

Data Mining techniques for the detection of fraudulent financial statements

Efstathios Kirkos ^{a,1}, Charalambos Spathis ^{b,*}, Yannis Manolopoulos ^{c,2}

^a Department of Accounting, Technological Educational Institution of Thessaloniki, P.O. Box 141, 57400 Thessaloniki, Greece

^b Department of Economics, Division of Business Administration, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

^c Department of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

Abstract

This paper explores the effectiveness of Data Mining (DM) classification techniques in detecting firms that issue fraudulent financial statements (FFS) and deals with the identification of factors associated to FFS. In accomplishing the task of management fraud detection, auditors could be facilitated in their work by using Data Mining techniques. This study investigates the usefulness of Decision Trees, Neural Networks and Bayesian Belief Networks in the identification of fraudulent financial statements. The input vector is composed of ratios derived from financial statements. The three models are compared in terms of their performances.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Fraudulent financial statements; Management fraud; Data Mining; Auditing; Greece

1. Introduction

Auditing nowadays has become an increasingly demanding task and there is much evidence that ‘book cooking’ accounting practices are widely applied. Koskivaara calls the year 2002, ‘the horrible year’, from a book-keeping point of view and claims that manipulation is still ongoing (Koskivaara, 2004). Some estimates state that fraud costs US business more than \$400 billion annually (Wells, 1997). Spathis, Doumpos, and Zopounidis (2002) claim that fraudulent financial statements have become increasingly frequent over the last few years. Management fraud can be defined as the deliberate fraud committed by management that causes damage to investors and creditors through material misleading financial statements. During the audit process, the auditors have to estimate the possibility of management fraud. The AICPA explicitly

acknowledges the auditors’ responsibility for fraud detection (Cullinan & Sutton, 2002). In order to develop his/her expectations, the auditor employs analytical review techniques, which allow for the estimation of account balances without examining relevant individual transactions. Fraser, Hatherly, and Lin (1997) classify analytical review techniques as non-quantitative, simple quantitative and advance quantitative. Advance quantitative techniques include sophisticated methods derived from statistics and artificial intelligence, like Neural Networks and regression analysis.

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 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 (Spathis, Doumpos, & Zopounidis, 2003).

* Corresponding author. Tel.: +30 2310 996688; fax: +30 2310 996452.

E-mail addresses: stkirk@acc.teithe.gr (E. Kirkos), hspathis@econ.auth.gr (C. Spathis), manolopo@csd.auth.gr (Y. Manolopoulos).

¹ Tel.: +30 2310 791209.

² Tel.: +30 2310 991912.

Data Mining (DM) is an iterative process within which progress is defined by discovery, either through automatic or manual methods. DM is most useful in an exploratory analysis scenario in which there are no predetermined notions about what will constitute an “interesting” outcome (Kantardzic, 2002). The application of Data Mining techniques for financial classification is a fertile research area. Many law enforcement and special investigative units, whose mission it is to identify fraudulent activities, have also used Data Mining successfully. However, as opposed to other well-examined fields like bankruptcy prediction or financial distress, research on the application of DM techniques for the purpose of management fraud detection has been rather minimal (Calderon & Cheh, 2002; Koskivaara, 2004; Kirkos & Manolopoulos, 2004).

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 Data Mining classification methods. The goal of this research is to identify the financial factors to be used by auditors in assessing the likelihood of FFS. One main objective is to introduce, apply, and evaluate the use of Data Mining methods in differentiating between fraud and non-fraud observations.

The aim of this study is to contribute to the research related to the detection of management fraud by applying statistical and Artificial Intelligence (AI) Data Mining techniques, which operate over publicly available financial statement data. AI methods have the theoretical advantage that they do not impose arbitrary assumptions on the input variables. However, the reported results of AI methods slightly or occasionally outperform the results of the statistical methods.

In this study, three Data Mining techniques are tested for their applicability in management fraud detection: Decision Trees, Neural Networks and Bayesian Belief Networks. The three methods are compared in terms of their predictive accuracy. The input data consists mainly of financial ratios derived from financial statements, i.e., balance sheets and income statements. The sample contains data from 76 Greek manufacturing companies. Relationships between input variables and classification outcomes are captured by the models and revealed.

The paper proceeds as follows: Section 2 reviews relevant prior research. Section 3 provides an insight into the research methodology used. Section 4 describes the developed models and analyzes the results. Finally, Section 5 presents the concluding remarks.

2. Prior research

In 1997, the Auditing Standards Board issued the *Statement on Auditing Standards (SAS) No. 82: Consideration of Fraud in a Financial Statement Audit*. This Standard requires auditors to assess the risk of fraud during each audit and encourages auditors to consider both the internal control system and management’s attitude toward controls, when making this assessment. Risk factors or “red

flags” that relate to fraudulent financial reporting may be grouped into the following three categories (SAS No. 82): (a) the management’s characteristics and influence over the control environment, (b) industry conditions, and (c) operational characteristics and financial stability. The *International Auditing Practices Committee (IAPC)* of the International Federation of Accountants approved the *International Statement on Auditing (ISA) 240*. This standard respects the auditor’s consideration of the risk that fraud and error may exist, and clarifies the arguments on the inherent limitations of an auditor’s ability to detect error and fraud, particularly management fraud. Moreover, it emphasizes the distinction between management and employee fraud and elaborates on the discussion concerning fraudulent financial reporting.

Detecting management fraud is a difficult task when using normal audit procedures (Porter & Cameron, 1987; Coderre, 1999). First, there is a shortage of knowledge concerning the characteristics of management fraud. Secondly, given its infrequency, most auditors lack the experience necessary to detect it. Finally, managers deliberately try to deceive auditors (Fanning & Cogger, 1998). For such managers, who comprehend the limitations of any audit, standard auditing procedures may prove insufficient. These limitations suggest that there is a need for additional analytical procedures for the effective detection of management fraud. It has also been noted that the increased emphasis on system assessment is at odds with the profession’s position regarding fraud detection, since most material frauds originate at the top levels of the organization, where controls and systems are least prevalent and effective (Cullinan & Sutton, 2002).

Recent studies have attempted to build models that will predict the presence of management fraud. Results from a logit regression analysis of 75 fraud and 75 no-fraud firms have indicated that no-fraud firms have boards with significantly higher percentages of outside members than fraud firms (Beasley, 1996). Hansen, McDonald, Messier, and Bell (1996) use a powerful generalized qualitative response model to predict management fraud based on a set of data developed by an international public accounting firm. The model includes the probit and logit techniques. The results indicate a good predictive capability for both symmetric and asymmetric cost assumptions.

Eining, Jones, and Loebbecke (1997) conducted an experiment to examine if the use of an expert system enhanced the auditors’ performance. They found that auditors using the expert system discriminated better among situations with varying levels of management fraud-risk and made more consistent decisions regarding appropriate audit actions. Green and Choi (1997) 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. A financial statement classified as fraudulent alerts the auditor to conduct further investigation. Fanning and Cogger (1998) used a Neural Network

to develop a fraud detection model. The input vector consisted of financial ratios and qualitative variables. They compared the performance of their model with linear and quadratic discriminant analysis, as well as logistic regression, and claimed that their model is more effective at detecting fraud than standard statistical methods.

Summers and Sweeney (1998) constructed a cascaded logit model to investigate the relationship between insider trading and fraud. They found that, in the presence of fraud, insiders reduce their holdings of company stock through high levels of selling activity. Abbot, Park, and Parker (2000) employed statistical regression analysis to examine if the existence of an independent audit committee mitigates the likelihood of fraud. They found that firms with audit committees, which consist of independent managers who meet at least twice per year, are less likely to be sanctioned for fraudulent or misleading reporting.

Bell and Carcello (2000) developed and tested a logistic regression model that estimates the likelihood of fraudulent financial reporting for an audit client, conditioned on the presence or absence of several fraud-risk factors. The significant risk factors included in the final model were: weak internal control environment, rapid company growth, inadequate or inconsistent relative profitability, management that places undue emphasis on meeting earnings projections, management that lies to the auditors or is overly evasive, ownership status (public vs. private) on the entity, and interaction term between a weak control environment and an aggressive management attitude towards financial reporting.

Spathis (2002) 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%. The results suggest that there is potential in detecting FFS through the analysis of published financial statements. Spathis et al. (2002) used the UTADIS method to develop a falsified financial statement detection model. The method operates on the basis of a non-parametric regression-based framework. They also used the discriminant analysis and logit regression methods as benchmarks. Their results indicate that the UTADIS method performs better than the other statistical methods as regards the training and validation sample. The results also showed that the ratios “total debt/total assets” and “inventory/sales” are explanatory factors associated with FFS.

3. Research methodology

3.1. Data

Our sample contained data from 76 Greek manufacturing firms (no financial companies were included). Auditors checked all the firms in the sample. For 38 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’ report 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 38 FFS firms were matched with 38 non-FFS firms. These firms were characterized as non-FFS based on the absence of any indication or proof concerning the issuing of FFS in the auditors’ reports, in financial and taxation databases and in the Athens Stock Exchange. This of course did not guarantee that the financial statements of these firms were not falsified or that FFS behavior would not be revealed in the future. It only guaranteed that no FFS had been found in an extensive relevant search. 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.

3.2. Variables

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 Spathis (2002); Spathis et al. (2002), Fanning and Cogger (1998); Persons (1995); Stice (1991); Feroz, Park, and Pastena (1991); Loebbecke, Eining, and Willingham (1989) and Kinney and McDaniel (1989) contained suggested indicators of FFS.

Financial distress may be a motivation for management fraud (Fanning & Cogger, 1998; Stice, 1991; Loebbecke et al., 1989; Kinney & McDaniel, 1989). In order to accommodate a ratio concerning financial distress, we employed the well-known Altman’s *Z* score. Altman in his pioneering work developed the ratio *Z* score to estimate financial distress (Altman, 1968). Since then, many researchers in studies relevant to bankruptcy prediction and financial distress have extensively used his *Z* score.

Persons (1995) claims that it is an open question whether a high debt structure is associated with FFS. A high debt structure may increase the likelihood of FFS, since it shifts the risk from equity owners and managers to debt owners. Managers may be manipulating the financial statements due to their need to meet debt covenants. This suggests that higher levels of debt may increase the probability of FFS. We have measured this by using the logarithm of Total Debt (LOGDEBT), the Debt to Equity (DEBTEQ) ratio and the Total Debt to Total Assets (TDTA) ratio.

Another motivation for management fraud is the need for continuing growth. Companies unable to achieve similar results to past performances may engage in fraudulent activities to maintain previous trends (Stice, Albrecht, & Brown, 1991). Companies who are growing rapidly may exceed the monitoring process ability to provide proper supervision (Fanning & Cogger, 1998). As a growth measure we use the Sales Growth (SALGRTH) ratio.

A number of accounts, which permit a subjective estimation, are more difficult to audit and thus are prone to fraudulent falsification. Accounts Receivable, inventory and sales fall into this category. Persons (1995); Stice (1991) and Feroz et al. (1991) claim that management may manipulate Accounts Receivable. The fraudulent activity of recording sales before they are earned may show up as additional accounts receivable (Fanning & Cogger, 1998). We check Accounts Receivable by using the ratio Account Receivable/Sales (RECSAL), the ratio Accounts Receivable/Accounts Receivable for two successive years (RETREND) and the digital variable REC11, which indicates a 10% change.

Many researchers suggest that management may manipulate inventories (Stice, 1991; Persons, 1995; Schilit, 2002). Reporting inventory at lower cost and recording obsolete inventory are some known tactics. We check inventory by using the ratios Inventory/Sales (INVSAL) and Inventory to Total Assets (INVTA). Gross Margin is also prone to manipulation. The company may not match its sales with the corresponding cost of goods sold, thus increasing gross margin, net income and strengthening the balance sheet (Fanning & Cogger, 1998). We test Gross Margin by using the ratio Sales minus Gross Margin (COSAL), the ratio Gross Profit/Total Assets (GPTA), the ratio Gross Margin/Gross Margin for two successive years (GMTREND) and the binary variable GM11, which indicates a 10% increase of the previous year's value.

Spathis (2002), in a logistic regression study on FFS prediction, claims that the ratios Net Profit/Total Assets (NPTA) and Working Capital/Total Assets (WCTA) present significant coefficients. Furthermore, Spathis et al. (2002) suggest that the ratio Net Profit/Sales (NPSAL) has also proven to be significant.

In the present study, some additional financial “red flag” statements are tested in relation to their ability to predict FFS. These ratios are: Logarithm of Total Assets (LTA), Working Capital (WCAP), the ratio of property plant & equipment (net fixed assets) to total assets (NFATA), sales to total assets (SALTA), Current Assets/Current Liabilities (CACL), Net Income/Fixed Assets (NIFA), Cash/Total Assets (CASHTA), Quick Assets/Current Liabilities (QACL), Earnings Before Interest and Taxes (EBIT) and Long Term Debt/Total Assets (LTDTA).

In total, we compiled 27 financial ratios. In an attempt to reduce dimensionality, we ran ANOVA to test whether the differences between the two classes were significant for each variable. If the difference was not significant (high

p -value), the variable was considered non-informative. Table 1 depicts the means, standard deviations, F -values and p -values for each variable.

As can be seen in Table 1, 10 variables presented low p -values ($p \leq 0.05$). These variables were chosen to participate in the input vector, while the remaining variables were discarded. The selected variables were DEBTEQ, SALTA, COSAL, EBIT, WCAP, ZSCORE, TDTA, NPTA, WCTA and GPTA.

3.3. Methods

Identifying fraudulent financial statements can be regarded as a typical classification problem. Classification is a two-step procedure. In the first step, a model is trained by using a training sample. The sample is organized in tuples (rows) and attributes (columns). One of the attributes, the class label attribute, contains values indicating the predefined class to which each tuple belongs. This step is also known as supervised learning. In the second step, the model attempts to classify objects which do not belong to the training sample and form the validation sample.

Data Mining proposes several classification methods derived from the fields of statistics and artificial intelligence. Three methods, which enjoy a good reputation for their classification capabilities, are employed in this research study. These methods are Decision Trees, Neural Networks and Bayesian Belief Networks.

3.3.1. Decision Trees

A Decision Tree (DT) is a tree structure, where each node represents a test on an attribute and each branch represents an outcome of the test. In this way, the tree attempts to divide observations into mutually exclusive subgroups. The goodness of a split is based on the selection of the attribute that best separates the sample. The sample is successively divided into subsets, until either no further splitting can produce statistically significant differences or the subgroups are too small to undergo similar meaningful division. There are several proposed splitting algorithms. In the Automatic Interaction Detection (AID) the most significant t -statistic as per analysis of variance is used. The chi-square AID uses chi-square statistic and the Classification and Regression Trees (CART) use an index of diversity (Koh & Low, 2004). In this study we employed the well-known ID3 algorithm. ID3 uses an entropy-based measure, known as information gain, in order to select the splitting attribute (Han & Camber, 2000).

The successive division of the sample may produce a large tree. Some of the tree's branches may reflect anomalies in the training set, like false values or outliers. For that reason tree pruning is required. Tree pruning involves the removal of splitting nodes in a way that does not significantly affect the model's accuracy rate.

In order to classify a previously unseen object, the attribute values of the object are tested against the splitting nodes of the Decision Tree. According to this test, a path

Table 1
P-values and statistics for input variables

Variables	Mean FFS	SD FFS	Mean non-FFS	SD non-FFS	F	p- value
<i>DEBTEQ</i>	2.706	3.531	1.075	0.937	7.56	0.007
<i>SALTA</i>	0.699	0.416	1.055	0.577	9.53	0.003
NPSAL	−0.459	2.434	0.067	0.159	1.77	0.188
RECSAL	1.755	5.897	0.456	0.349	1.84	0.179
NFATA	0.320	0.193	0.276	0.140	1.31	0.257
RETREND	1.273	0.491	1.733	3.028	0.86	0.358
REC11	0.605	0.495	0.711	0.460	0.92	0.340
GMTREND	2.564	9.014	18.715	108.484	0.84	0.363
GM11	0.395	0.495	0.579	0.500	2.60	0.111
SALGRTH	1.163	0.705	415.616	2554.464	1.00	0.320
INVSAL	0.359	0.656	0.179	0.159	2.70	0.105
INVTA	0.191	0.160	0.178	0.142	0.13	0.722
CASHTA	0.071	0.122	0.096	0.114	0.86	0.356
LTA	8.451	0.996	8.473	1.288	0.01	0.933
LOGDEBT	7.870	1.233	7.512	1.411	1.39	0.243
<i>COSAL</i>	3196.600	3066.188	5889.742	7734.285	3.98	0.050
<i>EBIT</i>	−33.624	1392.930	1093.253	2875.929	4.73	0.033
<i>WCAP</i>	369.171	2166.685	2280.737	4279.685	6.03	0.016
<i>ZSCORE</i>	0.778	0.936	1.990	0.730	39.59	0.000
NIFA	−0.531	1.996	−1.554	11.635	0.29	0.595
<i>TDTA</i>	0.629	0.243	0.437	0.196	14.31	0.000
<i>NPTA</i>	−0.021	0.095	0.074	0.065	26.11	0.000
CACL	2.113	6.095	2.302	2.267	0.03	0.859
<i>WCTA</i>	0.054	0.239	0.253	0.205	15.15	0.000
QACL	1.163	2.245	1.772	1.804	1.70	0.196
<i>GPTA</i>	0.144	0.121	0.274	0.140	18.78	0.000
LTDTA	0.065	0.138	0.041	0.077	0.87	0.353

Italics denote the selected variables.

is traced that will conclude with the object's class prediction. The main advantages of Decision Trees are that they provide a meaningful way of representing acquired knowledge and make it easy to extract IF–THEN classification rules.

3.3.2. Neural networks

Neural Networks (NN) is a mature technology with an established theory and recognized application areas. A NN consists of a number of neurons, i.e., interconnected processing units. Associated with each connection is a numerical value, called “weight”. Each neuron receives signals from connected neurons and the combined input signal is calculated. The total input signal for neuron j is $u_j = \sum w_{ij} * x_i$, where x_i is the input signal from neuron i and w_{ji} is the weight of the connection between neuron i and neuron j . If the combined input signal strength exceeds a threshold, then the input value is transformed by the transfer function of the neuron and finally the neuron fires (Han & Camber, 2000).

The neurons are arranged into layers. A layered network consists of at least an input (first) and an output (last) layer. Between the input and output layer there may exist one or more hidden layers. Different kinds of NNs have a different number of layers. Self-organizing maps (SOM) have only an input and an output layer, whereas a back-propagation NN has additionally one or more hidden layers.

After the network architecture is defined, the network must be trained. In backpropagation networks, a pattern is applied to the input layer and a final output is calculated at the output layer. The output is compared with the desired result and the errors are propagated backwards in the NN by tuning the weights of the connections. This process iterates until an acceptable error rate is reached.

Neural Networks make no assumptions about attributes' independence, are capable of handling noisy or inconsistent data and are a suitable alternative for problems where an algorithmic solution is not applicable. Since the backpropagation NNs has become the most popular one for the prediction and classification of problems (Sohl & Venkatachalam, 1995), we also chose to develop a back-propagation NN in the present study.

3.3.3. Bayesian Belief Networks

Bayesian classification is based on the statistical theorem of Bayes. Bayes theorem provides a calculation for the posterior probability. According to Bayes theorem, if H is a hypothesis—such as the object X belongs to the class C —then the probability that the hypothesis holds is $P(H|X) = (P(X|H) * P(H))/P(X)$.

If an object X belongs to one of i alternative classes, in order to classify the object a Bayesian classifier calculates the probabilities $P(C_i|X)$ for all the possible classes C_i and assigns the object to the class with the maximum probability $P(C_i|X)$.

Naive Bayesian classifiers make the class condition independence assumption, which states that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption simplifies the calculation of $P(C_i|X)$. If this assumption holds true, the Naive Bayesian classifiers have the best accuracy rates compared to all other classifiers. However, in many cases this assumption is not valid, since dependencies can exist between attributes.

Bayesian Belief Networks (BBN) allow for the representation of dependencies among subsets of attributes. A BBN is a directed acyclic graph, where each node represents an attribute and each arrow represents a probabilistic dependence. If an arrow is drawn from node A to node B, then A is parent of B and B is a descendent of A. In a Belief Network each variable is conditional independent of its non-descendents, given its parents (Han & Camber, 2000).

For each node X there exists the Conditional Probability Table, which specifies the conditional probability of each value of X for each possible combination of the values of its parents (conditional distribution $P(X|Parents(X))$). The probability of a tuple (x_1, x_2, \dots, x_n) having n attributes is

$$P(x_1, x_2, \dots, x_n) = \prod P(x_i | Parents(x_i))$$

The network structure can be defined in advance or can be inferred from the data. For classification purposes one of the nodes can be defined as the class node. The network can calculate the probability of each alternative class.

4. Experiments and results analysis

Three alternative models were built, each based on a different method. First, the Decision Tree model was constructed using the Sipina Research Edition software. The model was built with confidence level 0.05. We used the whole sample as a training set. Fig. 1 shows the constructed Decision Tree.

The model was tested against the training sample and managed to correctly classify 73 cases (general performance 96%). More specifically, the Decision Tree correctly classified all the non-fraud cases (100%) and 35 out of the 38 fraud cases (92%).

As can be seen in Fig. 1, the algorithm uses the variable Z score as first splitter. Thirty-five out of the 38 fraud companies present a considerably low Z score value (Z score < 1.49). Since Altman considered a Z score value of 1.81 as a cutoff point to define financial distress for US manufacturing firms (Altman, 2001), we can infer that companies in financial distress included in our sample tend to manipulate their financial statements.

As second level splitters, two variables associated with profitability (NPTA and EBIT) were used. No fraud companies with a high z score present high profitability, whereas fraud companies with a low z score present low profitability. Table 2 depicts the splitting variables in the order they appear in the Decision Tree.

In the second experiment we constructed the Neural Network model. We used the Nuclass 7 Non Linear Networks for Classification software to build a multi layer perceptron feed-forward Network. After testing a number of alternative designs and performing preliminary trainings, we chose a topology with one hidden layer containing five hidden nodes.

Table 2
The splitting variables

Variables
Z-SCORE
NPTA
EBIT
COSAL
DEBTEQ

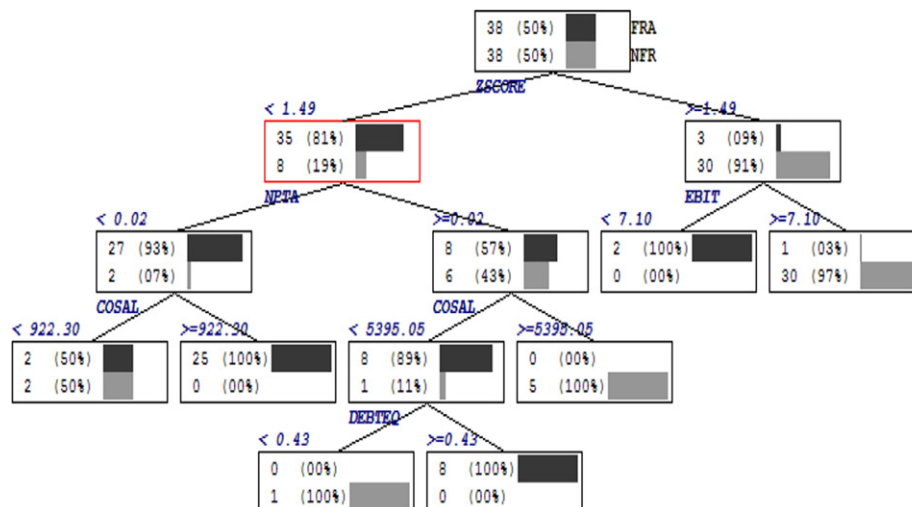


Fig. 1. The Decision Tree.

The selected network was trained by using the whole sample and was tested against the training set. The network succeeded in correctly classifying all the cases, thus achieving a performance of 100%. Unfortunately the software did not provide a transparent interface to the synaptic weights of the connections and thus an estimation of the importance of each input variable was not possible.

In the third experiment we developed a Bayesian Belief Network (BBN). The software we used was the **BN Power Predictor**. This software is capable of learning a classifier from data. The implemented algorithm belongs to the category of conditional independence test-based algorithms and does not require node ordering (Cheng & Greiner, 2001). Due to software limitations we had to perform value discretisation. After testing various discretisation methods (equal depth, equal width), we chose the supervised discretisation method. Unlike other discretisation methods, supervised entropy-based discretisation utilizes class information. This makes it more likely that the intervals defined may help to improve the classification accuracy (Han & Camber, 2000).

In order to train the Belief network we used the whole sample as a training set. After the training the network was tested against the training set. The network correctly classified 72 cases (performance 95%). In particular, it correctly classified 37 fraud cases (97%) and 35 non-fraud cases (92%). Fig. 2 shows the constructed Belief Network.

As can be seen in Fig. 2, the Belief Network seems to accommodate a more generalized aspect regarding financial statement falsification motivations. According to the network, fraud presents strong dependencies from the input variables Z-SCORE, DEBTEQ, NPTA, SALTA and WCTA. Each of these variables expresses a different aspect of a firm’s financial status. Z SCORE refers to financial distress, DEBTEQ to leverage; NPTA refers to profitability, SALTA to sales performance and WCTA to solvency.

Thus the Belief Network seems to record dependencies between financial statement falsification and a considerable

Table 3
The selected variables

Variables
Z-SCORE
NPTA
DEBTEQ
SALTA
WCTA

Table 4
Performance against the training set

Model	Fraud (%)	Non-fraud (%)	Total (%)
ID3	92.1	100.0	96.2
NN	100.0	100.0	100.0
BBN	97.4	92.1	94.7

number of a firm’s financial aspects. Table 3 shows the variables selected to participate in the belief network.

Table 4 depicts the performances of the three models on the training sample. The results indicate that the NN model is quite efficient in discriminating between FFS and non-FFS firms, followed by the BBN and ID3 models.

4.1. The models’ validation

Using the training set in order to estimate a model’s performance might introduce a bias. In many cases the models tend to memorize the sample instead of “learning” (data over fitting). To eliminate such a bias the performance of the models is estimated against previously unseen patterns. There are several approaches in model validation like dividing the sample into training and a separate hold out sample, 10-fold cross validation and hold one out validation. Although the three software packages we used embodied validation capabilities, it was impossible to follow a common validation procedure in terms of methodology and data for all the three software packages. Thus we

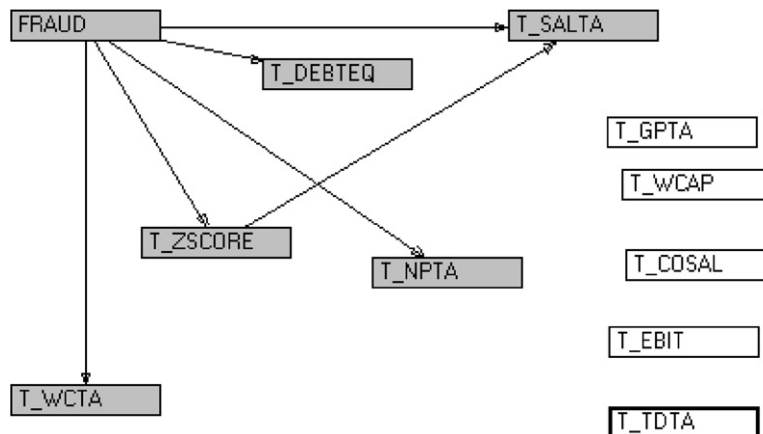


Fig. 2. The selected variables of the Belief Network.

Table 5
10-fold cross validation performance

Model	Fraud (%)	Non-fraud (%)	Total (%)
ID3	75.0	72.5	73.6
NN	82.5	77.5	80.0
BBN	91.7	88.9	90.3

had to manually divide the sample and create the training and validation sets.

We chose to follow a stratified 10-fold cross validation approach. In 10-fold cross validation, the sample is divided in 10-folds. In a stratified approach each fold contains an equal number of fraud and non-fraud cases. For each fold the model is trained by using the remaining nine folds and tested by using the hold out fold. Finally the average performance is calculated. Table 5 summarizes the 10-fold cross validation performances of the three models.

As expected the accuracy rates are lower for the validation set than the accuracy rates for the training set. However the performance of each of the three models is considerably different. The Decision Tree model, which manages to correctly classify 96% of the training set, presents a considerable decrease of its classification accuracy when tested against the validation sample. The model correctly classifies 73.6% of the total sample, 75% of the fraud cases and 72.5% of the non-fraud cases. The Neural Network model, which enjoys an absolute performance of 100% on the training set, manages to correctly classify 80% of the total validation sample, 82.5% of the fraud cases and 77.5% of the non-fraud cases. Finally, the Bayesian Belief Network model which has the lower accuracy for the training set succeeds in correctly classifying 91.7% of the fraud cases, 88.9% of the non-fraud cases and 90.3% of the total validation set.

In a comparative assessment of the models' performance we can conclude that the Bayesian Belief Network outperforms the other two models and achieves outstanding classification accuracy. Neural Networks achieve a satisfactorily high performance. Finally, the Decision Tree's performance is considered rather low.

In assessing the performance of a model, another important consideration is the Type I and Type II error rates. A Type I error is committed when a fraud company is classified as non-fraud. A Type II error is committed when a non-fraud company is classified as fraud. Type I and Type II errors have different costs. Classifying a fraud company as non-fraud may lead to incorrect decisions, which may cause serious economic damage. The misclassification of a non-fraud company may cause additional investigations at the expense of the required time. Although any model aims to reduce both Type I and Type II error rates, a model is supposed to be preferable when it presents a Type I error rate which is lower than its Type II error rate. In our experiments all the models present lower Type I error rates. Neural Network presents the greatest difference between Type I and Type II error rates.

5. Conclusions

Auditing practices nowadays have to cope with an increasing number of management fraud cases. Data Mining techniques, which claim they have advanced classification and prediction capabilities, could 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 three Data Mining techniques in detecting fraudulent financial statements by using published financial data. The methods employed were Decision Trees, Neural Networks and Bayesian Belief Networks.

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 Data Mining algorithms, can provide models capable of achieving considerable classification accuracies.

The present study contributes to auditing and accounting research by examining the suggested variables in order to identify those that can best discriminate cases of FFS. It also recommends certain variables from publicly available information to which auditors should be allocating additional audit time. The use of the proposed methodological framework could be of assistance to auditors, both internal and external, to taxation and other state authorities, individual and institutional, investors, the stock exchange, law firms, economic analysts, credit scoring agencies and to the banking system. For the auditing profession, the results of this study could be beneficial in helping to address its responsibility of detecting FFS.

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. The accuracy rates of the Neural Network model and the Decision Tree model were 80% and 73.6%, respectively. The Type I error rate was lower for all models.

The Bayesian Belief Network revealed dependencies between falsification and the ratios debt to equity, net profit to total assets, sales to total assets, working capital to total assets and Z score. Each of these ratios refers to a different aspect of a firm's financial status, i.e., leverage, profitability, sales performance, solvency and financial distress, respectively. The Decision Tree model primarily associated falsification with financial distress, since it used Z score as a first level splitter.

As usually happens, this study can be used as a stepping stone for further research. An important difference between the experiments is that the BBN model utilized discretized data due to software limitations. Data discretisation eliminates the effects of outliers at the cost of some information loss. Further research is required to address the topic of the impact of data discretisation on the models' performance and the topic of optimal discretisation algorithms. Research

is also needed to examine the circumstances under which DM methods perform better than other techniques.

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. Furthermore a particular study of the industry could reveal specific indicators. We hope that the research presented in this paper will therefore stimulate additional work regarding these important topics.

References

- Abbot, J. L., Park, Y., & Parker, S. (2000). The effects of audit committee activity and independence on corporate fraud. *Managerial Finance*, 26(11), 55–67.
- Altman, E. (1968). Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4), 589–609.
- Altman, E. (2001). *Bankruptcy, credit risk and high yield junk bonds, Part 1 Predicting financial distress of companies: revisiting the Z-score and Zeta® model*. Oxford, UK: Blackwell Publishing.
- American Institute of Certified Public Accountants (AICPA) (1997). *Consideration of fraud in a financial statement audit*. Statement on Auditing Standards No. 82: New York.
- Beasley, M. (1996). An empirical analysis of the relation between board of director composition and financial statement fraud. *The Accounting Review*, 71(4), 443–466.
- Bell, T., & Carcello, J. (2000). A decision aid for assessing the likelihood of fraudulent financial reporting. *Auditing: A Journal of Practice & Theory*, 9(1), 169–178.
- BN Power Predictor, J. Cheng's Bayesian Belief Network Software. <<http://www.cs.ualberta.ca/~jcheng/bnsoft.htm>> Accessed 17.03.2005.
- Calderon, T. G., & Cheh, J. J. (2002). A roadmap for future neural networks research in auditing and risk assessment. *International Journal of Accounting Information Systems*, 3(4), 203–236.
- Cheng, J., & Greiner, R. (2001). Learning Bayesian belief network classifiers: algorithms and systems'. In *Proceedings of the 14th biennial conference of the Canadian society on computational studies of intelligence: advances in artificial intelligence* (pp. 141–151).
- Coderre, G. D. (1999). *Fraud detection. Using data analysis techniques to detect fraud*. Global Audit Publications.
- Cullinan, P. G., & Sutton, G. S. (2002). Defrauding the public interest: a critical examination of reengineered audit processes and the likelihood of detecting fraud. *Critical Perspectives on Accounting*, 13(3), 297–310.
- Eining, M. M., Jones, D. R., & Loebbecke, J. K. (1997). Reliance on decision aids: an examination of auditors' assessment of management fraud. *Auditing: A Journal of Practice and Theory*, 16(2), 1–19.
- Fanning, K., & Cogger, K. (1998). Neural network detection of management fraud using published financial data. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 7(1), 21–24.
- Feroz, E., Park, K., & Pastena, V. (1991). The financial and market effects of the SECs accounting and auditing enforcement releases. *Journal of Accounting Research*, 29(Supplement), 107–142.
- Fraser, I. A. M., Hatherly, D. J., & Lin, K. Z. (1997). An empirical investigation of the use of analytical review by external auditors. *The British Accounting Review*, 29(1), 35–47.
- Green, B. P., & Choi, J. H. (1997). Assessing the risk of management fraud through neural-network technology. *Auditing: A Journal of Practice and Theory*, 16(1), 14–28.
- Han, J., & Camber, M. (2000). *Data mining concepts and techniques*. San Diego, USA: Morgan Kaufman.
- Hansen, J. V., McDonald, J. B., Messier, W. F., & Bell, T. B. (1996). A generalized qualitative—response model and the analysis of management fraud. *Management Science*, 42(7), 1022–1032.
- International Auditing Practices Committee (IAPC) (2001). *The auditor's responsibility to detect fraud and error in financial statements*. International Statement on Auditing (ISA) 240.
- Kantardzic, M. (2002). *Data mining: concepts, models, methods, and algorithms*. Wiley—IEEE Press.
- Kinney, W., & McDaniel, L. (1989). Characteristics of firms correcting previously reported quarterly earnings. *Journal of Accounting and Economics*, 11(1), 71–93.
- Kirkos, S., & Manolopoulos, Y. (2004). Data mining in finance and accounting: a review of current research trends. In *Proceedings of the 1st international conference on enterprise systems and accounting (ICESAcc)*, Thessaloniki, Greece (pp. 63–78).
- Koh, H. C., & Low, C. K. (2004). Going concern prediction using data mining techniques. *Managerial Auditing Journal*, 19(3), 462–476.
- Koskivaara, E. (2004). Artificial neural networks in analytical review procedures. *Managerial Auditing Journal*, 19(2), 191–223.
- Loebbecke, J., Eining, M., & Willingham, J. (1989). Auditor's experience with material irregularities: frequency, nature and detectability. *Auditing: A Journal of Practice & Theory*, 9, 1–28.
- Nuclass 7 Non Linear Networks for Classification, Neural Decision Lab, University of Texas at Arlington. <<http://www-ee.uta.edu/eeweb/IP/Software/Software.htm>> Accessed 27.03.2005.
- Persons, O. (1995). Using financial statement data to identify factors associated with fraudulent financial reporting. *Journal of Applied Business Research*, 11(3), 38–46.
- Porter, B., & Cameron, A. (1987). Company fraud—what price the auditor? *Accountant's Journal*(December), 44–47.
- Schilit, H. (2002). *Financial Shenanigans: How to detect accounting Gimmicks and fraud in financial reports*. New York, USA: McGraw-Hill.
- Sipina Research Edition, Data Mining, Université Lumière Lyon 2. <<http://eric.univ-lyon2.fr/~ricco/sipina.html>> Accessed 4.04.2005.
- Sohl, J. E., & Venkatachalam, A. R. (1995). A neural network approach to forecasting model selection. *Information & Management*, 29(6), 297–303.
- Spathis, C. (2002). Detecting false financial statements using published data: some evidence from Greece. *Managerial Auditing Journal*, 17(4), 179–191.
- Spathis, C., Doumpos, M., & Zopounidis, C. (2002). Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques. *The European Accounting Review*, 11(3), 509–535.
- Spathis, C., Doumpos, M., & Zopounidis, C. (2003). Using client performance measures to identify pre-engagement factors associated with qualified audit reports in Greece. *The International Journal of Accounting*, 38(3), 267–284.
- Stice, J. (1991). Using financial and market information to identify pre-engagement market factors associated with lawsuits against auditors. *The Accounting Review*, 66(3), 516–533.
- Stice, J., Albrecht, S., & Brown, L. (1991). Lessons to be learned—ZZZZBEST Regina and Lincoln savings. *The CPA Journal*(April), 52–53.
- Summers, S. L., & Sweeney, J. T. (1998). Fraudulent misstated financial statements and insider trading: an empirical analysis. *The Accounting Review*, 73(1), 131–146.
- Wells, J. T. (1997). *Occupational Fraud and Abuse*. Austin, TX: Obsidian Publishing.