

## A survey review on financial fraud detection using data mining techniques

Shivam Patel\* and Trushil Patel

U. and P. U. Patel Department of Computer engineering, CSPIT, CHARUSAT, Changa 388421, Gujarat, India.

**Corresponding author:** \*Shivam Patel, U. and P. U. Patel Department of Computer engineering, CSPIT, CHARUSAT, Changa 388421, Gujarat, India.

### Abstract

Financial fraud is a global hazardous problem and has affected the economy around the world. Due to the dramatic increase of financial fraud which results in loss of billions of dollars worldwide each year, several modern techniques in detecting fraud are continually evolved and applied to many business fields. Data mining is one of the most effective and powerful tool for detecting financial fraud had been used widely by the business analysts, researchers and detectors. This survey paper formalizes different types of financial frauds, presents a comprehensive investigation on existing fraud detection systems and financial fraud detection practices using data mining methods.

**Keywords:** Financial fraud, Fraud detection, existing fraud detection systems, data mining techniques

### Introduction

Financial fraud is an issue that has wide ranging consequences in both the finance industry and day to day life. Fraud not only causes unimaginable financial losses but also pushes the organization backwards in this cut - throat competitive world. Billions of dollars are lost yearly due to financial fraud; [17] Bank of America, for example, agrees to pay \$16.5 billion for resolving financial fraud case. Financial fraud encompasses various types of crimes and unlawful activities such as identity theft, asset misappropriation and many more.[1] Today, telecommunication market all over the world is facing a severe loss of revenue due to fierce competition and loss of income due to fraud.

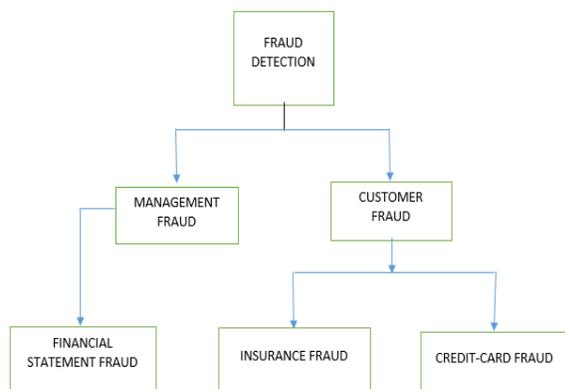
Traditional approaches of fraud detection relies on manual techniques such as auditing, which are inefficient and unreliable due to the complexities associated with the problem. Various data mining techniques are used in financial fraud detection which are efficient and reliable. Data mining being a 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 behaviour 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.

We present data mining techniques which are most appropriate for fraud analysis. We present automobile insurance example. 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.

Secondly, this paper presents a comprehensive analysis of data mining techniques used for detection of each type of financial fraud. Rest of the paper is covered the following topics: Different types of Financial Fraud, Existing fraud detection systems, different data mining techniques.

**Fraud detection**

Fraud may be defined as an intentional act in order to mislead or hoax or a person or an 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. In this paper, we had classified fraud into two categories namely management fraud and customer fraud where customer fraud is further categorised into two categories: insurance fraud and bank fraud.



**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. [15] Management fraud also known as financial statement fraud is a deliberate misstatement of material facts by the management in the books of accounts of a company which aims of deceiving the investors and creditors of the company. Thus, this illegitimate task performed by management has a severe negative impact on the economy of a company and ultimately deceiving the economy throughout the world as it significantly dampens the confidence of investors.

**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.

**Credit-card fraud:**

Revolution to a digital economy has lead to increasing usage of credit cards. According to RBI of INDIA, more than 6 crore of transactions worth ₹190989. That is roughly \$2964.974. 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, 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 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.

**Insurance fraud:**

An act performed by the insured person 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 it’s revenues to the frauds. Teris Roberts (MSF Access Campaign , Diagnostics

Advisor) has suggested an 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.

Insurance fraud detection is categorised into 4 different categories:

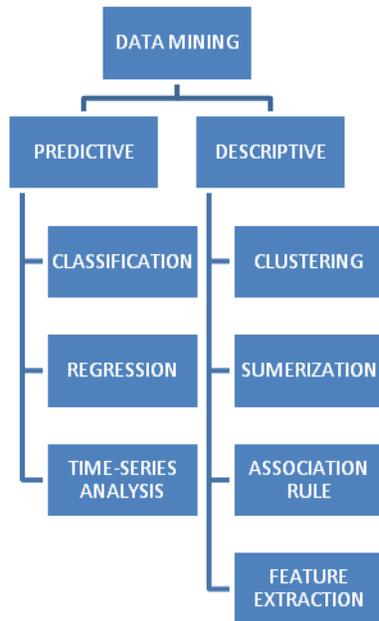
1. Home insurance fraud detection
2. Crop insurance fraud detection
3. Automobile insurance fraud detection
4. Health insurance fraud detection

**Existing fraud detection systems**

<b>SYSTEMS</b>	<b>RESULTS</b>
A fuzzy logic system	The result showed the chances of fraud and system predicted slightly better results than the auditors.
logic system	used Wang-Mendel algorithm to find out how health care providers committed fraud against insurance companies.
The EFD system	identified providers whose behaviour did not fit the rule.
The hot spots methodology	performed a process with 3 steps: a) k-means clustering ii) C4.5 algorithm the resulting decision tree can be converted to a rule set and iii) visualization tools for rule evaluation,
The credit fraud model	suggested a classification technique with fraud attribute and a clustering followed by a classification technique with no fraud attribute.
Kohonen's Self-Organizing Feature Map	used to categorize automobile injury claims depending on the size of fraud suspicion.
The distributed data mining model	The method was applied to credit card transactions.
The neural data mining approach	uses rule-based association rules to mine symbolic data and Radial Basis Function neural network to mine analog data.
SAS Enterprise Miner Software	depends on association rules, cluster detection and classification techniques to detect fraudulent claims.
ANN	The Bayesian Belief Network (BBN) and Artificial Neural Network (ANN) study used the STAGE algorithm for BBN in fraud detection and backpropagation for ANN. STAGE repeatedly alternates between two stages of search: running the original search method on objective function, and running hill-

	climbing to optimize the value function. The result shows that BBNs were much faster to train, but were slower when applied to new instances.
FraudFocus Software	automatically scores all claims. The scores are sorted in descending order of fraud potential and generate descriptive rules for fraudulent claims.
FairIsaac	recommended backpropagation neural networks for fraudulent credit card use.
The ASPECT group	focused on neural networks to train current user profiles and user profiles histories. A caller's current profile and the profile history are compared to find probable fraud.
Internal fraud detection by management	abnormal retail transactions by employees.
Home insurance, crop insurance, automobile insurance fraud detection, and health insurance A single metaclassifier	used to select the best base classifiers, and then combined with these base classifiers' predictions to improve cost savings (stacking bagging). Automobile insurance fraud detection data set was used to demonstrate the stacking-bagging problem.

**Different data mining techniques:**



Data mining tools take data and construct a representation of reality in the form of a model. The resulting model describes patterns and relationships present in the data.

**Predictive Data Mining Technique:**

The process of taking patterns discovered from the database and using them to predict the future. A predictive data-mining model predicts the value (numeric data/class label) of a specific attribute label using previously

known data. 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.

Classification models predict categorical class labels; and prediction models predict continuous valued functions. For example, we can build a classification model to categorize bank loan applications as either safe or risky, or a prediction model to predict the expenditures in dollars of potential customers on computer equipment given their income and occupation.

Regression is a data mining technique which comes under predictive data mining, predicts a number. Age, weight, distance, temperature, income, or sales could all be predicted using regression techniques. A regression task begins with a data set in which the target values are known.

#### **Time series analysis:**

A time series is a representation of collection of values obtained from sequential measurements over time. [15]Time-series data mining stops us from the desire to reify our natural visualization abilities, humans rely on numerous complicated schemes in order to perform task. We can actually derive a notion of shape and identify almost instantly the similarities between patterns on various time scales by avoiding focusing on small fluctuations. Major time-series-related tasks include query by content, anomaly detection, motif discovery, prediction, clustering, classification, and segmentation.

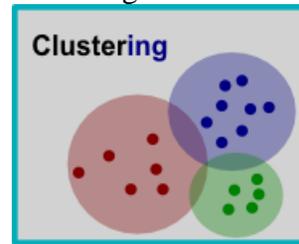
#### **Descriptive Data Mining Technique:**

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.

#### **Clustering:**

In the process of analyses of large data sets, one of the most undertaken concepts is the aggregation of similar objects within the dataset. Similar objects are clustered into with a dataset and thus the process is called clustering in data mining.



Summarization is a key data mining concept which involves techniques for finding a compact description of a dataset. Simple summarization methods such as tabulating the mean and standard deviations are often applied for data analysis, data visualization and automated report generation.

#### **Association rule:**

One of the reasons behind maintaining any database is to enable the user to find interesting patterns and trends in the data. For example, in a supermarket, the user can figure out which items are being sold most frequently. But this is not the only type of 'trend' which one can possibly think of. The goal of database mining is to automate this process of finding interesting patterns and trends. Once this information is available, we can perhaps get rid of the original database. The output of the data-mining process should be a "summary" of the database. This goal is difficult to achieve due to the vagueness associated with the term 'interesting'. The solution is to define various types of trends and to look for only those trends in the database. One such type constitutes the association rule.

In the rest of the discussion, we shall assume the supermarket example, where each record or tuple consists of the items of a single purchase. However the concepts are applicable in a large number of situations.

In the present context, an association rule tells us about the association between two or more items. For example: In 80% of the cases when people buy bread, they also buy milk. This tells us of the association between bread and milk. We represent it as - bread => milk | 80%

This should be read as - "Bread means or implies milk, 80% of the time." Here 80% is the "confidence factor" of the rule.

Feature extraction (FE) is one of the dimensionality reduction techniques.[14] FE extracts a subset of new features from the original feature set by means of some functional mapping keeping as much information in the data as possible. In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. The main goal of feature extraction is to obtain the most relevant information from the original data and represent that information in a lower dimensionality space.

### Conclusion

This paper begins with an overview of describing Financial Fraud and Concepts of Fraud Detection, followed by types of fraud detection including its two category: management fraud detection and customer fraud detection with further categorisation in them as financial statement fraud, credit-card fraud and insurance fraud and types of insurance fraud detection. We also studied the Existing Fraud Detection Systems. Data mining techniques are discussed in the later part of this paper where we studied different types of data mining techniques with detailed explanation of each technique which is useful in fraud detection.

### References

- [1]. Anita B. Desai et al, / (IJCSIT) "International Journal of Computer Science and Information Technologies", Vol. 4 (1),2013, 1 – 4
- [2]. 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.
- [3]. Rekha Bhowmik, "Data Mining Techniques in Fraud Detection" .Journal of Digital Forensics, Security and Law, Vol. 3(2)
- [4]. Barse, E., Kvarnstrom, H. & Jonsson, E. (2003). "Synthesizing Test Data for Fraud Detection Systems. Proc. of the 19th Annual Computer Security Applications Conference", 384-395.
- [5]. Bell, T. &Carcello, J. (2000). A Decision Aid for Assessing the Likelihood of Fraudulent Financial Reporting. Auditing: A Journal of Practice and Theory 10(1): 271-309.
- [6]. Belhadji, E., Dionne, G. &Tarkhani, F. (2000)." A Model for the Detection of Insurance Fraud." The Geneva Papers on Risk and Insurance 25(4): 517-538
- [7]. M. Cecchini, H. Aytug, G.J. Koehler. "Detecting Management Fraud in Public Companies.", Management Science, Volume 56, No. 10, 2010, pp. 1146 – 1160.
- [8]. 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.
- [9]. Pankaj Richhariya, Dr. Prashant K Singh, EnduDuneja,"A Survey on Financial Fraud Detection Methodologies" IRACST – International Journal of Commerce, Business and Management (IJCMB), Vol. 1, No.1, 2012

- [10]. Massey, K. Massey, Combating eFraud – a next generation approach, Financial Insights White Paper. 2005
- [11]. Major, J. & Riedinger, D. EFD: A Hybrid Knowledge/ Statistical-based system for the Detection of Fraud. Journal of Risk and Insurance 69(3): 309-324. 2002.
- [12]. Pathak, J., Vidyarthi, N. & Summers, S. A Fuzzy-base Algorithm for Auditors to Detect Element of Fraud in Settled Insurance Claims, Odette School of Business Administration. 2003.
- [13]. M. Pechenizkiy, S. Puuronen, A. Tsymbal “Feature Extraction For Classification In Data Mining Process” in International Journal "Information Theories & Applications" Vol.10
- [14]. Aastha Bhardwaj, Rajan Gupt. “Financial Frauds: Data Mining based Detection – A Comprehensive Survey” International Journal of Computer Applications (0975 – 8887) Volume 156 – No 10, December 2016
- [15]. Philippe Esling and Carlos Agon, Institut de Rechercheet Coordination Acoustique/Musique (IRCAM)
- [16]. Fox News. Bank of America pays \$16.5 bn to settle financial fraud case. Retrieved from Fox News Latino: <http://latino.foxnews.com/latino/news/2014/08/21/bank-america-pays165-bn-to-settle-financial-fraud-case/>, 2014.