM.
[6] Burges, C. (1998). A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery. 2(2):1-47.
[7] Calderon T.G., and Cheh J.J., (2002), 'A roadmap for future neural networks research in auditing and risk assessment', International Journal of Accounting Information Systems, Vol. 3, No. 4, pp. 203-236.
[8] Coderre G. D. (1999) Fraud Detection. Using Data Analysis Techniques to Detect Fraud. Global Audit Publications.
[9] Coffee, J. (2005), "A theory of corporate scandals: Why the USA and Europe differ". Oxford Review of Economic Policy, Vol. 21 (2), pp. 198-211.
[10] Cohen, W. (1995), "Fast Effective Rule Induction", Proceeding of International Confer-ence on Machine Learning, pp. 115-123.
[11] Fanning K. and Cogger K., (1998), 'Neural Network Detection of Management Fraud Using Published Financial Data', International Journal of Intelligent Systems in Account-ing, Finance & Management, Vol. 7, No. 1, pp. 21-24.
[12] Furnkranz, J. (1999), "Separate-and-Conquer Rule Learning", Artificial Intelligence Review, Vol. 13, pp. 3-54.
[13] Green B.P. and Choi J.H., (1997), 'Assessing the risk of management fraud through neuralnetwork technology', Auditing: A Journal of Practice and Theory, Vol. 16(1), pp.14-28.
[14] Jensen, F. (1996), An Introduction to Bayesian Networks, Springer.
[15] Kirkos S., Spathis C., Manolopoulos Y., (2006), Data Mining techniques for the detection of fraudulent financial statements, Expert Systems with Applications..
[16] Mitchell, T. (1997), Machine Learning, McGraw Hill.
[17] Murthy, S. (1998), "Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey", Data Mining and Knowledge Discovery, Vol. 2, pp. 345–389.
[18] Nieschwietz, R.J., Schultz, J.J. Jr and Zimbelman, M.F. (2000), "Empirical research on external auditors' detection of financial statement fraud", Journal of Accounting Litera-ture, Vol. 19, pp. 190-246.
[19] 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 infor-mation processing systems 11. MA: MIT Press.
[20] Quinlan, J. R. (1993), C4.5: Programs for machine learning, Morgan Kaufmann, San Francisco
[21] Seewald, A.K, 2002. How to Make Stacking Better and Faster While Also Taking Care of an Unknown Weakness, in Sammut C., Hoffmann A. (eds.), *Proceedings of the Nineteenth International Conference on Machine Learning (ICML 2002)*, Morgan Kaufmann Publishers, pp.554-561.
[22] Seewald, A. K., Furnkranz, J., 2001. An evaluation of grading classifiers. In Advances in Intelligent Data Analysis: *Proceedings of the Fourth International Symposium (IDA-01)*, pages 221–232, Berlin, Springer.
[23] Sikonja M. and Kononenko I. (1997), An adaptation of Relief for attribute estimation in regression, Proc. of ICML'97, pp. 296-304. Morgan Kaufmann Publishers.
[24] Spathis C., (2002), 'Detecting false financial statements using published data: some evidence from Greece', Managerial Auditing Journal, Vol. 17, No. 4, pp. 179-191.
[25] Spathis C., Doumpos M. and Zopounidis C., (2002), 'Detecting falsified financial state-ments: a comparative study using multicriteria analysis and multivariate statistical tech-niques', The European Accounting Review, Vol.11, No. 3, pp. 509-535.
[26] Ting, K., & Witten, I., 1999. Issues in Stacked Generalization, Artificial Intelligence Research 10, 271-289, Morgan Kaufmann.
[27] Wang, Y., Witten, I., 1997, Induction of model trees for predicting continuous classes, In *Proc. of the Poster Papers of the European Conference on ML*, Prague, 128–137.
[28] Watts, R. L., and J. L. Zimmerman, 1986, Positive Accounting Theory. Prentice-Hall.
[29] Witten I. & Frank E., Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, San Mateo, 2000.

Sotiris Kotsiantis received a diploma in mathematics, a Master and a Ph.D. degree in computer science from the University of Patras, Greece. His research interests are in the field of data mining and machine learning. He has more than 40 publications to his credit in international journals and conferences.

E. Koumanakos (MSc-Brunel University, Phd candidate-University of Patras) is credit underwriter in the National Bank of Greece. His research interests relate to financial reporting and auditing issues. He has more than 20 publications in international journals and conferences.

D. Tzelepis received his diploma in Business Economics from the Aristotle University of Thessaloniki, Greece and his Ph.D. degree in Economics Science from the University of Patras, Greece. His research interests are focused in the field of Financial Accounting and Reporting. He has more than 10 publications in international journals and conferences.

V. Tampakas is a Professor in ATEI of Patras, Greece. His research interests are in the field of text retrieval and accounting information systems. He has numerous publications in international journals and conferences.