Financial Frauds: Data Mining based Detection – A Comprehensive Survey

Aastha Bhardwaj
Assistant Professor
Vivekananda Institute of
Professional Studies
Pitampura, Delhi - 110034, India

Rajan Gupta, PhD
Associate Professor
Vivekananda Institute of
Professional Studies
Pitampura, Delhi - 110034, India

ABSTRACT
Financial fraud is a global problem and had affected the economy worldwide. Data mining being one of the most effective and powerful tool for detecting financial fraud had been used widely by the business analysts and researchers. This survey paper formalizes different types of financial frauds, summarizes the effective attributes for detecting each type of fraud, and present the latest developments on the use of data mining as a detection tool for financial frauds. The present survey analyses almost all published research work in the field of financial fraud detection for the period of 7 years starting from 2009. Its aim is to help researchers in identifying the suitable variables and data mining techniques by providing the landscape of research platforms for detection of financial fraud.

Keywords
Financial fraud, Management fraud, Customer fraud, Task relevant data, Data mining, Credit card fraud, Insurance fraud.

1. INTRODUCTION
Fraud not only causes unimaginable financial losses but also pushes the organization by many steps backwards in this cut - throat competitive world. Any deliberate act of deceit involving financial accounts or misappropriation of organizations' assets/ resources for personal enrichment is termed as "Financial Fraud". It is an organized crime, orchestrated with a view to grind one's own axe. Financial fraud encompasses various types of crimes and unlawful activities such as identity theft, asset misappropriation and many more. The fraudster resorts to various methods for duping the victim. Nonetheless, for committing a fraud the first condition being that perpetrator is more than conversant with working of the system thereby outwitting the structure. The degree of this unlawful activity is directly proportional to the access (gained through any means) one has to the system. Thus to counter this menace, system has to be proactive and accordingly implement fraud prevention and detection techniques.

In present scenario, implementing effective fraud prevention methods at first place and detection technique in case of failure of preventive measures is no more a competitive advantage but a reason that ensures the survival of the fittest. The chances of fraud may be reduced to a level by judging the accuracy of intention and legitimacy of financial transactions, which is almost impossible.

Data mining being a process of extracting knowledge by learning patterns from the available data has been used widely for developing fraud detection systems. It can identify useful and interesting patterns with efficacy, which can be used to find out any inconsistent behavior or fraudulent activity. Researchers from both industry and academia have designed a number of automated/semi -automated data mining systems for detection of financial frauds.

Data mining is known as gaining insights and identifying interesting patterns from the data stored in large databases in such a way that the patterns and insights are statistically reliable, previously unknown, and actionable [1]. Data mining is also defined as - a process that uses statistical, mathematical, artificial intelligence and machine learning techniques to extract and identify useful information and subsequently gaining knowledge from a large database [2]. Selection of task relevant data is one of the most important primitive for data mining tasks. Selecting task relevant attribute from large datasets is one of the major hurdle faced by researchers in order to design automated fraud detection systems.

In order to solve the problem of selecting effective attributes, this paper primarily aims to list the attributes already used by previous fraud detection experimental studies and automated systems.

This will assist researchers for selecting relevant attributes for designing more effective fraud detection systems.

Secondly, this paper presents a comprehensive analysis of data mining techniques used for detection of each type of financial fraud. This will lay foundation to provide scope for further research in the field of fraud detection.

The rest of the paper is organized as follows. Section 2 defines different types of financial frauds and helps in selecting task relevant data / attributes for detection of each type of fraud by analyzing the available research articles for the period of seven years. Section 3 discusses different data mining techniques* devised and implemented by various researchers for detecting financial frauds, followed by conclusion and future work (Section 4).

2. FINANCIAL FRAUDS
Fraud may be defined as, any intentional act in order to deceive or mislead another person or organization for financial benefits. This deliberate, illegal fraudulent activity may be defined and classified in number of ways depending on type of perpetrators.

Fraud committed by individuals external to the organization is termed as customer fraud or external fraud whereas, fraud committed by top - level management is known as management fraud or internal fraud (Fig I). In this paper, we had classified fraud into two categories namely management fraud and customer fraud.

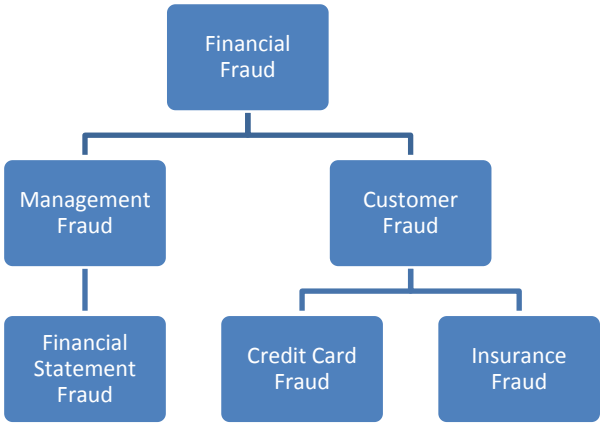

Fig I: Classification of Financial Fraud

2.1 Management Fraud

An intentional act committed by employees, internal auditors, executives, the board of directors, and managers, who may suffer a financial loss and or reputation loss, is termed as management fraud.

In management fraud, CEO's and executive managements are the perpetrator since they are capable of falsification of expenses, invoices, sales figure etc. Management fraud often called financial statement fraud is a deliberate misstatement of material facts by the management in the books of accounts of a company with the aim of deceiving investors and creditors. This illegitimate task performed by management has a severe impact on the economy throughout the world because it significantly dampens the confidence of investors [3].

Data mining methods are the most widely used technique for detection of financial statement fraud because data mining is capable of extracting novel patterns from large databases by building models, which can further be used for making crucial business decisions. Researchers had applied and evaluated a number of data mining techniques for differentiating fraud and non – fraud organizations. Data set for detection of fraudulent financial reporting consists of financial ratios from publicly available financial results of the organization.

For example, Belinna et al [4] used 148 financial reports out of which 24 were false reports. Cecchini et al [5] proposed a methodology to aid in detecting fraudulent financial reporting by utilizing only basic and publicly available financial data. Data sample consist of data of 205 fraudulent companies. The data was gathered by using accounting and auditing enforcement releases (AAERs). The fraud sample was matched with 6,427 non - fraudulent companies. Perols Johan L [6] compared the performance of six popular statistical and machine learning models in detecting financial statement fraud under different assumptions of misclassification costs and ratios of fraud firms to non - fraud firms. The fraudulent observations were located based on firms investigated by the SEC for financial statement fraud and reported in Accounting and Auditing Enforcement Releases (AAER) from the fourth quarter of 1998 through the fourth quarter of 2005. A total of 42 predictors were examined, only six were consistently selected and used by different classification algorithms. Ravisankar et al [7] predicted the occurrence of financial fraud by using six data mining techniques by analyzing data from 202 Chinese companies. Thirty-five financial ratios were considered for detecting fraudulent financial reporting. Gupta et al [3] detected fraudulent financial reporting by using three data mining techniques by analyzing data from 114 listed companies. 63 financial ratios were considered for fraud detection.

The summary of known research in the field of financial statement fraud detection, conducted till date reveals that input vector consists of financial ratios. Based on existing academic research and expert's knowledge, financial ratios that are relevant to the task of detecting management fraud are summarized in this section.

This section will help future studies and researchers in selecting the effective task relevant attributes / financial ratios.

The summary of the factors that should be considered for detecting financial statement fraud are given below. (Table 1).

2.1.1 Z-score

Financial distress may be a motivation for management fraud [8]. To measure the financial distress Z-score is developed by Altman [9]. It is a formula for estimating the financial status of a company and helpful in bankruptcy prediction. The formula for Z-score for public companies is given by:

Z-score= (Working capital / Total assets* 1.2) + (Retained earnings ÷Total assets* 1.4) + (Earnings before income tax ÷Total assets* 3.3) + (Book value of total / Liabilities * 0.6) + (Sales ÷Total assets* 0.999)

2.1.2 A high debt structure

A high debt structure may be an indicator for fraudulent financial reporting, because it shifts the risk from mangers to debt owners. Hence, we can state that higher levels of debt may increase the likelihood of financial statement fraud and one should carefully consider the financial ratios related to debt structure.

2.1.3 Continues growth

The need for continues growth may be another motivational factor for financial statement fraud [10]. So sales to growth ratio should be measured as a fraudulent financial statement indicator. Sales to growth = (Current Year's sales - Last Year's sales) / (Last Year's sales)

2.1.4 Other items

A company may manipulate accounts receivable, inventories, and gross margin. Accounts receivable may be manipulated by recording sales before they are earned. Inventory is also prone to manipulation. Mangers may manipulate inventory either by reporting inventory at lower cost or by obsolete inventory or both. A company may use gross margin as a factor for falsifying financial statement. The company may not match its sales with the corresponding cost of goods sold, thus increasing gross margin, net income and strengthening the balance sheet [8].

Table I: Task relevant attributes for detecting management fraud

S. No.	Financial Ratios / Items
1	Inventory / Total Assets [3], [11], [12],[13]
2	Asset Turnover [8], [12],[40]
3	Net profit/Total assets (ROA) [3], [7], [8], [13]
4	Gross Profit to Total Asset [7], [8], [12]
5	Total Liabilities to Equity [7], [8],[40]
6	Cash / Total Assets [3], [7],[40]
7.	Net Profit Margin [8], [12]
8.	Quick Assets / Current Liabilities [3], [14],[40]

2.2 Customer Fraud

Acquisition of goods/services resorting to unethical means or deceiving an organization by the customer for personal gains can be termed as customer fraud. In this type of fraud, a customer acquires the goods/services by unethical means or deceives an organization with an intention to commit financial loss. A customer can mislead various financial institution and insurance companies that will result two sub categories of customer fraud namely credit card fraud and insurance fraud.

2.2.1 Credit card fraud

Revolution from traditional commerce to ecommerce has compelled the use of credit card on a large scale. According to RBI [15], more than 6 crore of transactions worth Rs. 190989.13 Million went through in May 2015. Unfortunately, this intensifying usage also invites criminals to fraudulently use credit cards to earn money / acquire product or service by unethical means. According to the Nilson Report [16], fraud losses on credit cards, debit cards, and prepaid cards worldwide hit $16.31 billion in 2014 on a total card sales volume of $28.844 trillion. A study released in 2016 by New LexisNexis Risk Solutions [17] revealed that credit card fraud costed $7.6 billion. This rising number is an alarming call to provide some automatic intelligent system that can detect fraud before it is being committed.

Data mining is one of the most fascinating approach used widely for credit card fraud detection. It is defined as "a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid prediction" [18]. John [19] states that there is a fixed pattern to how credit-card owners consume their credit card on the internet. From this statement, it can be deduced that if a customer deviates from his normal course of behavior then there is something suspicious. Although not every asymmetrical action ensures fraud but chances are quite high of being deceptive. Keeping a regular watch on all activities of the user can prove to be beneficial in detecting a fraudulent act. The goal of a reliable detection system is to learn the behavior of users dynamically to minimize its own loss. Thus, systems that cannot evolve or ''learn", may soon become outdated resulting in large number of false alarms" [20]. Researchers have used number of attributes from the database, which defines the profile of a customer and pattern of his transactions. John[19]developed a neural network model, trained with attributes like merchants' websites, regular good and services purchased in past transactions of credit card holder; shipping address, email address and phone number of customer and geolocation of transaction. Whereas, Avinash et al [22] takes into consideration, the factors revealing the cardholder's spending behavior, i.e. columns related to his past transactions. Jyotindra et al's [23] model has taken 10 parameters like category of the purchase, same product purchased within short time, Late night transaction, Overseas transaction etc. from a dataset of Online shopping firm's credit card transaction data.

It is evident from the above discussion, that every past literature work has different parameters' name depending on the available transactions' database. Therefore, while selecting an attribute in consideration for future research work, focus should be on the type of information it is providing i.e. category in which it is being classified rather than restricting to the name of attributes in the database.

Various factors one should consider while selecting the effective attributes are broadly classified as under:

2.2.1.1 Income and spending/historical transactions' pattern:

Attributes which gives us the information regarding the spending behavior of a customer and pattern of the transactions.

Ex: Merchants' websites, regular good and services purchased in past credit cardholder's transactions [19], frequency of transactions [20], Number of the transactions, Same product purchased within short time, Time passed since the last transaction [23].

2.2.1.2 Customer Profile:

The attributes defining the characteristics and profile of a customer.

Ex: Email address and Phone number [19], Personal status & sex, Job, Foreign worker (Yes or No), Telephone and age [21].

2.2.1.3 Credibility:

Attributes revealing the assets and liabilities of the customer.

Ex: property, present address since, history of credit taken, present employment since, credit amount, instalment rate in %, saving account/bonds, existing credits at this bank, Number of people liable to provide maintenance for, guarantors, other instalment plans, housing [21].

2.2.1.4 Transaction related:

Details regarding transaction of the goods/services done by a customer plays an eminent role in determining any unusual activity. Summary of the attributes providing transactions related data are given below (Table II).

Table II: Task relevant data for detecting credit card fraud

S. No.	Attributes
1	Shipping address [19][20]
2	Amount of transaction [20][22] [23][24][38]
3	Geolocation of real-time transaction [19]
4	Billing address [20]

5	Location from which product is ordered [23]
6	Category of the purchase [23]
7	Time frame during which product is ordered [23]
8	Seller or Vendor[23]
9	With whom product is purchased[23]
10	Late night transaction[23]
11	Overseas transaction[23]
12	Transaction time[38]
13	Point of sale, currency, country, merchant type[24]

2.2.1.5 Hybrid:

Some variables are merged with already available data to form hybrid traits. This hybridity of divergent characteristics can prove useful in identifying new parameters. Ex: The transaction amount and the card ID is used to compute the average expenditure per week and per month of one card, the difference between the current and previous transaction and many others. Average daily/monthly spending of a customer, frequency of the transactions on that card [24].

2.2.2 Insurance fraud

An act performed by the insured person / beneficiary to apply for compensation by producing fake documents / reports is termed as insurance fraud. According to India forensic Research, every single insurance company loses 8.5% of its revenues to the frauds [25]. Teris Roberts [26]has suggested an suspicious activity assessment for insurance frauds wherein system takes care of risk factors like claim profile, policy profile, customer profile, entity profile and network profile and grades score to all the related entities(customer, broker etc.) at regular intervals.

Automobile insurance fraud includes staged automobile accident and a real accident with fabricated bills, thespian accidents, excessive repairs, and fictitious personal injuries all with one intention in mind i.e. false insurance claims resulting in financial loss to the companies [27]. Rekha [28] collected data from attributes like Policy Holder, Driver Rating, Vehicle Age, Price, and Report Filed. Liu et al [29] used a dataset of 5000 records with six attributes namely age, gender, claim amount, tickets, claim times, and accompanied with attorney. Lovera et al [30] suggests that staged accidents have several common characteristics. They occur in late hours and non-urban areas in order to reduce the probability of witnesses. Drivers are usually younger males; there are many passengers in the vehicles, but never children or elders. The police is always called to the scene to make the subsequent acquisition of means easier.

Thus, it can be concluded that attributes specifying the age and gender of driver, driver rating, age of passengers, and number of prior claims, price, and age of vehicle should be considered while selecting the dataset for further research work.

3. DATA MINING TECHNIQUES

Data mining is the process of discovering interesting/hidden patterns and extracting/mining knowledge from large amounts of data. It is popularly used to effectively find frauds because of its efficiency in discovering or recognizing unusual or unknown patterns in a collected set of dataset [19]. There are numerous techniques for finding interesting patterns from large data set. Broadly, these techniques can be classified into two categories: predictive and descriptive.

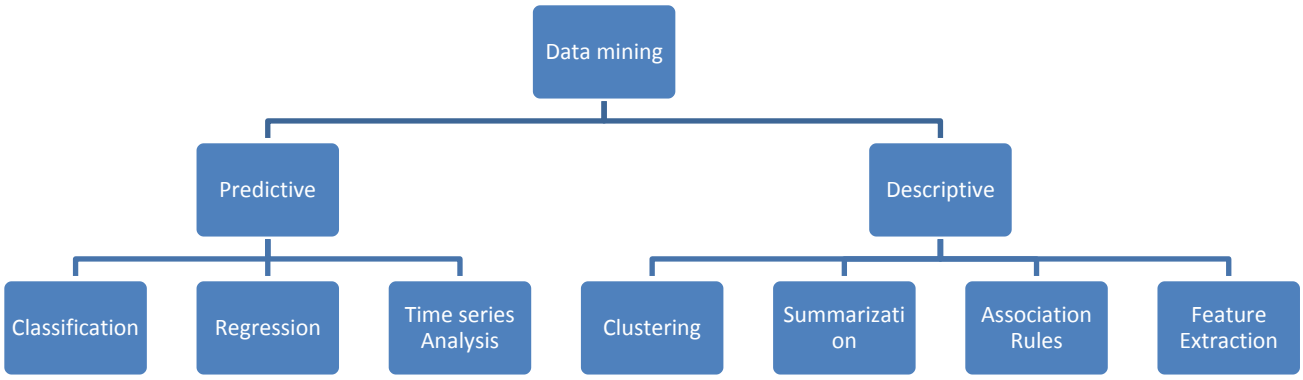

Fig II: Data Mining Techniques [31]

3.1 Predictive Data Mining Techniques

A predictive data-mining model predicts the value (numeric data/class label) of a specific attribute label using previously known data. As explained in the figure 2, there are multiple techniques that can be used for prediction and among different predictive data mining techniques; classification is considered as the best-understood technique of all data mining approaches [31].

Bhowmik R. [28] devised a model with accuracy of 78% using Naïve Bayesian Classification, and Decision Tree-Based algorithms to predict occurrence of Auto Insurance Fraud. Gupta et al [3] showed that Decision tree (CART) produces best sensitivity amongst Naive Bayesian Classifier and Genetic Programming by identifying more than 86% management fraud cases. Belinna et al [4] used CART to develop two FFS (financial fraud statements) detecting models: CART without industry benchmark and CART with industry benchmark. They found that both CART models achieve better accuracy in identifying management fraud cases and making predictions than Logit regression. CART with industry benchmark is slightly better than CART without

benchmark, but it does not always have superior performance. According to the results of Suduan Chen [40], the detection performance of the CHAID –CART) model is the most effective, with an overall accuracy of 87.97 %, the FFS detection accuracy is 92.69 %.

Cecchini et al [5] showed that in detecting management frauds, support vector machines using the financial kernel correctly labelled 80% of the fraudulent cases and 90.6% of the non-fraudulent cases on a holdout set. Johan Perols [6] showed that logistic regression and support vector machines perform well relative to an artificial neural network, bagging, C4.5, and stacking in detecting management frauds. Meenakshi et al [32] exhibited that classification by support vector machine has accuracy of 81% and by particle swarm optimization has accuracy of 95% in detecting credit card fraud. Luis Alexandre Rodrigues et al [39] created classifiers by combining C4.5, SVM and Naïve Bayes algorithm to find the most saving model to detect suspected cases of auto claim fraud.

Sanjay et al [35] proved that Bagging algorithm performs better classification of Credit card fraud detection technique as compared to Adaboost, Logitboost, CART and Dagging with 0.877 correct classification rate and 0.123 correct misclassification rate. Masoumeh et al [36] evaluated various methodologies like Naïve Bayes, Support Vector Machines, K-Nearest neighbour , Bagging classifier based on decision tree based on certain design criteria and introduced bagging classifier based on decision tree as the best classifier to construct credit card fraud detection model.

Junjie Wu et al [37] showed that COG(classification using local clustering) shows the best performances on both the normal and rare classes. COG can improve the prediction performances on rare classes, and this improvement is achieved without a big loss of the prediction power on normal classes. For data sets with imbalanced class distributions, the COG method can improve the performance of traditional supervised learning algorithms, such as support vector machines, on rare class analysis. COG-OS (COG with Over-Sampling) achieved better performances on both rare classes and normal class. It performs much better than pure SVMs and RIPPER on predicting the rare class as well as the normal class. They demonstrated that COG-OS is a prospective solution to the difficult classification problem induced by complex concepts and imbalanced class distributions. However, for the non-linear classifiers, such as C4.5 and RIPPER, COG shows no competitive results.

Neural network is one of the most popular data mining technique used by many researchers because of its ability to adapt and generalize [33]. The advantages of neural networks over other techniques is that they are capable of learning from the past, thus can improve results with passage of time. Francisca [34] used self-organizing map neural network (SOMNN) technique for credit card fraud (CCF) detection and detected over 95% of fraud cases without causing false alarm.

3.2 Descriptive Data Mining Techniques

A Descriptive mining model discovers and finds interesting patterns and relationships in bulk of data. It is normally used to generate correlation, frequency, cross tabulation, etc. Descriptive method can be defined as to discover regularities in the data and to uncover patterns [35].

Jyotindra et al[23] proposed transaction risk generation model (TRSGM) consists of five major components, namely, DBSCAN algorithm, Linear equation, Rules, Data Warehouse and Bayes theorem. Using first four components, suspicion level of each transaction determine the suspicion level of each incoming transaction based on the extent of its deviation from good pattern. The transaction is classified as genuine, fraudulent, or suspicious depending on this initial belief. Once a transaction is found to be suspicious, belief is further strengthened or weakened according to its similarity with fraudulent or normal transaction history using Bayes theorem.

Panigrahi S. et al [20] proposed a credit card fraud detection system (FDS), integration of three approaches, i.e. rule-based filter, Dempster–Shafer theory and Bayesian learning. They showed that combining rules using Dempster–Shafer theory gives good performance, giving up to 98% and less than 10% in terms of true positives and false positives respectively.

4. CONCLUSION & FUTURE SCOPE

Financial fraud either via customer or by the management is causing huge amount of monetary loss and has various negative consequences. Increasing magnitudes, frequency and drastically evolving ways of committing fraudulent financial transactions are giving the high sign for developing automated process rather than still relaying on humans. Frequency of fraudulent transactions is too less as compared to non-fraudulent ones. Instances galore where financial frauds have been kept a secret due to number of reasons. This scarcity of sample data available due to either beggarly occurrence of fraudulent transaction or not disclosed for confidentiality issues is one of the biggest challenge in designing a fraud detecting data-mining algorithm. These factors have given way to frauds of larger magnitudes and their occurrence has grown manifold. Devising a solution to this global problem involves the selection of task relevant data and implementing an effective data-mining algorithm. A number of variables are taken into account encompassing wide range of algorithms into it so that predictions so made can be more accurate. The past literature work has used several data mining techniques and various attributes have been chosen from the database for selecting the task relevant data. Since these attributes acts as an input vector to the fraud detection system, they must be selected after critical analysis of all the attributes present in the system. This paper analyzed each attribute and prepared a list of effective attributes for preparing more effective fraud detection systems. It will facilitate future research work by providing handy details regarding which attributes to be considered and selecting a better and effective data mining technique. It has the potential to assist the investigators by giving them a summary having accuracy of different data mining techniques and effectiveness of various variables to be used for finding task relevant data.

5. APPENDIX

S.no	Name	Author	Year	Technique	Dataset	Findings
1.	Credit card fraud detection: A fusion approach using Dempster Shafer theory and Bayesian learning. [20]	S. Panigrahi, A. Kundu, S. Sural, and A. Majumdar	2009	Density based clustering technique	Credit card data	Extensive simulation with stochastic models shows that fusion of different evidences has a very high positive impact on the performance of a credit card fraud detection system as compared to other methods.
2.	COG: local decomposition for rare class analysis. [36]	Junjie Wu · Hui Xiong · Jian Chen	2010	Local clustering (COG)	Data set is from a security company.	COG method can improve the performance of traditional supervised learning algorithms, such as support vector machines (SVMs), on rare class analysis. It has the capability in decomposing complex structures/concepts in the data into simple and linearly separable concepts, and thus enhancing linear classifiers on data sets containing linearly inseparable classes.
3.	Improving the Defense Lines: The Future of Fraud Detection in the Insurance Industry (with Fraud Risk Models, Text Mining, and Social Networks) [25]	Terisa Roberts, SAS Institute Inc.,	2010	Combination of text data analysis, social networks, and artificial intelligence	Major insurance company in South Africa	Improves the accuracy of fraud risk models and maintains an easy-to-implement and easy-to-interpret design.
4.	Data mining application in credit card fraud detection system.[33]	FRANCISCA NONYELUM OGWUELEKA	2011	Self-organizing map neural network (SOMNN) technique	Nigerian Bank's 18752 transactions of deposit and withdrawal	The receiver-operating curve (ROC) for credit card fraud (CCF) detection watch detected over 95% of fraud cases without causing false alarms The performance of CCF detection watch is in agreement with other detection software, but performs better than other statistical models and the two-stage clusters.
5.	A Data Mining with Hybrid Approach Based Transaction Risk Score Generation Model (TRSGM) for Fraud Detection of Online Financial Transaction.[23]	Dr. Jyotindra ,N. Dharwa ,Dr. Ashok R. Pate	2011	Hybrid containing data mining techniques, statistics and artificial intelligence.	Online shopping firm's Credit card transaction data	Proposed a flexible Model in which new rules can be easily added .It can detect new kind of fraud as well rather than only sticking past fraudulent dataset.
6.	Detecting Auto Insurance Fraud by Data Mining Techniques[27]	Rekha Bhowmik	2011	Naïve Bayesian classifier and Decision Tree-Based algorithm	Cases of automobile insurances	This model is strong with respect to class skew, making it a reliable performance metrics.
7.	Detection of	P. Ravisankar,	2011	Multilayer	Data from	PNN outperformed all the techniques

	Financial Statement Fraud and Feature Selection using Data Mining Techniques, Decision support system[7]	V. Ravi, G.Raghava Rao, I., Bose		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)	202 Chinese companies (101 were fraudulent and 101 were non fraudulent)	without feature selection, and GP and PNN outperformed others with feature selection and with marginally equal accuracies
8.	A Hybrid Data Mining Metaheuristic Approach for Anomaly Detection. [31]	Meenakshi, Sandeep Jaglan	2012	Hybrid of support vector machine and particle swarm optimization	Four training dataset of 100,200,500 and 700 transactions.	Used Support vector machine to classify the genuine and deceitful transaction and Particle swarm to optimize the high dimensional space.
9.	Application of Evolutionary Data Mining Algorithms to Insurance Fraud Prediction[28]	Jenn-Long Liu and Chien-Liang Chen	2012	Combination of GA-K means and MPSO-K means algorithms	5000 instances for car insurance claim	Accuracy of insurance fraud prediction can be enhanced by using the combination of two EvoDM algorithms than using only Kmeans algorithm.
10.	Prevention and Detection of Financial Statement Fraud – An Implementation of Data Mining Framework[3]	Rajan Gupta, Nasib Singh Gill	2012	Decision Tree, Naïve Bayesian Classifier, Genetic programming	Financial statements for 114 listed companies	Decision tree produces best sensitivity and Genetic programming best specificity as compared with other two methods.
11.	Meta Learning Algorithms for Credit Card Fraud Detection[34]	Sanjay Kumar Sen, ,Prof. Dr Sujata Dash	2013	Comparisons of machine learning algorithms	Credit card transactions	Showed that BAGGING machine learning classifier performance is better than Classification and Regression technique, Adaboost, Logitboost.
12.	Learned lessons in credit card fraud detection from a practitioner perspective[24]	Andrea DAL POZZOLO, Olivier CAELEN, Yann-Ael LE BORGNE, Serge WATERSCHOOT, Gianluca BONTEMPI	2014	Compared three approaches (static, update and forgetting) to learn from unbalanced and non-stationary credit card data streams	Logs of a subset of transactions from February 2012 to the twentieth of May 2013 from a Belgium payment service provider	Proposes AP, AUC and Precison Rank as correct performance measures for a fraud detection task. Proposed a way to include cardholder information into the transaction by computing aggregate variables on historical transaction of the same card.

13.	Application of credit card fraud detection: Based on bagging ensemble classifier.[35]	Masoumeh Zareapoor, Pourya Shamsolmoali	2014	Naïve Bayes, Support Vector Machines, K-Nearest neighbour , Bagging classifier based on decision tree.	Credit card dataset from UCSD-FICO competition	Bagging classifier based on decision tree detect fraudulent transactions than other methods
14.	”Credit Card Fraud Detection using Time Series Analysis”. [37]	R.Devaki V.Kathiresan S.Gunasekaran	2014.	Distance Based Method and time-series analysis	Online transactions via credit card in a bank	The approach used in the proposed work is very effective and also has decreased the false positive situation.
15.	Auto claim fraud detection using multi classifier system [38]	Luis Alexandre Rodrigues and Nizam Omar	2014	Decision Tree C4.5, Naive Bayes and Support Vector Machines (SVM)	The dataset has suspected 15.421 cases of frauds from 1994 to 1996 having	Combination of classifiers by C4.5, SVM and Naïve Bayes algorithm to find the most saving model to detect suspected cases of fraud.
16.	Detection of fraudulent financial statements using the hybrid data mining approach [40]	Suduan Chen	2016	Classification and regression trees (CART) and the Chi squared automatic interaction detector (CHAID)	Fraudulent and non-fraudulent financial statements of Companies from 2002 to 2013	The detection performance of the CHAID–CART model is the most effective

6. REFERENCES

[1] Elkan, C. (2001).Magical Thinking in Data Mining: Lessons from COIL Challenge 2000. Proc. of SIGKDD01,426-431.

[2] Turban, E., Aronson, J.E., Liang, T.P., &Sharda, R. (2007).” Decision Support and Business Intelligence Systems”, Eighth edition, Pearson Education, 2007.

[3] Gupta and Nasib S. Gill (2012), “Prevention and Detection of Financial Statement Fraud – An Implementation of Data Mining Framework”, International Journal of Advanced Computer Science and Applications, Volume 3 No. 8, pp. 150 – 156, Published by The Science and Information Organization, U.S.A.

[4] Belinna Bai, Jerome yen, Xiaoguang Yang, False Financial Statements: Characteristics of China Listed Companies and CART Detection Approach, International Journal of Information Technology and Decision Making , Volume 7, No. 2(2008), pp. 339 – 359.

[5] M. Cecchini, H. Aytug, G.J. Koehler, and P. Pathak. “Detecting Management Fraud in Public Companies.”, Management Science, Volume 56, No. 10, 2010, pp. 1146 – 1160.

[6] Johan Perols, Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms, A Journal of Practice &Theory, 30 (2), 19 (2011), pp. 19-50.

[7] P. Ravisankar, V. Ravi, G.Raghava Rao, I., Bose, Detection of Financial Statement Fraud and Feature Selection using Data Mining Techniques, Decision Support Systems, Volume 50 (2011), pp. 491 – 500.

[8] Fanning, K., &Cogger, K. (1998). Neural network detection of management fraud using published financial data. International Journal of Intelligent Systems in accounting, Finance & Management, 7(1), 21–24.

[9] E.I. Altman, Financial ratios, discriminant analysis and prediction of corporate bankruptcy, The Journal of Finance 23 (4) (1968) 589–609.

[10] Stice J., Albrecht S. and Brown L., (1991), ‘Lessons to be learned-ZZZZBEST, Regina, and Lincoln Savings’, The CPA Journal, April, pp. 52-53.

[11] Dalnial, Hawariah, et al. "Detecting Fraudulent Financial Reporting through Financial Statement Analysis." Journal of Advanced Management Science Vol 2.1 (2014).

[12] Spathis, C., M. Doumpos and C. Zopounidis. 2002. “Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques”. The European Accounting Review, 11 (3): 509-535.

[13] Kirkos, E., C. Spathis and Y. Manolopoulos. 2007. Data mining techniques for the detection of fraudulent financial statements. Expert Systems with Applications, 32 (4): 995-1003.

[14] H. Ali ATA, İbrahim H. SEYREK. "The use of data mining techniques in detecting fraudulent financial statements: an application on manufacturing firms.2009." The Journal of Faculty of Economics and Administrative Sciences Vol.14, No.2 pp.157-170.

[15] ATM & Card Statistics for May 2015 by RBI. Available at https://rbi.org.in/Scripts/ATMView.aspx?atmid=51

[16] Report available at https://www.nilsonreport.com/upload/pdf/Global_Card_Fraud_Damages_Reach_16B_-_PYMNTS.com.pdf

[17] Report available at https://nilsonreport.com/upload/pdf/Card_Fraud_Costing_Issuers_10.9_Billion_Annually_-_Yahoo_Finance.pdf

[18] Edelstien, H.A. (1999). Introduction to data mining and knowledge discovery. (2nd Ed.), Two Crows Corporation.

[19] John Akhilomen."Data Mining Application for Cyber Credit-card Fraud Detection System"; Journal of Engineering Science and Technology Vol. 6, No. 3 (2011) 311 - 322 . Proceedings of the World Congress on Engineering 2013 Vol III, WCE 2013, July 3 - 5, 2013, London, U.K

[20] Suvasini Panigrahi , Amlan Kundu , Shamik Sural , A.K. Majumdar." Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning". Information Fusion 10 (2009) 354–363.

[21] Azeem Ush Shan Khan, Nadeem Akhtar and Mohammad Naved Qureshi "Real-Time Credit-Card Fraud Detection using Artificial Neural Network Tuned by Simulated Annealing Algorithm". DOI: 02.ITC.2014.5.65 © Association of Computer Electronics and Electrical Engineers, 2014

[22] Avinash Ingole, Dr. R. C. Thool ."Credit Card Fraud Detection Using Hidden Markov Model and Its Performance". International Journal of Advanced Research in Computer Science and Software Engineering. June 2013.

[23] Dr. Jyotindra N. Dharwa Dr. Ashok R. Patel. A Data Mining with Hybrid Approach Based Transaction Risk Score Generation Model (TRSGM) for Fraud Detection of Online Financial Transaction. International Journal of Computer Applications (0975 – 8887) Volume 16– No.1, February 2011.

[24] Andrea DAL POZZOLO, Olivier CAELENb, Yann-A¨el LE BORGNEa, Serge WATERSCHOOT, Gianluca BONTEMPI. "Learned lessons in credit card fraud detection from a practitioner perspective", 2014.

[25] India Forensic research available at http://indiaforensic.com/certifications/india-loses-6-25-billion-to-insurance-frauds-an-indiaforensic-research.

[26] Terisa, R. "Improving the defense lines: the future of fraud detection in the insurance industry (with fraud risk models, text mining, and social networks)." SAS Global forum, Washington. 2010.

[27] E.W.T. Ngai, H. Yong, Y.H. Wong, C. Yijun and S. Xin, "The application of data mining techniques in financial fraud detection: A Classification Framework and an Academic Review of Literature". Decision Support Systems, Volume 50, Issue 3, February 2011.

[28] Bhowmik, Rekha. "Detecting auto insurance fraud by data mining techniques." Journal of Emerging Trends in Computing and Information Sciences 2.4 (2011): 156-162.

[29] Jenn-Long Liu and Chien-Liang Chen ."Application of Evolutionary Data Mining Algorithms to Insurance Fraud Prediction". Proceedings of 2012 4th International Conference on Machine Learning and Computing IPCSIT vol. 25 (2012) © (2012) IACSIT Press, Singapore.

[30] Šubelj, Lovro, Štefan Furlan, and Marko Bajec. "An expert system for detecting automobile insurance fraud using social network analysis." Expert Systems with Applications 38.1 (2011): 1039-1052.

[31] Pradnya P. Sondwale."Overview of Predictive and Descriptive Data Mining Technique". International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, April 2015.

[32] Meenakshi, Sandeep Jaglan." A Hybrid Data Mining Metaheuristic Approach for Anomaly Detection". International Journal of Advanced Research in Computer Science and Software Engineering,2013.

[33] Comparative Analysis of Data Mining Techniques on Educational Dataset. Sumit Garg,Arvind K.Sharma. International Journal of Computer Applications, Volume 74– No.5, July 2013.

[34] FRANCISCA NONYELUM OGWUELEKA."Data mining application in credit card fraud detection system" . Journal of Engineering Science and Technology Vol. 6, No. 3 ,2011.

[35] Sen, Sanjay Kumar, and Sujata Dash. "Meta Learning Algorithms for Credit Card Fraud Detection." Meta 6.6 (2013): 16-20.

[36] Zareapoor, Masoumeh, and Pourya Shamsolmoali. "Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier." Procedia Computer Science, Pages 679-685(2015).

[37] Wu, Junjie, Hui Xiong, and Jian Chen. "COG: local decomposition for rare class analysis." Data Mining and Knowledge Discovery 20.2 (2010): 191-220.

[38] R.Devaki V.Kathiresan S.Gunasekaran,"Credit Card Fraud Detection using Time Series Analysis". International Journal of Computer Applications, 2014.

[39] Rodrigues, Luis Alexandre, and Nizam Omar. "Auto claim fraud detection using multi classifier system." Journal of Computer Science & Information Technology (2014).

[40] Chen, Suduan. "Detection of fraudulent financial statements using the hybrid data mining approach." SpringerPlus 5.1 (2016)