



A Comparative Study on Fraud Detection in Financial Statement using Data Mining Technique

Mrs. R.Meenatkshi¹, Mrs. K.Sivaranjani²

¹ Research Scholar, Department of Computer Science, Bishop Heber College Tiruchirappalli, Tamilnadu, India

² Asst. Professor, Department of Information Technology, Bishop Heber College, Tiruchirappalli, Tamilnadu, India

Abstract— Financial losses due to financial statement frauds (FSF) are increasing day by day in the world. The industry recognizes the problem and is just now starting to act. Although prevention is the best way to reduce frauds, fraudsters are adaptive and will usually find ways to circumvent such measures. Several data mining algorithms have been developed that allow one to extract relevant knowledge from a large amount of data like fraudulent financial statements to detect FSF. A careful reading of the financial statements can indicate whether the company is running smoothly or is in crisis. 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. In a nutshell, the financial statements are the mirrors of a company's financial status. It comprises the analysis of various classification techniques of financial fraud detection. This paper compares classification techniques and produce the result based on accuracy level.

Keywords— Data mining, Fraud Detection, Financial Fraud, Financial Statements-Means clustering, Multi-Level Feed Forward Network, Genetic Programming

I. INTRODUCTION

Every day, news of financial statement fraud is adversely affecting the economy worldwide. Considering the influence of the loss incurred due to fraud, effective measures and methods should be employed for prevention and detection of financial statement fraud. Data mining methods could possibly assist auditors in prevention and detection of fraud because data mining can use past cases of fraud to build models to identify and detect the risk of fraud and can design new techniques for preventing fraudulent financial reporting. Cost of financial statement fraud is very high both in terms of finance as well as the goodwill of the organization and related country. In order to curb the chances of fraud and to detect the fraudulent financial reporting, number of researchers had used various techniques from the field of statics, artificial intelligence and data mining.

Financial statement frauds (FSF) have received considerable attention from the public, the financial community and regulatory bodies because of several high profile frauds reported at large corporations such as Enron, Lucent, and WorldCom and Satam computers over the last few years. Falsifying financial statements primarily consist of elements manipulating by overstating assets, profit, or understating liabilities. Detecting management fraud using normal audit procedures is a difficult task. First, there is a shortage of knowledge concerning the characteristics of management fraud. Second, most auditors lack the experience necessary to detect it. Finally, financial managers and accountants are deliberately trying to deceive the auditors. For such managers, who understand the limitations of an audit, standard auditing procedures may be insufficient. These limitations suggest the need for additional analytical procedures for the effective detection of false financial statements. Statistics and data mining methods have been applied successfully to detect activities such as money laundering, e-commerce credit card fraud, telecommunications fraud, insurance fraud, and computer intrusion etc. However, FSF is complicated and detecting them is difficult. People tend to question about how to do it and how effective they are. The main objective this paper is to provide a comprehensive review on financial fraud detection (FFD) process. Selected data-mining-based methods that have been used in FFD were examined.

II. RELATED WORK

Data mining has been applied in many aspects of financial analysis. Few areas where data mining techniques have already being used include: bankruptcy prediction, credit card approval, loan decision, money-laundering detection, stock analysis, etc. However, research related to the use of data mining for detection of financial statement fraud is limited. The main objective of this research is to predict the occurrence of financial statement fraud in financial as accurately as possible using intelligent techniques. There has been a limited use of data mining techniques for detection of financial statement fraud.

[1] Ravisankar applied six data mining techniques namely Multilayer Feed Forward Neural Network (MLFF)→Sensitivity (50.6%), Specificity (65.6%), Support Vector Machines (SVM)→ Sensitivity (45.6%), Specificity (65.6%), Genetic Programming (GP), Group Method of Data Handling (GMDH), Logistic Regression (LR), and Probabilistic Neural Network (PNN) to identify financial that resort to financial statement fraud on a data set obtained from 202 Chinese financial. They found Probabilistic neural network as the best techniques without feature selection. Multilayer Feed Forward Neural Network and PNN outperformed others with feature selection and with marginally equal accuracies Sensitivity (65.6%), Specificity (74.6%).

[2] Neural Network based support systems was proposed by Koskivaara. He demonstrated that the main application areas of were detection of material errors, and management fraud. He also investigated the impact of various preprocessing models on the forecast capability of NN when auditing financial accounts Sensitivity (55.6%), Specificity (75.6%) .

[5] Sohl and Venkatachalam used back-propagation NN for the prediction of financial statement fraud Sensitivity (65.6%), Specificity (81.6%).

[8] Kirkos used three Data Mining classification methods namely Decision Trees, Neural Networks and Bayesian Belief Networks Sensitivity (79.9%), Specificity (85.58%).

[9] Zhou & Kapoor in applied four data mining techniques namely regression, decision trees, neural network and Bayesian networks in order to examine the effectiveness and limitations of these techniques in detection of financial statement fraud. They explore a self adaptive framework based on a response surface model with domain knowledge to detect financial statement fraud Sensitivity (45.6%), Specificity (69.6%).

In [10] effectiveness of CART on identification and detection of financial statement fraud was examined by Belinna et al Sensitivity (78.5%), Specificity (85.6%) and found CART as a very effective technique in distinguishing fraudulent financial statement from non-fraudulent. Juszczak et al.[11] apply many different classification techniques in a supervised two-class setting and a semi-supervised one-class setting in order to compare the performances of these techniques and settings Sensitivity (85.6%), Specificity (80.6%).

Further, Zhou & Kapoor [12] in applied four data mining techniques namely regression, decision trees, neural network and Bayesian networks in order to examine the effectiveness and limitations of these techniques in detection of financial statement fraud. They explore a self – adaptive framework based on a response surface model with domain knowledge to detect financial statement fraud Sensitivity (63.6%), Specificity (84.6%).

Recently, Johan Perols [14] compares the performance of six popular statistical and machine learning models in detecting financial statement fraud. The results show, somewhat surprisingly, that logistic regression and support vector machines perform well relative to an artificial neural network in detection and identification of financial statement fraud Sensitivity (45.6%), Specificity (65.6%). The review of the existing literature reveals that the research conducted till date is solely in the field of detection and identification of financial statement fraud and a very little or no work has been done in the field of prevention of fraudulent financial reporting.

III. PERFORMANCE ANALYSIS

3.1. Evolution of financial statement

Financial statement fraud in particular has cast rapidly increasing adverse impact not only on individual investors but the overall stability of global economies. Although there are minor variations in its definition, a financial statement fraud is defined by the Association of Certified Fraud Examiners as —The intentional, deliberate, misstatement or omission of material facts, or accounting data which is misleading and, when considered with all the information made available, would cause the reader to change or alter his or her judgment or decision. Another motivation for

management fraud is the need for continuing growth. Financial unable to achieve similar results to past performances may engage in fraudulent activities to maintain previous trends. Financial who are growing rapidly may exceed the monitoring process ability to provide proper supervision. As a growth measure here the Sales Growth (SALGRTH) ratio is used, 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.

3.2 .Financial Statement Fraud

- 1) Manipulation of financial records.
- 2) Intentional omission of events, transactions, accounts, or other significant information from which financial statements.
- 3) Misapplication of accounting principles, policies, and procedures used to measure, recognize, report, and disclose business transactions. This software is a tool for the auditor in detection of fraudulent financial statements. Traditionally, auditors are responsible for detecting financial statement fraud. With the appearance of an increasing number of financial that resort to these unfair practices, auditors have become overburdened with the task of detection of fraud. Various techniques of data mining are being used to lessen the workload of the auditors.

Despite the increased of time and effort that has been spent to detect the same, the number of detected frauds and the detection rate have largely decreased. When the executives who are involved in financial fraud are well aware of the fraud detection techniques and software, which are usually public information and are easy to obtain, they are likely to adapt the methods in which they

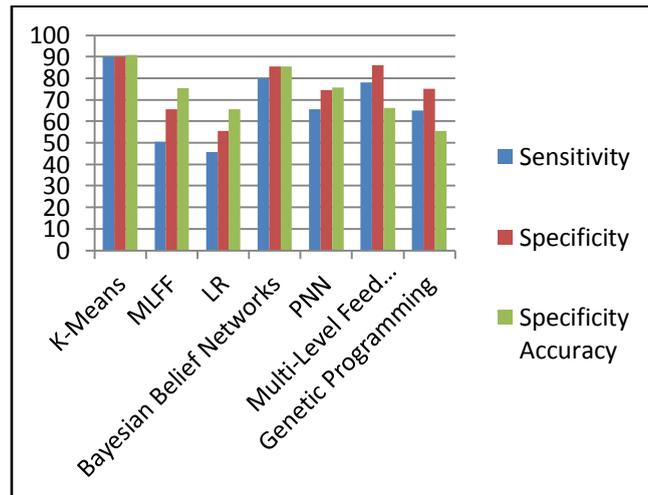
commit fraud and make it difficult to detect the same, especially by existing techniques. There exists an urgent need for new methods that is not only efficient but effective to catch up with these probable newly emerged or adaptive financial shenanigans. This paper provides an overview of existing financial shenanigans and their trend, and new framework to detect evolutionary financial statement fraud is suggested.

The final step of the framework is used for measuring the performance and judging the efficiency of data mining methods. Performance of association rules generated in this study has been measured with the help of accuracy, time taken, and memory space (Table 2). The algorithm for MLFF, GP, GMDH, PNN, Bayesian Belief Networks, K-means, Multi-Level Feed Forward Network, Genetic Programming and is given below. The algorithms are namely, MLFF, GP, GMDH, PNN, Bayesian Belief Networks, K-means, Multi-Level Feed Forward Network, and Genetic Programming. According K-means to the accuracy level will be calculated.

Table 1 : General analysis of sensitivity, specificity, accuracy Fraud using the above 7 algorithms.

Algorithms	Sensitivity	Specificity	Accuracy
K-Means	89.9%	90.1%	90.01%
LP	50.6%	65.6%	75.5%
GMDH	45.6%	55.6%	65.5%
Bayesian Belief Networks	79.9%	85.58%	85.60%
PNN	65.6%	74.6%	75.6%
Multi-Level Feed Forward Network	78%	86%	66.16%
Genetic Programming	65%	75.25%	55.63%

Chart1: General analysis of sensitivity, specificity, accuracy Fraud using the above 7 algorithms.



IV. CONCLUSION

Data mining framework for prevention and detection of financial statement fraud in this study. The framework used in this research follow the conventional flow of data mining. These informative variables are being used for implementing association rule mining for prevention and three predictive mining techniques namely K-means, Multi-Level Feed Forward Network, Genetic programming for detection of financial statement fraud-means produces best accuracy(98.05) specificity as compared with other two methods. These techniques will detect the fraud in case of failure of prevention mechanism. Hence, the framework used in this research is able to prevent fraudulent financial reporting and detect it if management of the organization is capable of perpetrating financial statement fraud despite the presence of anti fraud environment.

REFERENCES

[1] Alhadj, R , Agyemang, M., Barker, K. (2006). A comprehensive survey of numeric and symbolic outlier mining techniques, *Intelligent Data Analysis* 10 (6) (2014) 521–538.

[2] Erol, M. 2008. The expectations from auditing against corruptions (errors and tricks) in the enterprises. Suleyman Demirel University the Journal of Faculty of Economics and Administrative Sciences. Aug 2010, pp 409-424.

[3] Eick, S.G. & Fyock, D.E. (2011). Visualizing corporate data, *AT&T Technical Journal* 75 (1) (2012) 74–86.

[4] FBI, Federal Bureau of Investigation, Financial Crimes Report to the Public Fiscal Year, Department of Justice, United States, 2014, http://www.fbi.gov/publications/financial/fcs_report2007/financial_crime_2014.htm.

[5] Fawcett, T., & Provost, F. (2010). Adaptive fraud detection, *Data Mining and Knowledge Discovery* 1 (3) (2012) 291–316.

[6] Han, J., & Kamber, M. (2014). *Data Mining: Concepts and Techniques*, Second edition, organ Kaufmann Publishers, 2006, pp. 285–464.

[7] E. Koskivaara, Artificial neural networks in auditing: state of the art, *The ICFAI Journal of Audit Practice*. International Auditing Practices Committee (IAPC) (2013). The auditor’s responsibility to detect fraud and error in financial statements. *International Statement on Auditing (ISA) 240*. Nov 1998, pp 1734-1778.

[8] Mitra, S., Mitra, P. & Pal, S.K., (2012). Data mining in soft computing framework: a survey, *IEEE Transactions on Neural Networks* 13 (1) (2014) 3–14.

[9] Persons O.S. (2015). Using Financial Statement Data to Identify Factors Associated with Fraudulent Financial Reporting .*Journal of Applied Business Research*,2011,11(3):38-46.

- [10] P. Ravisankar, V. Ravi, G. Raghava Rao, I. Bose, Decision Support Systems: Detection of financial statement fraud and feature selection using data mining techniques. Aug 2014,pp 309-324.
- [11] Spathis, C. T. (2012). Detecting false financial statements using published data: some evidence from Greece, *Managerial Auditing Journal* 17 (4) (2013) 179–191
- [12] Cerda, L., Sánchez, D., Vila, M.A., & Serrano, J.M. (2015). Association rules applied to credit card fraud detection, *Expert Systems with Applications* 36 (2) (2014) 3630–3640.
- [13] J.E. Sohl, A.R. Venkatachalam, A neural network approach to forecasting model Selection , *Information & Management* Jan 2014,pp 297–303.
- [14] Baesens, B., Derrig, R.A., & Dedene, G. Viaene, S., (2014). A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection, *The Journal of Risk and Insurance* 69 (3) (2013) 373–421.
- [15] Milne.P, Takeuchi J.Yamanishi, K. Williams, (2014). On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms, *Data Mining and Knowledge Discovery* 8 (3) (2014) 275–300.