IJCTA, 9(27), 2016, pp. 407-413 ISSN: 0974-5572
© International Science Press

Fraud Detection in Financial Statement using Data Mining Technique and Performance Analysis

R. Meenatkshi* and Sivaranjani**

ABSTRACT

Financial losses due to financial statement frauds are increasing day by day in the world. Financial fraud is an issue with distant reaching consequences in the investment industry, government, corporate sectors, and for ordinary consumers the business recognizes the difficulty and is just now starting to act. Although prevention is the best way to decrease frauds, fraudsters are adaptive and will regularly find ways to circumvent such measures. Detecting fraud is important once preclusion device has failed. Several data mining algorithms have been developed that allow one to extract applicable knowledge from a huge amount of data like fraudulent financial statements to detect FSF. Financial statements are account of economic flows of a business. Generally, they include balance sheets, income statements, cash flow statements, statements of retained income, and some other statements. In a nutshell, the financial statements are the mirrors of a company's financial status. This paper presents implementation of two data mining techniques namely K-Means Clustering Algorithm and Genetic Programming. Performance of both the techniques is analyzed and results are presented comprehensively.

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

1. INTRODUCTION

Financial statement frauds have conventional considerable awareness from the public, the financial people and regulatory body because of a number of high profile frauds reported at huge corporation such as Enron, Lucent, and WorldCom and computers over the last few years. Falsifying economic statements mainly consist of elements manipulating by overstate assets, profit, or understating liability. detecting organization fraud using usual audit procedures is a difficult task. First, there is a shortage of knowledge concerning the characteristics of management fraud. Second, most auditor lack the knowledge necessary to detect it. Finally, financial managers and accountants are purposely trying to deceive the auditors. For such managers, who understand the restrictions of an audit, standard auditing procedures may be unsatisfactory. These limitations suggest the need for extra analytical procedures for the effective revealing of false financial statement. Information and data mining methods have been apply successfully to detect activities such as cash laundering, e-commerce credit card fraud, tele-communications fraud, insurance fraud, and computer interruption etc. FSF is complicated and detect them is difficult. People tend to inquiry about how to do it and how efficient they are. The main purpose this study is to provide a comprehensive review on financial fraud detection (FFD) process. Select data-mining-based method that have been use in FFD were examined.

* Research Scholar, Department of Computer Science, Bishop Heber College, Tiruchirappalli, Tamilnadu, India, *Email: meenuvisva@gmail.com.*

** Asst. Professor, Department of information technology, Bishop Heber College, Tiruchirappalli, Tamilnadu, India, *Email: sivaranjanik2000@yahoo.in.*

2. REVIEW OF LITERATURE

2.1. Financial Fraud

Financial fraud can be generally defined as an intentional act of deception involving financial transactions for principle of personal grow Fraud is a offense, and is also a civil law violation. Many fraud cases involve complicated financial business conducted by 'white collar criminals' such as business professionals with specialized knowledge and criminal intent.

2.2. Evolution of Financial Statement

Financial statement fraud within particular has transmit rapidly increasing adverse impact not only on entity financier but the overall constancy of global economies. Although there are minor variation in its meaning, a financial statement fraud is defined by the organization of Certified Fraud Examiners as ¯The on purpose, deliberate, misstatement or omission of material facts, or accounting data which is misleading and, when careful with all the information made available, would cause the reader to revise or alter his or her judgment or decision.

One more motivation for organization fraud is the need for continuing growth. Financial unable to achieve similar results to past performance may engage in fraudulent activities to maintain previous trends. Financial who are growing rapidly may exceed the monitor process capability to provide proper supervision. As a growth measure here the Sales Growth (SALGRTH) ratio is used, a number of accounts, which allow a one-sided estimation, are more difficult to audit and thus are prone to false misrepresentation. Financial records Receivable, inventory and sales fall into this category.

2.3. Financial Statement Fraud

- Manipulation of financial report.
- Intentional exclusion of events, transactions, accounts, or other important information from which monetary statements are prepared.
- Misapplication of accounting principles, procedures used to measure and policies, recognize, report, and disclose business transactions.

This Application is a tool for the auditor in exposure of falsified financial statements. Traditionally, auditors are accountable for detecting financial statement fraud. With the appearance of an increasing number of financial that option to these unmerited practices, auditors have become overburdened with the mission of detection of fraud. Various technique of data mining are being used to reduce the workload of the auditors.

Even though the improved of time and effort that has been spent to detect the same, the number of detected frauds and the finding rate have mostly decreased. When the executives who are involved in financial fraud are well attentive of the fraud detection techniques and application, which are usually public information and are easy to obtain, they are likely to adjust the method in which they commit fraud and make it difficult to detect the same, especially by obtainable technique. There exist an urgent call for new methods that is not only efficient but effective to catch up with these possible newly emerge or adaptive financial shenanigans. This paper provides an overview of existing financial mischief and their trend, and new structure to detect evolutionary financial statement fraud is suggested.

3. ENHANCEMENT PROCESS

3.1. K-Means Clustering Algorithm

K- Means is one of the simplest unverified learning algorithms that explain the well known clustering problem. The procedure follows a easy and easy way to categorize a given figures set through a certain number of clusters

(assume k clusters) fixed a priori. The major idea is to define k centroids, one for every cluster. These centroids should be placed in a wiliness way because of different location causes different result. So, the better option is to place them as much as probable far away from each other. The after that step is to take each point belonging to a given data set and associate it to the adjoining centroid. When no point is pending, the first pace is completed and an early group age is done. At this tip it is needed to recalculate _k‘new centroids as barycentre of the clusters resultant from the previous step. After these k new centroids are obtain, a new required has to be done between the same data set points and the adjacent new centroid. A loop has been generate. As a end result of this loop it may be noticed that the k centroids change their position step by step until no more change are done. In other words centroids do not move any more. Finally, this algorithm aim at minimizing an objective function, in this case a square error function. The objective function is

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j} - c_j \right\|^2$$

Where $\left\| x_i^{(j} - c_j \right\|^2$ is a chosen distance compute?

Between a data point $x_i^{(j)}$ and the cluster centre C_j is an pointer of the distance of the *n* data point from their respective cluster centers.

3.2. Multi-Layer Feed forward Network

Multi Layer Feed neural networks, trained with a back-propagation learning algorithm, are the most popular neural networks. They are applied to a wide multiplicity of chemistry related troubles. A MLFF neural network consists of neurons that are ordered into layers. The primary layer is called the input layer, the last layer is called the output layer, and the layers between are secreted layers For the formal explanation of the neurons mapping function r is used, which assign for each neuron I a subset T(i) subset of V which consists of all ancestors of the given neuron. A subset T(i) subset of V than consists of all predecessor of the given neuron i. Each neuron in a particular layer is connected with all neurons in the after that layer. The relation between the i[th] and j[th] neuron is characterized by the weight coefficient wij and the i[th] neuron by the threshold coefficient ²i. The weight coefficient reflects the degree of importance of the given correlation in the neural network. The output value (activity) of the *i*[th] neuron x_i is determined by the following equations.

$$\xi_i = \vartheta_i + \sum \omega_{ij} x_j \, j \in r_i$$

$$x_i = f(\xi_i$$

4. EXPERIMENT AND RESULTS

The ultimate step of the structure is used for measuring the performance and judging the efficacy 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 K-means, Multi-Level Feed Forward Network, Genetic Programming and is given below. The 3 data mining methods used for detection of economic statement fraud are compared on the basis of two method K-means produces best accuracy, time taken second.

4.1. Comparison of Various Dataset financial record set sample data

4.2. Comparison of Various Dataset using Algorithms.

Performance matrix indicating the accuracy, time taken of the three methods used in this study is given in Table 2.

Table 1
Sample Dataset Financial

S no	*Financial Items / Ratios*	*Financial Amount*	*Ratio*
1.	Debt	1.345	0.028
2.	Inventory / Primary business income	3.031	0.001
3	Inventory/Total assets	17.468	0
4	Net profit/Total assets	3.035	0.001
5	Accounts receivable/Primary business income	6.099	0.018
6	Primary business income/Total assets	3.038	0.001
7	Primary business income/Fixed	3.055	0.001
8	Cash/Total assets	2.918	0.001
9	Inventory/Current liabilities	6.744	0.001
10	Total debt/Total assets	2.851	0.001
11	Long term debt/Total capital and reserves	4.266	0.014
12	Deposits and cash/Current assets	2.932	0.001
13	Capitals and reserves/Total debt	2.213	0.003
14	Gross profit/Primary business profit	3.847	0.008
15	Accounts Receivable / Sales	1.702	0.021
16	Working Capital / Total Assets	2.906	0.001
17	Sales / Total Assets	12.818	0.003
18	Net income / Fixed Assets	3.038	0.001
19	Quick assets / Current Liabilities	1.839	0.05
20	Revenue /Total Assets	12.818	0.003

Figure 1: Fraud Detection using K-Means & MLP in financial statement

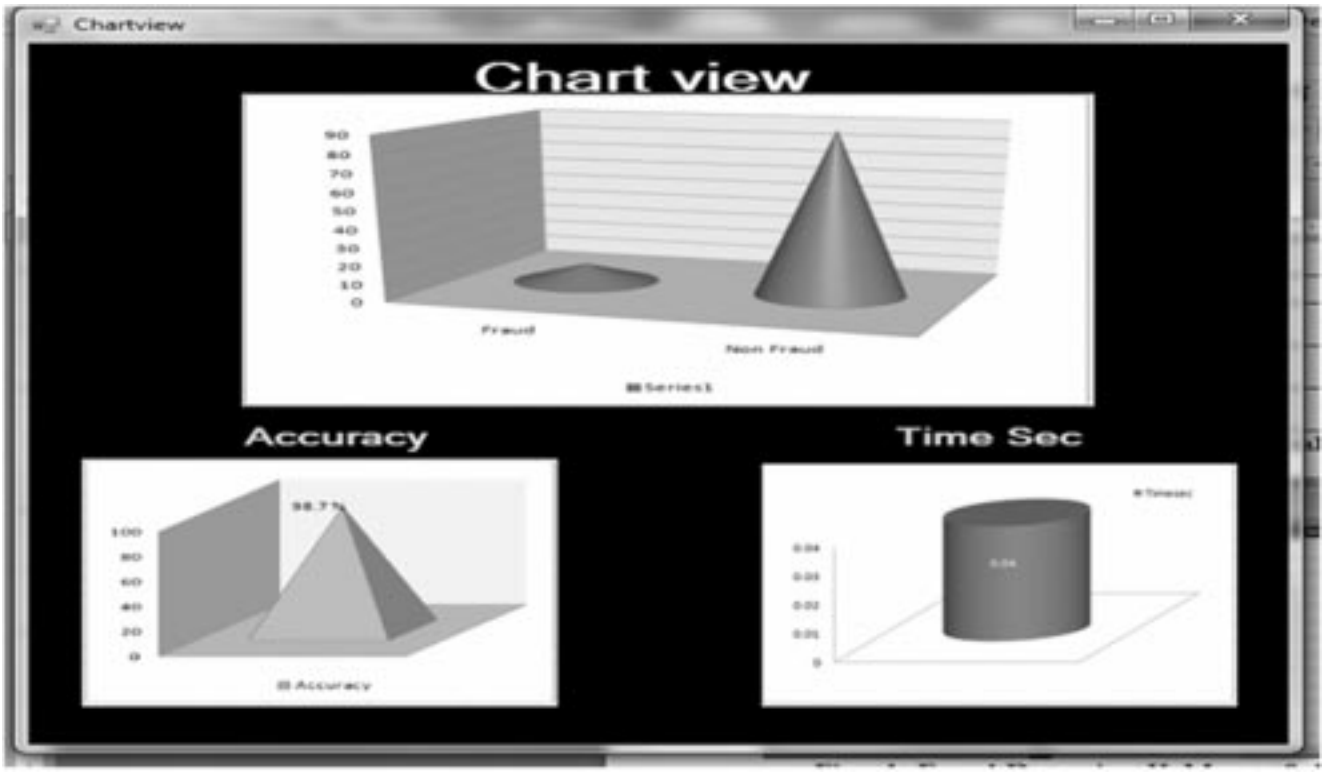

Figure 2: Chart view of deducting fraud in financial statement using K-Means & MLP we get this Accuracy and Timesec

Table 2
Accuracy and time taken (in sec) in k-means& MLP

Algorithms	*Accuracy (%)*	*Time Taken (in secs.)*
K-Means &MLP	98.7%	0.04

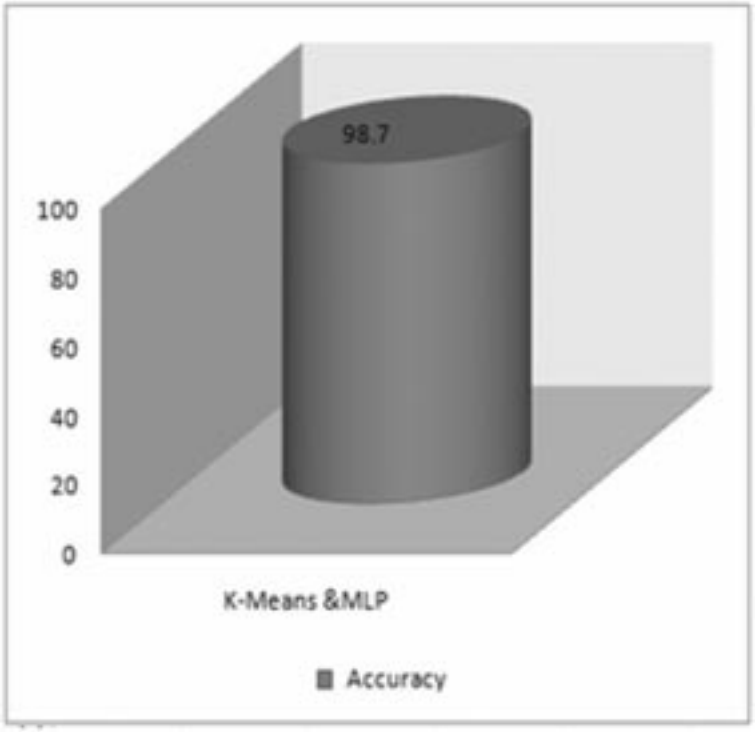

Figure 3: Accuracy

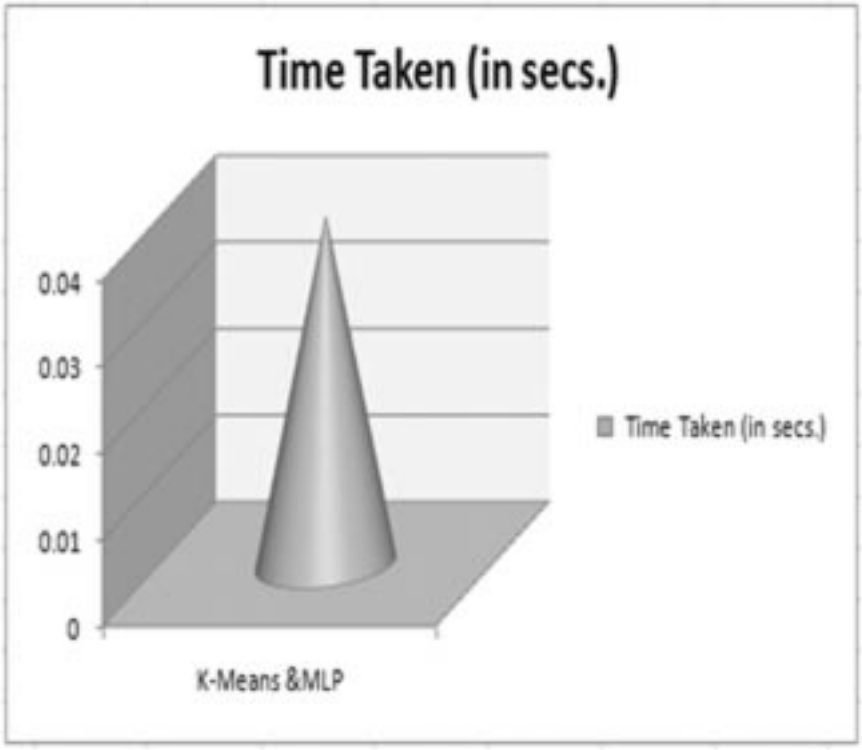

Figure 4: Time Taken Sec

5. CONCLUSION

Data mining structure 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(98.7%), time taken second(0.04) and specificity as compared with other two methods. These techniques will detect the fraud in case of failure of prevention mechanism. Hence, the frame used in this study is able to prevent fraudulent financial reporting and detect it if management of the association is capable of perpetrate financial statement fraud despite the presence of anti fraud environment.

REFERENCES

[1] Agyemang, M., Barker, K., & Alhajj, R. (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., Pal, S.K., & Mitra, P. (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] Sánchez, D., Vila, M.A., Cerda, L., & 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] Viaene, S., Derrig, R.A., Baesens, B., & Dedene, G. (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] Yamanishi, K., Takeuchi, J., Williams, G., & Milne, P. (2014). On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms, Data Mining and Knowledge Discovery 8 (3) (2014) 275–300.