Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 32 (2007) 995–1003
Expert Systems with Applications
www.elsevier.com/locate/eswa

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.

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 bookkeeping 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.

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2006.02.016

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