

A Comparative Study to Detect Fraud Financial Statement using Data Mining and Machine Learning Algorithms

Ankita Sharma¹, Mayank Patel², Manish Tiwari³

¹M.Tech. Student(IV sem) , Dept. of Computer Science, GITS, Udaipur, Rajasthan ,India

²Professor, Dept. of Computer Science, GITS, Udaipur, Rajasthan, India

³Professor, Dept. of Computer Science, GITS, Udaipur, Rajasthan, India

Abstract — Financial frauds are increasing day by day within the world. The enterprise acknowledges the problem and is just now beginning to act. Many data mining algorithms are developed that permit one to extract relevant data from a large quantity of information like false financial statements. In recent years, machine learning has developed and received major attention in the predictive analytics in audit analysis. A careful reading of the financial statements will indicate whether or not the corporate is running smoothly or is in crisis. Financial statements are records of economic flows of a business. Generally, they include balance sheets, income statements, cash flow statements, statements of retained earnings, and a few different statements. During a shell, the financial statements are the mirrors of a company's financial status. This paper contains the analysis of various classification techniques to detect frauds, can sometimes find ways in which to avoid such measures and compares classification techniques and produce the result based on accuracy level.

Index Terms — Data mining, Fraud Detection, Financial Fraud, Financial Statements- Means clustering, Machine Learning.

I. INTRODUCTION

Financial statement frauds (FSF) have received wide attention from the general public, the financial community and restrictive bodies due to many high profile frauds reportable at massive companies like Enron, Lucent, and WorldCom and Satyam computers over the previous few years. Finding financial statements primarily contains elements manipulating by overstating assets, profit, or understating liabilities. Detecting management fraud using traditional audit procedures could be a tough task. First, there's a shortage of knowledge regarding the characteristics of management fraud. Second, most auditors lack the expertise necessary to find it. Finally, financial managers and accountants are deliberately trying to deceive the auditors for such managers, who perceive the restrictions of an audit, standard auditing procedures could also be limited. These limitations recommend the requirement for extra analytical procedures for the effective detection of false financial statements. Statistics and data mining strategies are applied with success to find activities like money laundering, e-commerce master card fraud, telecommunications fraud, insurance fraud, and laptop intrusion etc. However, FSF is difficult and detecting them is tough. Individuals tend to question regarding the way to do it and the way effective they're. The main objective this paper is to provide a comprehensive review on financial fraud detection (FFD) method. Selected data-mining- based strategies that are utilized in FFD were examined.

II. RELATED WORK

Data mining has been applied in several aspects of financial analysis. Few areas wherever data processing techniques have already getting used include: bankruptcy prediction, master card approval, loan decision, money-laundering detection, stock analysis, etc. However, analysis associated with the utilization of data mining for detection of financial statement fraud is restricted. The most objective of this analysis is to predict the occurrence of financial statement fraud in financial as accurately as possible using intelligent techniques. There has been a restricted use of data mining techniques for detection of financial statement fraud.

[1] Ravisankar applied six data mining techniques specifically 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 methodology of data Handling (GMDH), logistic Regression (LR), and Probabilistic Neural Network (PNN) to spot financial that resort to financial statement fraud on an information set obtained from 202 Chinese financial. They found Probabilistic neural network because the best techniques without feature choice. Multilayer Feed Forward Neural Network and PNN outperformed others with feature choice and with marginally equal accuracies Sensitivity (65.6%), Specificity (74.6%).

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

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

[4] Kirkos used 3 data mining classification strategies specifically decision Trees, Neural Networks and bayesian Belief Networks Sensitivity (79.9%), Specificity (85.58%).

[5] Zhou & Kapoor in applied four data processing techniques specifically regression, call trees, neural network and theorem networks so as to look at the effectiveness and limitations of those techniques in detection of financial statement fraud. They explore a self adjustive framework supported a response surface model with domain data to find plan fraud Sensitivity (45.6%), Specificity (69.6%).

[6] 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 really effective technique in identifying fallacious financial statement from non-fraudulent.

[7] Juszczak et al. apply many alternative classification techniques during a supervised two- class setting and a semi-supervised one-class setting so as to check the performances of those techniques and settings Sensitivity (85.6%), Specificity (80.6%).

[8] Further, zhou & Kapoor in applied four data mining techniques specifically regression, decision trees, neural network and bayesian networks so as to look at the effectiveness and limitations of those techniques in detection of financial statement fraud. They explore a self – adaptive framework supported a response surface model with domain data to find financial statement fraud Sensitivity (63.6%), Specificity (84.6%).

[9] Johan Perols compares the performance of six widespread statistical and machine learning models in detection plan 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 analysis conducted until date is alone within the field of detection and identification of financial statement fraud and a really little or no work has bee n tired the sector of prevention of fraudulent financial reporting.

[10] Recently, Arpit Tiwari & Nishata Hooda, taken Data of 777 different firms are collected from six distinct sectors. They used various machine learning algorithmic characteristics, which will be capable of predicting whether any fraud has been committed in any firm or not. Ten classification models like SVM, Random Forest, J48, Bayes net, etc. are compared in terms of their accuracy measure, FP rate, TP rate, F-Measure, error rate, Mathew's Correlation Coefficient (MCC), and Area Under the Curve (AUC).

III. PERFORMANCE ANALYSIS

3.1. Evolution of financial statement

Financial statement fraud mostly impact not only on individual investors or organization however generally effect worldwide economies. A financial statement fraud is describe 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 of any business or individual. Financial unable to realize similar results to past performances could interact in fallacious activities to take care of previous trends. Financial who are growing quickly could exceed the observation method ability to provide correct supervising. As a growth live here the Sales Growth ratio is employed, 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 financial trem to calculate financial difference form actual.

3.2 Finance statement fraud

1) Manipulation of financial records.

2) Intentional omission of events, transactions, accounts, or other significant information from which financial statements.

Misapplication of accounting principles, policies, and procedures used to measure, recognize, report, and disclose business transactions. This software system could be a tool for the auditor in detection of deceitful money statements. Traditionally, auditors are responsible for detecting financial statement fraud. With the appearance of an increasing variety of economic that resort to those unfair practices, auditors became burdened with the task of detection of fraud. Various techniques of data mining are getting used to reduce the employment of the auditors. This techniques makes auditing easy to generate perfect audit report .Despite the raised of time and effort that has been spent to find a similar, the number of detected frauds and the detection rate have largely decreased. When the executives who are concerned in financial fraud are cognizant of the fraud detection techniques and software system, that are usually public info and are simple to get, they are seemingly to adapt the strategies in which they commit fraud and build it troublesome to observe identical, particularly by existing techniques. There exists an urgent need for brand new strategies that's not only economical but effective to catch up with these probable new emerged or adaptative money challenges. This paper provides an summary of existing financial challenges and their trend, and new framework to observe evolutionary financial statement fraud is usually recommended.

Table1 Comparison of Algorithms in terms of Sensitivity, Accuracy, Specificity

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%
RANDOM FOREST	93%	92%	94.78%
DECISION TREE	59.35%	62.15%	93.68%
GENETIC PROGRAMMING	65%	75.25%	55.63%

The final step of the framework is employed for measure the performance and judgment the efficiency of data mining strategies. Performance of association rules generated during this study has been measured with the assistance 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 particularly, 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.

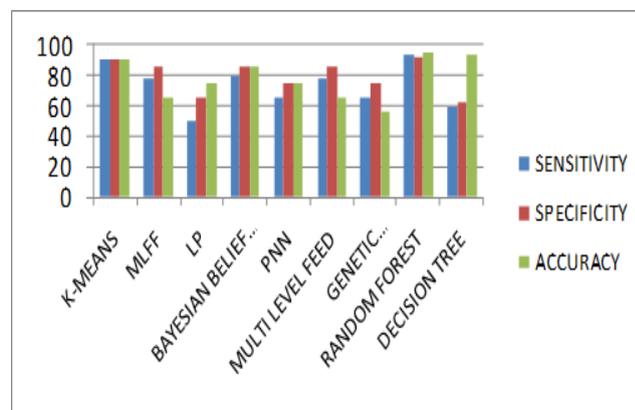


Chart1: General analysis of sensitivity, specificity, accuracy of the above algorithms.

IV. CONCLUSION

Data mining and machine learning framework for prevention and detection of financial statement fraud during this study. The framework employed in this analysis follow the traditional flow of data mining. Various techniques and algorithms are employed for detection of fraud audit work in firm. Used techniques particularly K-means, Multi-Level Feed Forward Network, Genetic programming, machine learning algorithms that is random forest , decision tree for detection of financial statement fraud-means produces best accuracy (94.78) specificity as compared with different strategies. These techniques can find the fraud in case of failure of prevention mechanism. Hence, the framework employed in this analysis is ready to prevent fallacious financial reporting and detect it if management of the organization is capable of finding financial statement fraud despite the presence of anti fraud atmosphere.

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 ICAI 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. P. Ravisankar, V. Ravi, G. Raghava Rao, Bose, *Decision Support Systems: Detection of financial statement fraud and feature selection using data mining techniques*. Aug 2014, pp 309- 324.
- 9) Spathis, C. T. (2012). Detecting false financial statements using published data: some evidence from Greece, *Managerial Auditing Journal* 17 (4) (2013) 179–191
- 10) 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.
- 11) J.E. Sohl, A.R. Venkatachalam, A neural network approach to forecasting model Selection , *Information & Management* Jan 2014, pp 297–303.
- 12) 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.
- 13) 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.
- 14) Arpit Tiwari and Nishta Hooda (2018) *Machine Learning Framework for Audit Fraud Data Prediction*, Volume 7, Issue VI, JUNE/2018