



Application of Data Mining Techniques for Financial Accounting Fraud Detection Scheme

¹S Sowjanya Chintalapati,
M.Tech IIInd Yr, Dept of CSE,
NOVA College, Vijayawada, India

²G.Jyotsna,
Asst Professor, Dept of MBA,
RVR College of Engg. & Tech Hyderabad, India

Abstract: *Data mining techniques are providing great aid in financial accounting fraud detection, since dealing with the large data volumes and complexities of financial data are big challenges for forensic accounting. The implementation of data mining techniques for fraud detection follows the traditional information flow of data mining, which begins with feature selection followed by representation, data collection and management, pre - processing, data mining, post-processing, and performance evaluation. Data mining methods have the capability of detecting fraud because these techniques can use past cases of fraud to build models, which identify and detect the risk of fraud. Financial statement fraud, one of the financial frauds, has reached the epidemic proportion globally. Collapses of high profile companies have left a dirty smear on the effectiveness of corporate governance, quality of financial reports, and credibility of audit functions. Financial statement fraud has become a critical issue in the businesses around the world. The aim of this contribution is to show some data mining techniques for fraud detection and prevention with applications in credit card and telecommunications, within a business of mining the data to achieve higher cost savings, and also in the interests of determining potential legal evidence.*

Keywords: *Data Mining; Auditing; Fraud Detection, Financial Fraud, Fraud triangle.*

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. The requirement of detecting, defining and reporting financial accounting fraud has increased [1]. 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 Oxford English Dictionary [2] defines fraud as "wrongful or criminal deception intended to result in financial or personal gain". In academic literature fraud is defined as leading to the abuse of a profit organization's system without necessarily leading to direct legal consequences [3]. Although the literature is missing a universally accepted definition of financial fraud, researcher has defined it as "a deliberate act that is contrary to law, rule, or policy with intent to obtain unauthorized financial benefit" [4] and "intentional misstatements or omission of amount by deceiving users of financial statement, especially investors and creditors"

[5]. the accounting fraud is executed by making falsified financial accounting statements where the numbers are manipulated by overstating assets, spurious entries related to sales and profit, misappropriation in taxes, or understating liabilities, debts, expenses or losses [1]. The accounting fraud is also defined by accounting professionals as "deliberate and improper manipulation of the recording of data in financial statements in order to achieve an operating profit of the company and appear better than it actually is" [6].

Economically, financial fraud is becoming an increasingly serious problem and effective detecting accounting fraud has always been an important but complex task for accounting professionals [7]. The internal auditing of financial matters in the companies has become an increasingly demanding activity and there are many evidence that "book cooking" accounting practices are world-wide applied for doing financial frauds [8]. The detection of accounting fraud using traditional internal audit procedures is a difficult or sometimes an impossible task [9]. First, the auditors usually lack the required knowledge concerning the characteristics of accounting fraud. Second, as the fraudulent manipulation of accounting data is so infrequent, most of the auditors lack the experience and expertise needed to detect and prevent frauds. Finally, the other concern people of finance department like Chief Financial Officer (CFO), financial managers and accountants are intentionally trying to deceive the internal or external auditors [10]. While knowing the limitations of an audit, finance and accounting managers have concluded that traditional and standard auditing procedures are insufficient to detect frauds. These limitations of financial auditing suggest the need for additional automatic data analysis procedures and tools for the effective detection of falsified financial statements. It is necessary to take into account the cost of the fraud detection and the cost of fraudulent behavior, because stopping a fraud of few dollars can require a very expensive system. This is possible by introducing a decision layer on top of the system in order to decide

the action taking into account factors like the amount of transaction and the risk associated to user doing the transaction. The development of new detections methods is more difficult due to the severe limitation on privacy and on exchange of ideas. Moreover, data sets are not available and results are often not disclosed to the public. The planning audit strategies is a posteriori fraud detection problem with prevention purpose of analyzing historical audit data and constructing models of planning effectively future audits. An application is fiscal and insurance domain, where audits are intended to detect tax evasion and fraudulent claims. A case study is presented by Bonchi (1999) which illustrates how techniques based on classification can be used to support the task of planning audit strategies.

The fraud detection methods in online auction (Shah, 2002) are based on statistical methods and association analysis in order to detect shilling that occurs when the seller tries to hike up the prices in auction by placing buy bids under distinct aliases or through associates. Apart fraud, the detection efforts may be further motivated by the need to understand the behavior of customers to enable provision of matching services and to improve operations.

2. What Is A Fraud?

Fraud is an intentional act meant to induce another person to part with something of value, or to surrender a legal right. It is a deliberate misrepresentation or concealment of information in order to deceive or mislead. Fraud can range from minor employee theft and unproductive behavior to misappropriation of assets and fraudulent financial reporting. In different situational contexts, fraud can take somewhat different forms. For example, bribery, embezzlement, securities fraud, health care fraud, money-laundering scams, insurance fraud, software piracy, internet fraud, telemarketing fraud, mortgage foreclosure scams, and identity theft -- these all have their own special characteristics. There are at least as many types of fraud as there are types of people who commit it. But in each instance, fraud involves deception. Someone knowingly lies in order to obtain an unlawful benefit, or an unfair advantage.

Some examples of fraud include:

- any dishonest or fraudulent act;
- Forgery or alteration of a check, bank draft, or financial document;
- Misappropriation of assets;
- Deliberate impropriety in the handling or reporting of money or financial transactions.
- Wrongfully using influence in a business transaction to receive a benefit (such as bribery, kickbacks, and bid-rigging);
- Profiteering as a result of insider information;
- Disclosing insider information to another person in order for them to secure unlawful gain.

Abuse is behavior that is deficient or improper when compared with behavior that a prudent person would consider reasonable and necessary business practice given the facts and circumstances. Instances of abuse are not fraud or illegal acts, but they are harmful, and they need to be minimized.

The fraud triangle is a model for explaining the factors that cause someone to commit occupational fraud. It consists of three components which, together, lead to fraudulent behavior:

1. Perceived unshareable financial need
2. Perceived opportunity
3. Rationalization



Fig 1: The Fraud Triangle

3. Data Mining Approach

Data mining analyzes the huge volumes of transactions and billing data and seeks out patterns, trends and clusters that reveal fraud. The main steps for implementing this approach for fraud detection within a business organization are:

1. Analyze the fraud objectives and the potential fraudsters, in order to converting them into data mining objectives;
2. Data collection and understanding;
3. Data cleaning and preparation for the algorithms;
4. Experiment design;
5. Evaluation results in order to review the process.

Relevant technical problems are due to:

1. Imperfect data not collected for purpose of data mining, so they are inaccurate, incomplete, and irrelevant data attributes;
2. Highly skewed data, there are many more legitimate than fraudulent examples, so by predicting all examples to be legal a very high success rate is achieved without detecting any fraud;
3. Higher chances of over fitting that occurs when model high accuracy arises from fitting patterns in the training set that are not statistically reliable and not available in the score set. To handle with skewed data the training set is divided into pieces where the distribution is less skewed (Chan, 1998).

A typical detection approach consists in outlier detection where the non-fraudulent behavior is assumed as normal and identifies outliers that fall far outside the expected range should be evaluated more closely.

Statistic techniques used for this approach are:

1. Predict and Classify

- Regression algorithms: neural networks, CART, Regression, GLM;
- Classification algorithms (predict symbolic Outcome): CART, logistic regression;

2. Group and Find Associations

- Clustering/Grouping algorithms: K-means, Kohonen, Factor analysis;
- Association algorithms: GRI, Capri Sequence.

Many existing fraud detection systems operate by: supervised approaches on labelled data, hybrid approaches on labelled data, semi-supervised approaches with legal (non-fraud) data, unsupervised approaches with unlabelled data (Phua, 2005). The classification framework, which is shown in Fig. 1, is based on a literature review of existing knowledge on the nature of data mining research [19, 20], fraud detection research [18].

A classification framework for financial fraud is suggested in [7] based on the financial crime framework of the U.S. Federal Bureau of Investigation [21], which is one of the established frameworks for financial fraud Detection. Fig. 2 consists of two layers, the first comprising the six data mining application classes of classification, clustering, prediction, outlier detection, regression, and visualization[22,23], supported by a set of algorithmic approaches to extract the relevant relationships in the data [14].

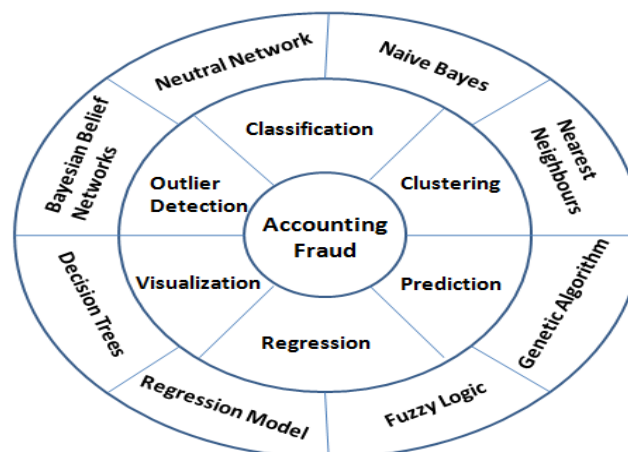


Fig2: Conceptual Framework for Application of Data Mining to Financial Accounting Fraud Detection.

A brief description of the conceptual framework with references is provided and of the six data mining application classes (classification, clustering, outlier detection, prediction, regression and visualization), each component of is discussed in more detail in the following sections.

3.1 Classification of Data Mining Applications

Each of the six data mining application classes is supported by a set of algorithmic approaches to extract the relevant relationships in the data. These approaches can handle different classes of problems. The classes are presented below.

Classification - Classification builds up and utilizes a model to predict the categorical labels of unknown objects to distinguish between objects of different classes. These categorical labels are predefined, discrete and unordered [24]. The research literature describes that classification or prediction is the process of identifying a set of common features (patterns), and proposing models that describe and distinguish data classes or concepts [17]. Common classification

techniques include neural networks, the Naïve Bayes technique, decision trees and support vector machines. Such classification tasks are used in the detection of credit card, healthcare and automobile insurance, and corporate fraud, among other types of fraud, and classification is one of the most common learning models in the application of data mining in fraud detection.

Clustering: Clustering is used to partition objects into previously unknown conceptually meaningful groups (i.e. clusters), with the objects in a cluster being similar to one another but very dissimilar to the objects in other clusters. Clustering is also known as data segmentation or partitioning and is regarded as a variant of unsupervised classification [24]. Cluster analysis decomposes or partitions a data set (single or multivariate) into dissimilar groups so that the data points in one group are similar to each other and are as different as possible from the data points in other groups [1]. It is suggested that data objects in each cluster should have high intra-cluster similarity within the same cluster but should have low inter-cluster similarity to those in other clusters [17]. The most common clustering techniques are the K-nearest neighbour, the Naïve Bayes technique and self-organizing maps.

Prediction: Prediction estimates numeric and ordered future values based on the patterns of a data set [19]. It is noted that, for prediction, the attribute, for which the value being predicted is continuous-valued (ordered) rather than categorical (discrete-valued and unordered). This attribute is referred as the predicted attribute [24]. Neural networks and logistic model prediction are the most commonly used prediction techniques.

Outlier Detection: Outlier detection is employed to measure the distance between data objects to detect those objects that are grossly different from or inconsistent with the remaining data set [24]. Data that appear to have different characteristics than the rest of the population are called outliers [26]. The problem of outlier/anomaly detection is one of the most fundamental issues in data mining. A commonly used technique in outlier detection is the discounting learning algorithm [27].

Regression: Regression is a statistical methodology used to reveal the relationship between one or more independent variables and a dependent variable (that is continuous-valued) [24]. Many empirical studies have used logistic regression as a benchmark [28]. The regression technique is typically undertaken using such mathematical methods as logistic regression and linear regression, and it is used in the detection of credit card, crop and automobile insurance, and corporate fraud.

Visualization: Visualization refers to the easily understandable presentation of data and to methodology that converts complicated data characteristics into clear patterns to allow users to view the complex patterns or relationships uncovered in the data mining process [14]. The researchers have exploited the pattern detection capabilities of the human visual system by building a suite of tools and applications that flexibly encode data using colour, position, size and other visual characteristics. Visualization is best used to deliver complex patterns through the clear presentation of data or functions [29].

3.2 Classification of Data Mining Techniques for Financial Accounting Fraud Detection

To determine the main algorithms used for financial accounting fraud detection, we present a Review of data mining techniques identified in literature applied to the detection of financial fraud. The most frequently used techniques are logistic models, neural networks, the Bayesian belief network, and decision trees, all of which fall into the classification category. These four techniques are discussed in more detail in the following paragraphs.

Regression Models: The regression based models are mostly used in financial accounting fraud detection. The majority of them are based on logistic regression, stepwise-logistic regression, multi criteria decision making method and exponential generalized beta two (EGB2) [7]. Logistic model is a generalized linear model that is used for binomial regression in which the predictor variables can be either numerical or categorical [30]. It is principally used to solve problems caused by insurance and corporate fraud.

Some of the research has suggested logistic regression based model to predict the presence of financial statement fraud [30]. Statistical method of logistic regression can detect falsified financial statements efficiently [30]. Some researchers have also developed generalized qualitative response model based on Probit and Logit techniques to predict financial statement fraud. That model was based on a dataset collected by an international public accounting company and needs testing for generalization [14]. Cascaded Logit model has also proposed to investigate the relationship between insider trading and possibility of fraud. The study in [14] found that, when the fraud is being executed, insiders, i.e. top executives and managers, reduce their stock holdings through high stock selling activity. The other methods like statistical regression analysis are also useful to test if the existence of an independent audit committee mitigates or reduces the likelihood of fraud. Literature also describes that organizations with audit committees, formed by independent managers, meeting no more than twice per year, are less likely to be sanctioned for fraudulent financial reporting [26].

The regression analysis using Logit model can be used for empirical analysis of financial indexes which can significantly predict financial fraud [29]. Logistic analysis and clustering analysis jointly can be used to establish a detecting model of fraud from four aspects of financial indexes, company governance, financial risk and pressure and related trading. After cluster filtering significant variables, prediction model can be established with methods of Standardization, non-Standardization Bayes and Logistic [4].

The logistic regression based accounting fraud detecting models are common in literature since the model based on logistic regression can reach up to 95.1% of detecting accuracy with significant expectation effect.

Neural Networks: The neural networks are non-linear statistical data modeling tools that are inspired by the functionality of the human brain using a set of interconnected nodes [11]. Neural networks are widely applied in classification and clustering, and its advantages are as follows. First, it is adaptive; second, it can generate robust models; and third, the classification process can be modified if new training weights are set. Neural networks are chiefly applied to credit card, automobile insurance and corporate fraud.

Literature describes that neural networks can be used as a financial fraud detection tool. The neural network fraud classification model employing endogenous financial data created from the learned behavior pattern can be applied to a test sample. The neural networks can be used to predict the occurrence of corporate fraud at the management level.

Researchers have explored the effectiveness of neural networks, decision trees and Bayesian belief networks in detecting fraudulent financial statements (FFS) and to identify factors associated with FFS [8].

The study in [10] revealed that input vector consisted of financial ratios and qualitative variables, was more effective when fraud detection model was developed using neural network. The model was also compared with standard statistical methods like linear and quadratic discriminant analysis, as well as logistic regression methods [10].

The generalized adaptive neural network architectures and the adaptive logic network are well received for fraud detection. The hybrid techniques like fuzzy rule integrated with a neural network (neuro-fuzzy systems) are also proposed. The literature describes that the integrated fuzzy neural network outperformed traditional statistical models and neural networks models reported in prior studies.

Bayesian Belief Network: The Bayesian belief network (BBN) represents a set of random variables and their conditional independencies using a directed acyclic graph (DAG), in which nodes represent random variables and missing edges encode conditional independencies between the variables [8]. The Bayesian belief network is used in developing models for credit card, automobile insurance, and corporate fraud detection. The research in [8] described that Bayesian belief network model correctly classified 90.3% of the validation sample for fraud detection. Bayesian belief network outperformed neural network and decision tree methods and achieved outstanding classification accuracy [8].

Decision Trees: A decision tree (DT) is a tree structured decision support tool, where each node represents a test on an attribute and each branch represents possible consequences. In this way, the predictive model attempts to divide observations into mutually exclusive subgroups and is used for data mining and machine learning tasks [8]. Decision trees are predictive decision support tools that create mapping from observations to possible consequences [24]. Predictions are represented by leaves and the conjunctions of features by branches. Decision trees are commonly used in credit card, automobile insurance, and corporate fraud.

Nearest Neighbor Method: Nearest neighbour method is a similarity based classification approach. Based on a combination of the classes of the most similar k record(s), every record is classified. Sometimes this method is also known as the k-nearest neighbour technique [24]. K-nearest neighbour method is used in automobile insurance claims fraud detection and for identifying defaults of credit card clients.

Fuzzy logic and Genetic Algorithm: Genetic algorithms are used in classifier systems to represent and modeling the auditor decision behavior in a fraud setting. Genetic algorithm along with binary support vector system (BSVS) which is based on the support vectors in support vector machines (SVM) are used to solve problems of credit card fraud that had not been well identified.

Fuzzy Logic is a mathematical technique that classifies subjective reasoning and assigns data to a particular group, or cluster, based on the degree of possibility the data has of being in that group. The expert fuzzy classification techniques enable one to perform approximate reasoning that can improve performance in three ways. First, performance is improved through efficient numerical representation of vague terms, because the fuzzy technology can numerically show representation of a data item in a particular category. The second way performance is enhanced is through increased range of operation in ill-defined environments, which is the way that fuzzy methodology can show partial membership of data elements in one or more categories that may not be clearly defined in traditional analysis. Finally, performance is increased because the fuzzy technology has decreased sensitivity to "noisy" data, or outliers. A multilevel fuzzy rule-based system is proposed in [9] to rank state financial management. The authors used fuzzy set theory to represent imprecision in evaluated information and judgments.

A fuzzy logic model has been implemented in [5] for fraud detection in an Excel spreadsheet. By using the fuzzy logic model to develop clusters for different statements representing red flags in the detection of fraud, non-financial data was included with financial statement variables for the analysis. The model consist of different financial variables like leverage, profitability, liquidity, cash flow and a variable designed to represent a company's risk of fraud. Fuzzy logic efficiently modeled the variable, which was developed to quantify fraud risk factors. The model predicted frauds with 86.7% accuracy [5]. The same model was adapted in [1] to develop a model for detection of financial statement fraud. The proposed model used a combination of different financial statement data.

Fuzzy logic based expert system has been developed to identify and evaluate whether elements of fraud are involved in insurance claims settlements. The fuzzy logic based expert system was developed for auditors to identify fraud in settled claimed insurance. The system was able to cut costs by detecting fraudulent filings.

Genetic programming with fuzzy logic production rules is used to classifying data. The study in [5] has proposed and tested a system to detect frauds on real home insurance claims and credit card transaction data. The study on genetic programming for fraud detection lacks benchmarking with the existing methods and techniques. A genetic algorithm based approach to detect financial statement fraud. It was found that exceptional anomaly scores are valuable metrics for characterizing corporate financial behavior and that analyzing these scores over time represents an effective way of detecting potentially fraudulent behavior.

Expert Systems: Researchers in the field of Expert systems have examined the role of Expert Systems in increasing the detecting ability of auditors and statement users. By using expert system, they could have better detecting abilities to accounting fraud risk under different context and level and enable auditors give much reliable auditing suggestions through rational auditing procedure. The research has confirmed that the use of an expert system enhanced the auditors' performance. With assistance from expert system, the auditors discriminated better, among situations with different levels of management fraud-risk. Expert System aided in decision making regarding appropriate audit actions.

The financial accounting fraud detection research is classified as per data mining application and data mining techniques. Some researchers have tried to apply a combination of many data mining techniques like decision trees, neural networks, Bayesian belief network, K-nearest neighbour. The main objective is to apply a hybrid decision support system using stacking variant methodology to detect fraudulent financial statements.

4. Data Mining Based Framework For Fraud Detection

The research related with application of data mining algorithms and techniques for financial accounting fraud detection is a well studied area. The implementation of these techniques follows the same information flow of data mining processes in general. The process starts with feature selection then proceeds with representation, data collection and management, pre-processing, data mining, post-processing, and in the end performance evaluation. This paper has proposed an expanded generic data mining framework. This framework considers specific characteristics of fraud detection techniques for financial accounting fraud (Refer Fig.3).

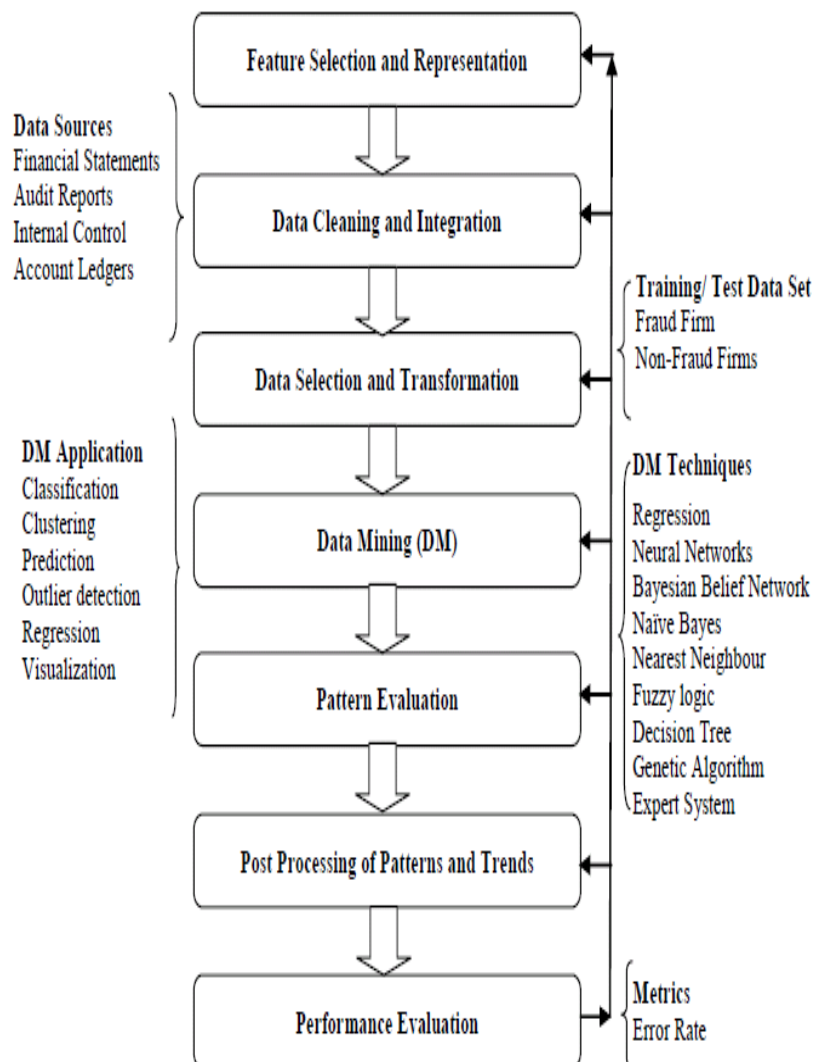


Fig 3: The Data Mining Based Framework for Financial Accounting Fraud Detection.

5. Conclusion:

This paper reviewed the literature describing use of data mining algorithms including statistical test, regression analysis, Neural Network, decision tree, Bayesian network etc for financial accounting fraud detection. Regression Analysis is widely used for fraud detection since it has great explanation ability. The researchers have not made any comparison so far, related with detecting effect and accuracy of Neural Network compared to regression model. The advantages of Neural Network are that there are no strict requests for data and it has a strong generalization and adjustment. After correct allocation and proper training, Neural Network may perform great classification comparing with regression model. But due to special inner hidden structure, it is impossible for researchers to track the formation process of output conclusion. There are other issues also related with Neural Network like no clear explanation on connecting weight, complex accuracy and statistical reliability checking procedure, and lack of explanation.

Future Study

This paper suggests that using only financial statements data may not be sufficient for detections of fraud. The importance of data mining techniques in the detection of financial fraud has been recognized. The future work may be proposing a comprehensive classification framework or a systematic review of data mining application in financial accounting fraud detection.

References

- [1] Yue, X., Wu, Y., Wang, Y. L., & Chu, C. (2007). A review of data mining-based financial fraud detection research, international conference on wireless communications Sep, Networking and Mobile Computing (2007) 5519–5522.
- [2] Oxford Concise English Dictionary, 11th Edition, Oxford University Press, 2009.
- [3] Phua, C., Lee, V., Smith, K. & Gayler, R. (2005). A comprehensive survey of data mining-based fraud detection research, Artificial Intelligence Review (2005) 1–14.
- [4] Wang, J., Liao, Y., Tsai, T. & Hung, G. (2006). Technology-based financial frauds in Taiwan: issue and approaches, IEEE Conference on: Systems, Man and Cyberspace Oct (2006) 1120–1124.
- [5] Wang, S. (2010). A Comprehensive Survey of Data Mining-Based Accounting-Fraud Detection Research. International Conference on Intelligent Computation Technology and Automation, vol. 1, pp.50-53, 2010.
- [6] Accounting Fraud Definition and Examples that are retrieved from <http://www.accountingelite.com/accounting-tips/accounting-fraud-definition-and-examples-free-accounting-fraud-article/>
- [7] Ngai, E.W.T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2010). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature, Decision Support System (2010), doi:10.1016/j.dss.2010.08.006.
- [8] Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements, Expert Systems with Applications 32 (4) (2007) 995–1003.
- [9] Fanning, K., Cogger, K., & Srivastava, R. (1995). Detection of management fraud: a neural network approach. International Journal of Intelligent Systems in Accounting, Finance & Management, vol. 4, no. 2, pp. 113–26, June 1995.
- [10] Fanning, K., & Cogger, K. (1998). Neural network detection of management fraud using published financial data. International Journal of Intelligent Systems in Accounting, Finance & Management, vol. 7, no. 1, pp. 21- 24, 1998.
- [11] Silverstone, Howard, & Sheetz, M. (2004). Forensic Accounting and Fraud Investigation for Non-Experts. Hoboken, John Wiley & Sons, 2004.
- [12] Bologna, Jack & Lindquist, R. J. (1987). Fraud Auditing and Forensic Accounting. New York: John Wiley & Sons, 1987.
- [13] Elkan, C. (2001). Magical Thinking in Data Mining: Lessons from COIL Challenge 2000. Proc. of SIGKDD01, 426-431.
- [14] Turban, E., Aronson, J.E., Liang, T.P., & Sharda, R. (2007). Decision Support and Business Intelligence Systems, Eighth edition, Pearson Education, 2007.
- [15] Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C.J. (1992). Knowledge discovery in databases: An overview, AI Magazine 13 (3) (1992) 57–70.
- [16] Bolton, R. J., & Hand, D.J. (2002). Statistical fraud detection: a review, Statistical Science 17 (3) (2002) 235–255.
- [17] Zhang, D., & Zhou, L. (2004). Discovering golden nuggets: data mining in financial application, IEEE Transactions on Systems, Man and Cybernetics 34 (4) (2004) Nov.
- [18] Derrig, R. A. (2002). Insurance fraud, The Journal of Risk and Insurance 69 (3) (2002) 271–287.
- [19] Ahmed, S.R. (2004). Applications of data mining in retail business, International Conference on Information Technology: Coding and Computing 2 (2) (2004) 455–459.
- [20] Mitra, S., Pal, S.K., & Mitra, P. (2002). Data mining in soft computing framework: a survey, IEEE Transactions on Neural Networks 13 (1) (2002) 3–14.
- [21] FBI, Federal Bureau of Investigation, Financial Crimes Report to the Public Fiscal Year, Department of Justice, United States, 2007, http://www.fbi.gov/publications/financial/fcs_report2007/financial_crime_2007.htm.
- [22] Fawcett, T., & Provost, F. (1997). Adaptive fraud detection, Data Mining and Knowledge Discovery 1 (3) (1997) 291–316.

- [23] Sánchez, D., Vila, M.A., Cerda, L., & Serrano, J.M. (2009). Association rules applied to credit card fraud detection, *Expert Systems with Applications* 36 (2) (2009) 3630–3640.
- [24] Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*, Second edition, Morgan Kaufmann Publishers, 2006, pp. 285–464.
- [25] Persons O.S. (1995). Using Financial Statement Data to Identify Factors Associated with Fraudulent Financial Reporting. *Journal of Applied Business Research*, 1995, 11(3):38-46.
- [26] Agyemang, M., Barker, K., & Alhaji, R. (2006). A comprehensive survey of numeric and symbolic outlier mining techniques, *Intelligent Data Analysis* 10 (6) (2006) 521–538.
- [27] Yamanishi, K., Takeuchi, J., Williams, G., & Milne, P. (2004). On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms, *Data Mining and Knowledge Discovery* 8 (3) (2004) 275–300.
- [28] Viaene, S., Derrig, R.A., Baesens, B., & Dedene, G. (2002). A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection, *The Journal of Risk and Insurance* 69 (3) (2002) 373–421.
- [29] Eick, S.G. & Fyock, D.E. (1996). Visualizing corporate data, *AT&T Technical Journal* 75 (1) (1996) 74–86.
- [30] Spathis, C. T. (2002). Detecting false financial statements using published data: some evidence from Greece, *Managerial Auditing Journal* 17 (4) (2002) 179–191.

AUTHORS BIBLIOGRAPHY:



Sarada Sowjanya.C is currently perusing her M.Tech in NOVA College of Engineering and Technology for Women, Ibrahimpatnam, and Vijayawada. She did her BE from RASTRASANT TUKADOGI MAHARAJ NAGPUR UNIVERSITY, NAGPUR (Maharatsra Dt). Her areas of interest is: Data Mining, Computer Networks, Data Base Management Systems.



G.Jyotsna completed her MBA from Jawaharlal Nehru Technological University Kukatpally, Hyderabad (JNTUH). She did her B.Tech from Chaitanya Bharathi Institute of Technology, (CBIT) Hyderabad. Her area of interest is: HR marketing.