Decision Support Systems 50 (2011) 491–500

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Detection of financial statement fraud and feature selection using data mining techniques

P. Ravisankar [a], V. Ravi [a,*], G. Raghava Rao [a], I. Bose [b]

[a] *Institute for Development and Research in Banking Technology, Castle Hills Road #1, Masab Tank, Hyderabad 500 057, AP, India*
[b] *School of Business, The University of Hong Kong, Pokfulam Road, Hong Kong*

ARTICLE INFO

Article history:
Received 20 November 2009
Received in revised form 14 June 2010
Accepted 3 November 2010
Available online 12 November 2010

Keywords:
Data mining
Financial fraud detection
Feature selection
t-statistic
Neural networks
SVM
GP

ABSTRACT

Recently, high profile cases of financial statement fraud have been dominating the news. This paper uses data mining techniques such as Multilayer Feed Forward Neural Network (MLFF), Support Vector Machines (SVM), Genetic Programming (GP), Group Method of Data Handling (GMDH), Logistic Regression (LR), and Probabilistic Neural Network (PNN) to identify companies that resort to financial statement fraud. Each of these techniques is tested on a dataset involving 202 Chinese companies and compared with and without feature selection. PNN outperformed all the techniques without feature selection, and GP and PNN outperformed others with feature selection and with marginally equal accuracies.

1. Introduction

Financial fraud is a serious problem worldwide and more so in fast growing countries like China. Traditionally, auditors are responsible for detecting financial statement fraud. With the appearance of an increasing number of companies that resort to these unfair practices, auditors have become overburdened with the task of detection of fraud. Hence, various techniques of data mining are being used to lessen the workload of the auditors. Enron and Worldcom are the two major scandals involving corporate accounting fraud, which arose from the disclosure of misdeeds conducted by trusted executives of large public corporations. Enron Corporation [17] was an American energy company based in Houston, Texas. Before its bankruptcy in late 2001, Enron was one of the world's leading electricity, natural gas, pulp and paper, and communications companies, with revenues amounting to nearly $101 billion in 2000. Long Distance Discount Services, Inc. (LDDS) began its operations in Hattiesburg, Mississippi in 1983. The company's name was changed to LDDS WorldCom [18] in 1995, and later it became WorldCom. On July 21, 2002, WorldCom filed for Chapter 11 bankruptcy protection in the largest such filing in US history at that time.

Financial statements are a company's basic documents to reflect its financial status [3]. A careful reading of the financial statements can indicate whether the company is running smoothly or is in crisis. If the company is in crisis, financial statements can indicate if the most critical thing faced by the company is cash or profit or something else. All the listed companies are required to publish their financial statements every year and every quarter. The stockholders can form a good idea about the companies' financial future through the financial statements, and can decide whether the companies' stocks are worth investing. The bank also needs the companies' financial statements in order to decide whether to grant loans to them. In a nutshell, the financial statements are the mirrors of the companies' financial status. Financial statements are records of financial flows of a business. Generally, they include balance sheets, income statements, cash flow statements, statements of retained earnings, and some other statements. A detailed description of the items listed in the various financial statements is given below:

• Balance sheet

A balance sheet is a statement of the book value of an organization at a particular date, usually at the end of the fiscal year. A balance sheet has three parts: assets, liabilities, and shareholders' equity. The

* Corresponding author. Tel.: +91 40 23534981x2042; fax: +91 40 23535157.
E-mail addresses: ravisankar_hcu@yahoo.co.in (P. Ravisankar), rav_padma@yahoo.com (V. Ravi), graghavarao.cse@gmail.com (G. Raghava Rao), bose@business.hku.hk (I. Bose).

0167-9236/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2010.11.006

difference between the assets and the liabilities is known as the 'net assets' or the 'net worth' of the company.

• Income statement

Income statements, also called Profit and Loss Statement for companies indicate how net revenue (money received from the sale of products and services before expenses are subtracted, also known as the 'top line') is transformed into net income (the result after all revenues and expenses have been accounted for, also known as the 'bottom line'). The purpose of the income statement is to show managers and investors whether the company made or lost money during the period under consideration.

• Cash flow statement

A cash flow statement is a financial statement that shows incoming and outgoing funds during a particular period. The statement shows how changes in balance sheet and income accounts affect cash and cash equivalents. As an analytical tool the statement of cash flows is useful in determining the short-term viability of a company, particularly its ability to pay bills.

• Statement of retained earnings

The statement of retained earnings, also known as 'statement of owners' equity' and 'statement of net assets' for non-profit organizations, explains the changes in company's retained earnings over the reporting period. It breaks down changes affecting the account, such as profits or losses from operations, dividends paid, and any other items charged or credited to retained earnings. Next, we will describe the key characteristics of financial fraud that can be observed through the financial ratios calculated on the basis of the financial statements published by companies.

1.1. Financial ratios

Financial ratios are a valuable and easy way to interpret the numbers found in financial statements. They can help to answer critical questions such as whether the business is carrying excess debt or inventory, whether customers are paying according to terms, whether the operating expenses are too high, and whether the company assets are being used properly to generate income.

• Liquidity

Liquidity measures a company's capacity to pay its liabilities in short term. There are two ratios for evaluating liquidity. They are:

1) Current ratio = Total current assets / Total current liabilities
2) Quick ratio = (Cash + Accounts receivable + Any other quick assets) / Current liabilities

The higher the ratios the stronger is the company's ability to pay its liabilities as they become due, and the lower is the risk of default.

• Safety

Safety indicates a company's vulnerability to risk of debt. There are three ratios for evaluating liquidity. They are:

1) Debt to equity = Total liabilities / Net worth
2) EBIT/Interest = Earnings before interest and taxes / Interest charges
3) Cash flow to current maturity of long-term debt = (Net profit + Non-cash expenses) / Current portion of long-term debt

• Profitability

Profitability ratios measure the company's ability to generate a return on its resources. There are four ratios to evaluate a company's profitability. They include:

1) Gross profit margin = Gross profit / Total sales
2) Net profit margin = Net profit / Total sales
3) Return on assets = Net profit before taxes / Total assets
4) Return on equity = Net profit before taxes / Net worth

• Efficiency

Efficiency evaluates how well the company manages its assets. There are four ratios to evaluate the efficiency of asset management:

1) Accounts receivable turnover = Total net sales / Accounts receivable
2) Accounts payable turnover = Cost of goods sold / Accounts payable
3) Inventory turnover = Cost of goods sold / Inventory
4) Sales to total assets = Total sales / Total assets

Financial statement fraud may be perpetrated to increase stock prices or to get loans from banks. It may be done to distribute lesser dividends to shareholders. Another probable reason may be to avoid payment of taxes. Nowadays an increasing number of companies are making use of fraudulent financial statements in order to cover up their true financial status and make selfish gains at the expense of stockholders. The fraud triangle is also known as Cressey's Triangle, or Cressey's Fraud Triangle. The fraud triangle seeks to explain what must be present for fraud to occur. The fraud triangle describes the probability of financial reporting fraud which depends on three factors: incentives/pressures, opportunities, and attitudes/rationalization of financial statement fraud [37,38]. The fraud triangle is depicted in Fig. 1, and it is discussed below.

When financial stability or profitability is threatened by economic, industry, or entity operating conditions, or excessive pressure exists for management to meet debt requirements, or personal net worth is materially threatened, the management will face the incentives or pressures to resort to fraudulent practice. Pressure can come in the form of peer pressure, living a lavish lifestyle, a drug addiction, and many other aspects that can influence someone to seek gains via financial fraud. When there are significant accounting estimates that are difficult to verify, or there is oversight over financial reporting, or high turnover or ineffective accounting internal audit, there are opportunities for fraud. For instance, a cashier can steal money out of the cash register because it is there. If the cashier is required to drop all cash into an underground safe for which he does not know the combination, opportunity will not exist. When inappropriate or inefficient communication and support of the entity's values is evident, or a history of violation of laws is known, or management has a practice of making overly aggressive or unrealistic forecasts, then there are risks of fraudulent reporting due to attitudes/rationalization. Rationalization is a grey area in the fraud triangle. Opportunities and incentives exist or they don't. Rationalization depends on the individuals and the circumstances they are facing [19,37]. Understanding the fraud triangle is essential to evaluating financial fraud. When someone is able to grasp the basic concept of the fraud triangle, they are able to better understand financial frauds, how they occur, why they occur, and what to do to stop them.

1.2. Variables related to financial statement fraud

Based on expert's knowledge, intuition, and previous research, it is important to identify some key financial items that are relevant for detection of financial statement fraud. These are listed below:

• Z-score: The Z-score is developed by Altman [2]. It is a formula for measurement of the financial health of a company and works as a tool to predict bankruptcy. It is used to detect financial statement

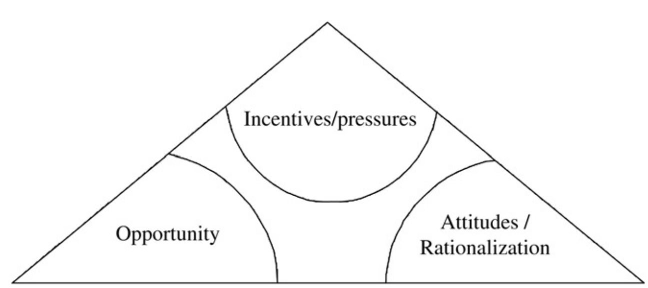

Fig. 1. Components of the fraud triangle.