A complete literature review on financial fraud detection applying data mining techniques

Subhas Barman*, Utpal Pal, Md. Asif Sarfaraj, Biswajit Biswas, Animesh Mahata and Palash Mandal

Jalpaiguri Government Engineering College,
Jalpaiguri-735102, West Bengal, India
Email: subhas.barman@gmail.com
Email: log4utpal2@gmail.com
Email: md.a.sarfaraj120@gmail.com
Email: biswajitbiswas957@gmail.com
Email: m.animesh93@gmail.com
Email: palashmandal2294@gmail.com
*Corresponding author

Abstract: Financial fraud is defined as unlawful or criminal duplicity attempted to result to organisational or personal gain. It is a big threat to the economics of a firm, corporate sector, government or ordinary customers in the form of credit card fraud, insurance fraud, and financial statement fraud. Several approaches exist in the literature of financial fraud detection. But, due to inefficiency of those approaches, researchers leverage data mining techniques to detect financial fraud. This paper aims to build a systematic academic review of financial fraud detection approaches based on data mining techniques in the recent years. In the practice, different data mining techniques namely: K-nearest neighbour, decision tree, fuzzy logic, logistic model, Bayesian belief network, naïve Bayes, Beneish M-Score model, Benford's law, Altman Z-score have been applied to improve accuracy of fraud detection. In this paper, existing financial fraud detection techniques are compared with the advantages and limitations of the techniques.

Keywords: financial fraud; financial fraud detection; money laundering; data mining technique.

Reference to this paper should be made as follows: Barman, S., Pal, U., Sarfaraj, M.A., Biswas, B., Mahata, A. and Mandal, P. (2016) 'A complete literature review on financial fraud detection applying data mining techniques', *Int. J. Trust Management in Computing and Communications*, Vol. 3, No. 4, pp.336–359.

Biographical notes: Subhas Barman received his BTech in Computer Science and Engineering Department from Kalyani University, West Bengal, India and MTech in Information Technology from Indian Institute of Technology, Kharagpur. He currently is an Assistant Professor in the Department of Computer Science and Engineering, Jalpaiguri Government Engineering College, Jalpaiguri, West Bengal. His current research includes information and network security, data mining and financial frauds.

Utpal Pal is a final semester student of BTech in Computer Science and Engineering at Jalpaiguri Government Engineering College, West Bengal, India. His current research includes financial fraud detection and data mining techniques.

Md. Asif Sarfaraj is a final semester student of BTech in Computer Science and Engineering at Jalpaiguri Government Engineering College, West Bengal, India. His current research includes financial fraud detection and data mining techniques.

Biswajit Biswas is a final semester student of BTech in Computer Science and Engineering at Jalpaiguri Government Engineering College, West Bengal, India. His current research includes financial fraud detection and data mining techniques.

Animesh Mahata is a final semester student of BTech in Computer Science and Engineering at Jalpaiguri Government Engineering College, West Bengal, India. His current research includes financial fraud detection and data mining techniques.

Palash Mandal is a final semester student of BTech in Computer Science and Engineering at Jalpaiguri Government Engineering College, West Bengal, India. His current research includes financial fraud detection and data mining techniques.

1 Introduction

Fraud is a crime, and is also a civil law violation. In recent year financial fraud, including credit card fraud, bank fraud, insurance fraud, financial statement fraud, corporate fraud, security fraud, money laundering have attracted a great deal of concern and attention. According to Oxford Dictionary, fraud is a wrongful or criminal deception intended to result in financial or personal gain. Fraud is a cautious act that mock the law or policy of an organisation which earns the fraudsters an unauthorised financial benefit (Lisic et al., 2015). Fraud may affect day to day living cost of peoples, as well can reduce assurance in the industry and can thwart economics (West and Bhattacharya, 2016).

Financial fraud is becoming a serious problem in the economic society. It makes a horrific threat to the economic condition of an organisation or even the government. According to Forbes magazine, The MF Global, a brokerage firm, led by former Goldman Sachs Chairman and former New Jersey Senator then Governor Jon Corzine, had $41 billion in assets before failing in October 2011 (Yang, 2014). There are several other cases which cost a huge amount of damage to the economy. For example, investment bank Lehman, with $600 billion in assets, failed in late 2008. It was the largest bankruptcy in history and a spark to the worldwide financial crisis. New York money Manager Bernard Madoff's $65 billion Ponzi scheme was the largest fraud ever by an individual. It was exposed in December 2008 when Madoff, now doing 150 years in prison and $170 billion in restitution, confessed to his sins. Credit card fraud alone accounts for billions of dollars of lost revenue each year (Bhattacharyya et al., 2011).

According to a BBC news reported in 2007, the amount of fraudulent insurance claim is approximately 1.6 billion pounds a year in UK.

Financial frauds can be classified into many categories. Among them, credit card fraud, insurance fraud, financial statement fraud, money laundering, mortgage fraud, security and commodities fraud etc. reflect greater effect on economy. A large number of research paper has been published on credit card fraud, insurance fraud, financial statement fraud due to their relative large impact on world economy. Recently, research on money laundering is also attracting a great amount of interest.

Traditional approaches for detecting frauds became inefficient due to introducing of new methods of committing fraud by fraudsters. Data mining can be applied to this problem because techniques used in data mining-based approaches can detect small deviation in large datasets (Ngai et al., 2011). There are varieties of frauds committed and also there exist a large number of data mining techniques which can be used for detecting frauds, and continuous research is going on to find the best solution for a particular problem. An efficient approach to finding the solution of financial fraud problem is to applying the data mining techniques for classifying suspicious transaction, which are further investigated for unfitting transaction category or fraud (Ngai et al., 2011). Data mining can be applied to the early level of fraud detection in most cases and the complex ones can be reviewed manually.

In general, there are six data mining classes: clustering, classification, prediction, outlier detection, regression, visualisation, which are commonly used to detect financial frauds. Several techniques have been evolved surrounding these six classes, among them neural network, logistic model, decision tree, naïve Bayes, genetic programming, support vector machine produces excellent result for detecting financial frauds.

Practically, there are two categories of data mining approaches, one is statistical and another is computational. Statistical methods are those that involves mathematical computations such as naïve Bayes theorem and logistic regression. On the other side, computational methods involve modern techniques such as neural network and decision trees. Both of these approaches are useful for financial fraud detection. The statistical methods are more adaptive to problems, whereas, computational methods are quite inflexible to few problems in financial fraud domain.

The main objective of this paper is to present a review of existing research trends in financial fraud detection and categorise them according to their performance, fraud type and data mining technique used for fraud detection. Our survey includes a comparison between different fraud detection techniques and a summary of performance on the basis of accuracy of fraud detection techniques.

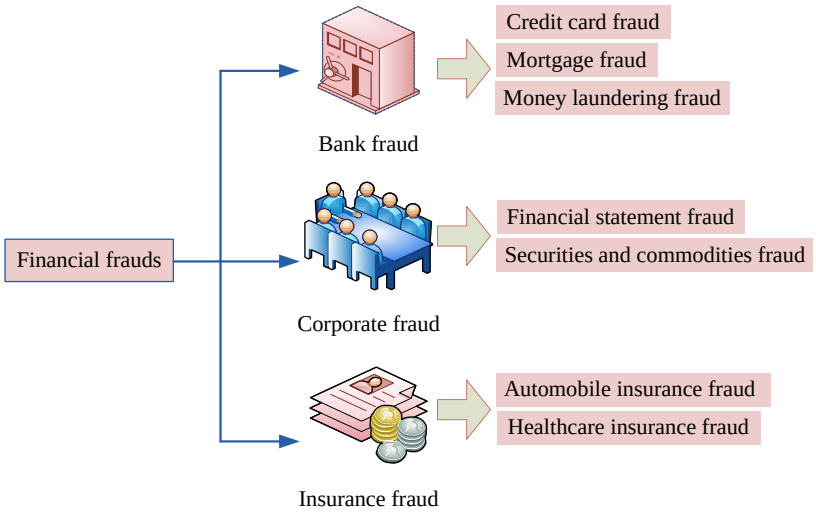
The remaining part of this paper is as follows: Section 2 describes different types of financial frauds. Different data mining techniques used in financial fraud detection are discussed in Section 3. Section 4 includes the detailed survey on financial fraud detection and the challenges in financial fraud detection are stated in Section 5. Finally, the survey is concluded in Section 6.

2 Types of financial frauds

There exist several types of financial frauds. But in terms of damage made by these frauds to the economy, few major names are credit card, financial statement fraud,

insurance fraud, money laundering etc. A classification is provided in Figure 1. They are briefly described below.

Figure 1 Different types of financial frauds (see online version for colours)



2.1 Credit card fraud

Credit card fraud generally comes under bank fraud category. Generally, unauthorised access and use of a person's credit card without his consent is known as credit card fraud. According to Duman and Ozcelik (2011), credit card frauds are of two types. In the first one, the fraud is pulled off by some group of fraudsters. This type of credit card fraud is known as counterfeit fraud. Effect of this action is huge, which can affect ten to hundreds of cardholders of a bank. The second type of fraud involves illegal use of a stolen or lost card. This type of fraud does not affect a mass of cardholders rather on one or two cardholders at a time. There are several other ways in which a fraudsters can use a card illegally (Duman and Ozcelik, 2011). Phishing is a practice where a fraudsters represents himself as a financial officer to obtain the card details of a person. Often they use a similar device to swipe machine or an interface of an ATM which can directly collect all the details about the card (Bhattacharyya et al., 2011). Recently growing online activity has also given a rise to credit card fraud. An easy approach can be taken for detecting credit card fraud is to analyse a customer's usual spending structure and point out the anomalies (Duman and Ozcelik, 2011).

2.2 Financial statement fraud

Financial statement are the collections of reports released by an organisation about its financial results, financial condition, cash flows, and their expenses, loans, profits etc. along with some comments and proposition on business and expected future issues on business, by the management (Ravisankar et al., 2011; Glancy and Yadav, 2011). Financial statements are basically the certificates of an organisation which decides, whether it is running smoothly or in crises. It also helps the investor and bank officials

to make decision whether it is profitable or not to invest in the organisation. The stock holders can review the company's financial future based on the financial statement (Ravisankar et al., 2011). Financial statement fraud often termed as corporate fraud.

Financial statement fraud, involves reshaping the original financial statement to make the company look more bankable. The auditors are responsible for committing this types of fraud, which often occurs due to managerial pressure. The main reason behind financial statement fraud includes escalating stock price, to get bank loan, attracting investors etc. Another reason may be to avoid tax payment. It is very common nowadays and increasing in number of making fraudulent statement to cover up original financial condition of organisation and gain more profit (Ravisankar et al., 2011). Financial statement fraud is very hard to diagnose, as it is usually performed by experienced people.

2.3 Insurance fraud

Insurance fraud is any act committed with the wrongful intention to attain a fraudulent payback from an insurance company. This type of fraud can be committed at any stage during the insurance process (e.g. application, rating, eligibility, billing, claims etc.) and can be carried out by personnel in that chain like consumer agents, brokers, insurance company employees etc. (Subelj et al., 2011). Insurance frauds occur in the companies of automobile insurance, healthcare insurance, crop insurance etc. (Ngai et al., 2011).

In the process of a fraudulent automobile claim, customer may submit reports of a fabricated injury, or took accident or may produce an excessive repair cost. Unnecessary medical services, duplicate claims are generally the way of the committing health care insurance fraud. In crop insurance fraud, the customer amplifies their losses, due to natural disaster or declining agriculture prices (Ngai et al., 2011).

2.4 Mortgage fraud

Mortgage fraud is a type of fraud in which the intent is to materially misrepresent or omit information on a mortgage loan application to obtain a loan or to obtain a financial advantage that would have not obtained it the borrower known the truth. This fraud is committed through manipulation of a property or mortgage documents. The fraudulent document may cover up the original value of property for the purpose of the gaining higher amount of loan from lender (Ngai et al., 2011).

2.5 Money laundering

Money laundering is the form of financial fraud where fraudster tries to disguise the illegal or dirty money and makes them clean or lawful money. Money laundering is an act where people deliberately engage in financial transactions with the revenue earned by some improper or illegal exercise to cover up the nature of revenue or the source to make the property appear legitimate (Genzman, 1997).

2.6 Securities and commodities

Securities and commodities frauds may happen in the form of Ponzi scheme, late-day trading, market manipulation, advanced fee fraud, pyramid schemes, foreign currency exchange, broker embezzlement etc. (Ngai et al., 2011).

3 Data mining techniques used for financial fraud detection

The application of data mining techniques in detection of financial fraud merely based on six data mining classes, and several techniques have been introduced to support for building these classes.

3.1 Data mining classes

In data mining technique, we need to analyse a large set of data and make a suitable point of view where we can easily find an anomaly or any distortion which can be useful. Data mining classes are these foundations of view point. Six data mining classes are found to be useful in case of financial fraud detection problems.

3.1.1 Clustering

Cluster analysis or clustering is a task, in which a set of objects, that are grouped in such a way that objects in the same group (known as cluster) are more similar (in some sense) to each other than to those in other groups (clusters) (Han and Kamber, 2006). It mainly performs task of analysing in data mining. Clustering is also known as unsupervised learning used for data segmentation or partitioning. Clustering is mainly done on multivariate dataset which results into groups of data where points in one group is similar to each other but separate from each other from points from other groups (Yue et al., 2007). Few frequently used clustering techniques are the K-nearest neighbour, the naïve Bayes technique and self-organising map techniques etc.

3.1.2 Classification

Classification model can predict categorical class labels of unknown objects to differentiate between objects of different classes. By classification process we can identify a set of common features and models that describe and differentiate data classes (Zhang and Zhou, 2004). Few frequently used classification techniques are neural networks, the naïve Bayes method, decision trees and support vector machines etc. This classification classes are used to detect the credit card fraud, healthcare and automobile insurance fraud, and corporate fraud, among other types of fraud. It is one of the most important approaches in terms of application of data mining in financial fraud detection (FFD).

3.1.3 Prediction

Prediction model can predict continuous valued functions, i.e., predicts unknown or missing values. For predicting, the characteristics of the object for which the values

are being predicted, needs to be continuous rather than categorical or discrete valued (Han and Kamber, 2006). This attribute is mainly denoted as the predicted attribute. Commonly used prediction techniques are neural networks and logistic model prediction etc.

3.1.4 Outlier detection

An outlier is an observation which misguides so much from the other observations as to arouse suspicions that it was produced by a separate mechanism. According to Agyemang et al. (2005), when characteristics of a data appear different from other remaining set of data in the same population then the data is known as outliers. Most frequently used outlier detection class is the discounting learning algorithm.

3.1.5 Regression

Regression is a data mining function that can predict a number. Profit, sales, mortgage values, house costs, square footage, temperature, or distance etc. could all be easily predicted by using regression techniques. For example, a regression model can be used to predict the cost value of a house, depending upon some attributes like location, area size, number of rooms, and other useful factors. A regression task starts with a dataset, in which the target values are known, it means target values will be provided. Regression is a statistical method which is mainly used to explain the relationship among the discrete valued dependent or independent variables (Han and Kamber, 2006). The regression model is mainly undertaken using some mathematical methods and it is basically used for detection of credit card fraud, crop and auto-mobile insurance fraud, and corporate fraud. Commonly, there are three types of regression models: linear, polynomial, and logistic regression.

3.1.6 Visualisation

Visualisation is a very common and popular term that describes any effort to assist people makes out the significance of data, by placing it, in a visual context. According to Shaw, Subramaniam and Welge, visualisation is the easily acceptable and understandable presentation of data. It includes the methodology to extract clear patterns and relationship between data from complex data characteristics to make it easy to understand for users using data mining process (Shaw et al., 2001). Visualising is a model that should allow a user to discuss and briefly explain the logic behind the model.

3.2 Data mining techniques

The above six classes are supported by a large number of data mining techniques for each class (see Figure 2). These techniques are used to formulate the applicable relationships in the data. A data mining based framework for financial fraud detection process is given in Figure 3.

Figure 2 Data mining classes and techniques (see online version for colours)

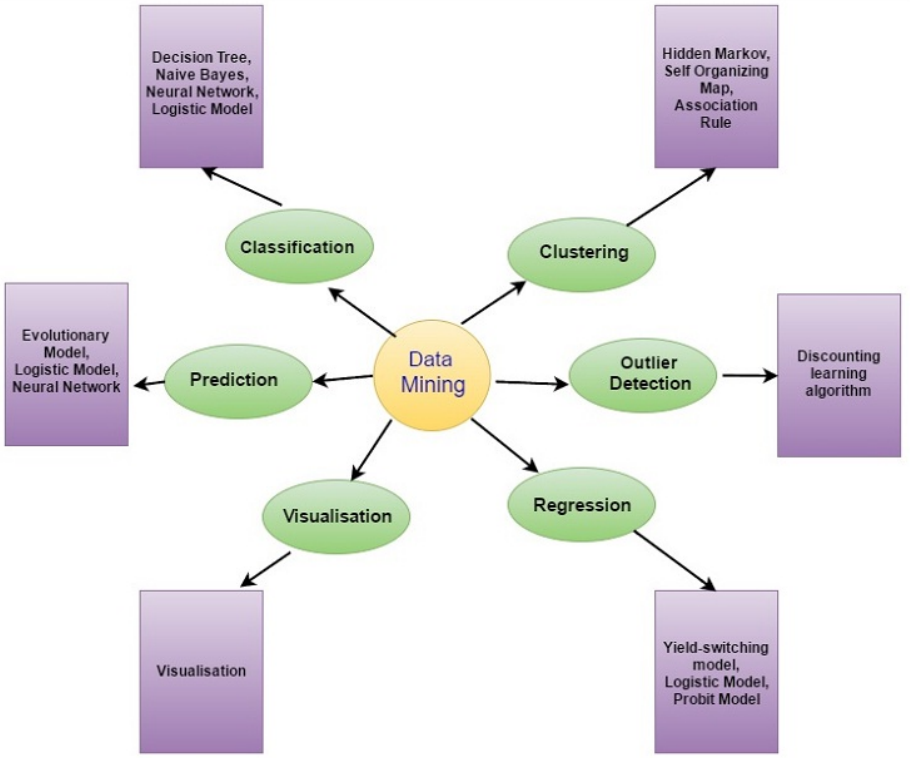
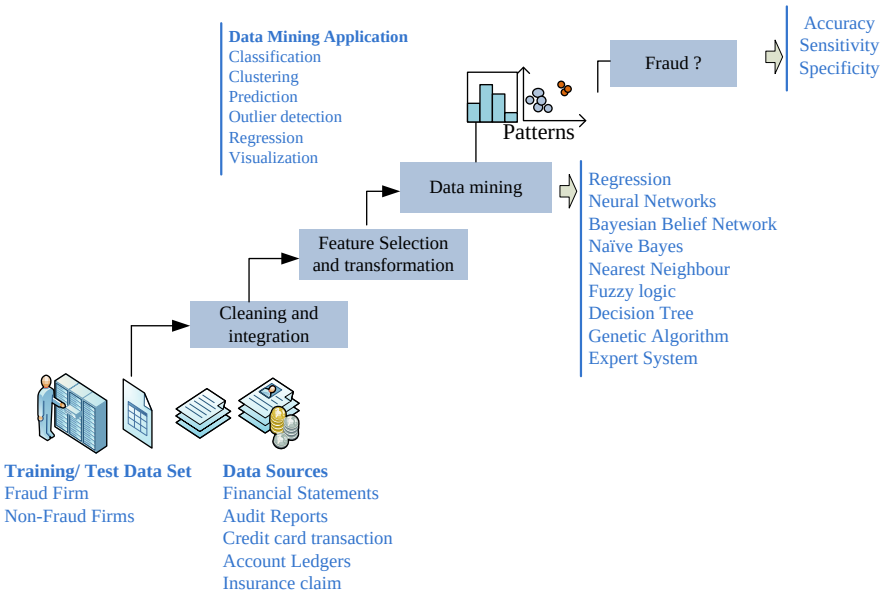


Figure 3 Data mining-based framework for financial fraud detection (see online version for colours)



In this framework, source data (either training dataset or test dataset) is extracted from different homogeneous or heterogeneous sources of financial data. The framework uses pre-processing steps like cleaning and integration to make the source data suitable for the system. The features, on the basis of what the classification of the financial data is to be carried out, are selected and transformed in input of a uniform format. The transformed inputs are processed and analysed with the data mining techniques. Data mining techniques are applied to the inputs/ features to make the classification of the training data. Now, test datasets are verified and evaluated the result on the basis of accuracy, sensitivity and specificity.

Each technique works differently with the data. Probabilistic neural network is used to solve classification and mapping of data, whereas multi-layered feed forward neural network begins as a network of nodes, arranged in three layers-input, hidden as nodes to buffer input and output for the model respectively. A logistic model is a linear model which is mainly used for binomial regression, where the variables that are predictor can be either numerical or categorical. This model is mainly and vastly used for solving problems related to auto-mobile insurance and corporate fraud. The naïve Bayesian classifiers assume that the impact of an attribute value on a given class is independent of the values of the other attributes. This assumption is known as conditional independence. In case of support vector machine, it implements nonlinear class boundaries by mapping input vectors nonlinear way into high dimensional space. Genetic programming is mainly a searching methodology, which is belonged to the class of evolutionary computation. It can randomly produce initial population. Then the initial population is manipulated by using various genetic operations for generating new population. K-means is a clustering algorithm, it tries to partition n object or n point into k clusters according to their similarity in which each object or each point belongs to the cluster with the nearest mean and the mean value of the object in a cluster can be declared as the centroid of the cluster. The objective of K-means algorithm is to minimise total intra-cluster variance or squared error function.

The working principle of decision tree and fuzzy logic has been discussed with German dataset below.

3.2.1 Decision tree

A decision tree (Pal and Paul, 2003) is a classifier that is used to construct a tree structured decision support tool. The decision tree consists of nodes where each node represents a test on an attribute and each branch represents possible consequences. In this way, the decision tree model is used to divide the observation into mutually exclusive subgroups. The decision tree is a predictive model that is used to create mapping from observations to possible outcomes where predictions are represented by leaves and the combinations of features by branches. Decision trees are generally used in credit card, auto-mobile insurance and corporate fraud. Though this technique is useful but there are many techniques which can be used for providing a better framework. Let's see an example. In our example, we construct a decision tree (i.e., Figure 4) using few attributes (e.g., checking status, foreign worker, duration and own telephone) of German dataset. The sample data values of the selected attributes of German dataset are shown in Table 1.

Figure 4 Decision tree (see online version for colours)

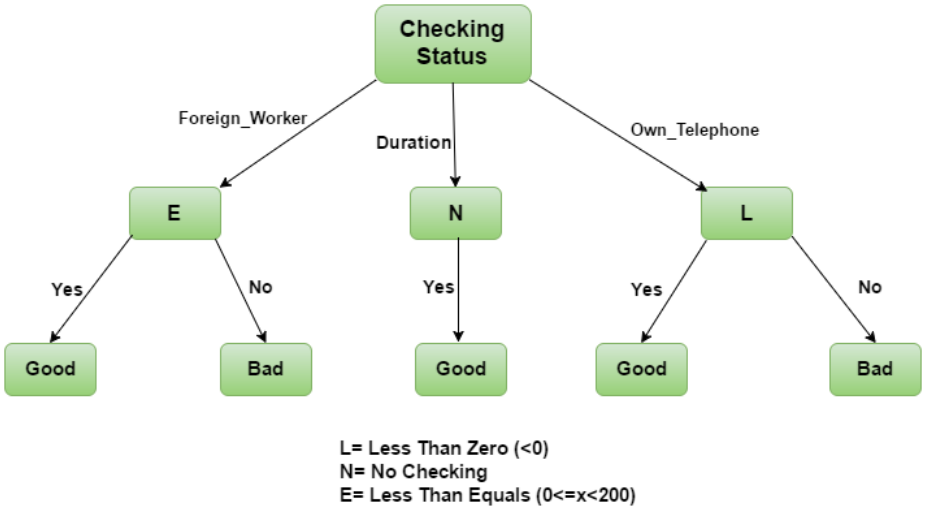


Table 1 German data

Checking status	*Duration*	. . . . .	*Own telephone*	*Foreign worker*	*Class*
0<=X<200	up_2_years	. . . . .	None	Yes	Bad
0<=X<200	lo_1_year	. . . . .	None	Yes	Bad
<0	up_2_years	. . . . .	None	Yes	Bad
0<=X<200	lo_1_year	. . . . .	Yes	Yes	Good
<0	1_2_years	. . . . .	None	Yes	Bad
<0	1_2_years	. . . . .	None	Yes	Good
<0	1_2_years	. . . . .	None	No	Bad
No checking	1_2_years	. . . . .	None	Yes	Good
<0	up_2_years	. . . . .	None	No	Good
0<=X<200	1_2_years	. . . . .	Yes	Yes	Bad
No checking	1_2_years	. . . . .	Yes	Yes	Good
No checking	lo_1_year	. . . . .	Yes	Yes	Good
<0	lo_1_year	. . . . .	None	Yes	Good
<0	lo_1_year	. . . . .	None	No	Good

We analysed the data to classify that a customer is belonging in good or bad class with the help of the decision tree. It means, when a customer applies for a bank loan, the bank has to check the customer's profile with respect to decision rules produced from the decision tree and make a decision accordingly. As per the dataset, two type of decisions, good or bad, can be occurred.

- if the customer's profile is good then the customer can easily repay the loan, then not providing him a loan, is loss of business to the bank
- if the customer's profile is bad then it is to be risky to the bank to provide him a loan.

INPUT: In data partitions, D, that is a set of training tuples and with their associated class labels, also attribute set, the set of candidate attributes. Attribute selection technique is a method to determine the partitioning criterion that 'best' partitions the data tuples into separate classes. This criterion contains of a partitioning attribute and most probable, either a partition point or splitting subset.

OUTPUT: A decision tree, A is separate valued; in this case the result of the test at node N assembles directly to the familiar values of A. A branch is produced for each known value, a_j, of A and marked or labelled with that value partition D_j is the subset in D which has value a_j of A. There is same value of A for all of the tuples in a provided partition. Then A may not be considered to use in any future partitioning of tuples. So that's why, it is displaced from attribute list or set. A is continuous valued: for this case, the test of node N has two possible results, according to conditions, $A \leq$ split point and $A >$ split point, respectively. Three possibilities for partitioning tuples based on this partitioning method, shown with instance.

Let assume A be the splitting attributes

a if A is discrete valued, then one branch is gone up for each known value of A

b if A is continuous valued, the two branches are increased, according to $A \leq$ splitting point and $A >$ splitting point

c if A is a discrete valued and a binary tree must be formed the test will be in the form of $A \in SA$, where SA is the partitioning subset for A.

3.2.2 Fuzzy logic

The main disadvantage of rule-based classification is the involvement of sharp cut off for continuous distribution. For example, in customer credit application, this rule essentially says that for customer who have had a job for two or more years and who have a high income (i.e., of at least \$50,000) are approved.

$$IF(year_{employed} \geq 2)\ AND\ (income \geq \$50k)\ THEN\ credit = approved \quad (1)$$

By the rule [equation (1)], a customer who has had a job for at least two years, only they will receive credit if his or her income is say, \$50,000 but not if it is \$49,000. Harsh thresholding may be seemed unfair. Here, we categorised income into some categories such as low-income, medium-income, high-income. Now, we can apply fuzzy logic for fuzzy thresholded boundaries for defining each category from Figure 5. Here, fuzzy logic uses truth values between 0.0 and 0.1 for representing the degree of membership that certain value is provided in a category. Each category represents a fuzzy set. With fuzzy logic, we can describe that an income of \$49,000 is more or less, high, although not as high as in income of \$50,000. Fuzzy logic gives us graphical tools for helping users for the purpose of converting attribute values to fuzzy truth values.

Fuzzy set theory is also known as possibility theory. It is an alternative of traditional two value logic which is probability theory. It deals with improper measurement of date and also fuzzy set theory permits us for dealing with vague facts, means that are inexact for this example, the income value \$49,000 belongs to 60th medium and high fuzzy sets, but to separate degrees using fuzzy logic notation.

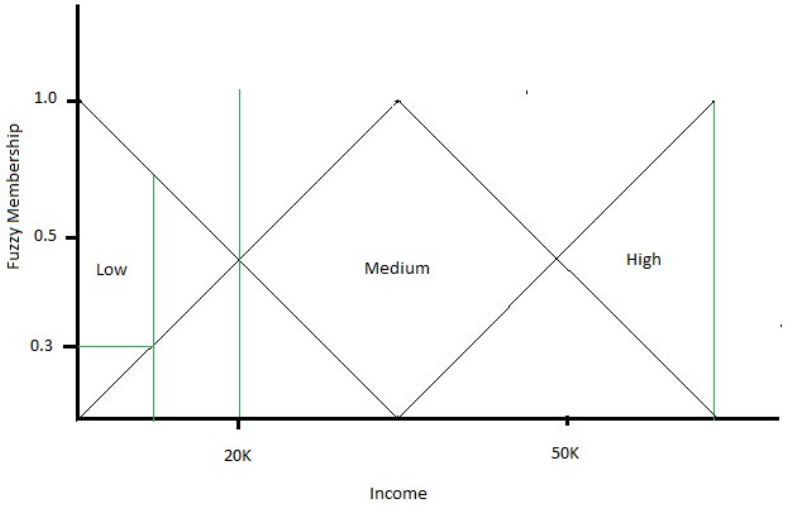
$m_{medium-income}$ (\$49k) = 0.15 and $m_{high-income}$ (\$49k) = 0.96

Whereas m denotes the membership function, which is operating on the fuzzy sets of medium and high-income, respectively.

Here fuzzy truth values for income defined as the degree of membership of income values represent three classes that are low, medium and high. Each classes acts as a fuzzy set. Also a given income value x can have membership in more than one fuzzy set. The membership value of x in each fuzzy set does not have to total to 1.

The fuzzy logic has been used in vast area for classification including market research, finance, health care and environmental engineering.

Figure 5 Fuzzy logic (see online version for colours)



4 Existing work on financial fraud detection

To find out the current trend and method used in financial fraud detection, we classify our survey in three categories. First, we discuss our observation regarding the existing work on financial fraud detection with respect to fraud types. Secondly, we present the existing work as reported in the literature of financial fraud detection with respect to the data mining techniques used for detecting frauds. Thirdly, we study the performance of the existing work as reported by the researchers in the literature of financial fraud detection using data mining techniques. This section describes these classifications in details.

4.1 Classification based on financial fraud types

It is being noticed that the frauds committed in past are deviating in nature for each type. So, detecting problem domain has become difficult based upon past detection. By classifying the types of fraud which has been detected, we can analyse the techniques which are more convenient and generally used in most cases of detecting a single type

of fraud. In addition, we can draw the general streamline of investigation based on the scope and success rate of common practices.

4.1.1 Existing work on credit card fraud detection

For every technique applied in different fraud types, the feature selection will vary based on the problem domain. In case of credit card fraud, researchers typically select independent variables or sometimes aggregate values. There are two different types of approach has been noticed. First one is to configuring the data mining technique with respect to frauds trends and the other one is to configure the technique with respect to transaction behaviour of customer. For example, Duman and Ozcelik (2011) used customer behaviour features like transaction statistics, regional transaction statistics, daily transaction amount statistics, daily number of transaction statistics.

In contrast, Bhattacharyya et al. (2011) used independent values such as account number, date of transaction, transaction currency and aggregated value like transaction amount per day along with transaction amount at a single point of service (POS).

All the techniques used to detect credit card fraud become successful in term of accuracy including scatter search, genetic algorithm, support vector machine self-organisation maps, artificial immune system, decision tree, bayesian belief network, neural network etc. Yeh and Lien (2007) used six data mining techniques such as discriminant analysis, logistic regression, Bayes classifier, Nearest neighbour, Artificial neural network and classification trees. According to Yeh and Lien (2007), artificial neural network is the only one that can accurately estimate the real probability of fraud. In another research, Duman and Ozcelik (2011) used both genetic algorithm and scatter search (GASS) in a combined and modified manner on a 1,050 numbers of fraud transactions along with some legitimate transactions, it produced nearly accurate result for all the transactions.

4.1.2 Existing work on financial statement fraud detection

In case of financial statement fraud, there are some specific types of frauds committed for an individual company. According to Kirkos et al. (2007) financial statement fraud costs us business around $400 billion annually. Kirkos et al. used financial statement of Greek manufacturing firms as dataset and reported 73.6% accuracy for fraud detection. Many researchers select various features to detect the financial statement frauds. For example, Koh and Low (2004) used some suitable ratios such as total income to total assets, net interest payment to gross before tax and interests payments and market value to assets. Some other features which are useful for the detection of fraud include gross margin index, assets quality index, sales growth index, depreciation growth index, primary business income, amounts receivable, net profit, loan index etc.

Due to varying nature of financial statement fraud, many researchers applied different techniques to carry out fraudulent statement. Ravisankar et al. used few data mining techniques on financial statement of 202 companies listed on Chinese stock exchanges to detect financial statement fraud. Ravisankar et al. (2011) reported that among the data mining techniques, probabilistic neural network detected highest accuracy of 98.09%. Aris et al. (2013) compared Benford's law vs Beneish model on same data and both techniques be almost equally useful. In another research on Enron

company's Beneish M-score model and Altman's Z-score model proved to be accurate finding the fraud as early as possible (Mahama, 2015).

4.1.3 Existing work on insurance fraud detection

Insurance fraud has vast effect in economy. Many researchers report different techniques to detect insurance frauds (Viaene et al., 2007; Bermudez et al., 2008). As the insurance frauds are of different types and it may be taken place at any point of the insurance process, the feature selection for this type fraud detection becomes difficult. But, in many cases analysing the network between insurance policy holder, the broker, and insurance company officials produced fruitful result.

Although, lack of research in insurance fraud domain, it has been noticed that social network analysis (Subelj et al., 2011) is a handful technique to detect auto-mobile insurance fraud. In some cases, modified Bayesian belief network and logistic regression can detect insurance fraud to some degree (Bermudez et al., 2008). Note that Viaene et al. (2007) and Bermudez et al. (2008), both approaches used audited motor insurance claims from Spanish insurance companies as the dataset.

4.1.4 Existing work on money laundering detection

A very few research has been conducted on money laundering detection. According to Gao and Ye (2007), outlier detection and money laundering network analysis may help in anti-money laundering research.

4.2 Classification based on fraud detection techniques used

To understand each type of fraud to be detected we must look into data mining techniques and classify according to each technique used to detect a particular fraud. Classification on the basis of techniques used on fraud detection is helpful to understand the importance of each technique for a given problem. This section describes the classification framework. On the other hand, we can easily place any gaps in research by looking at algorithms which have been investigated partially. We classify the researches on financial fraud detection based on the techniques used and the classification is shown in Table 4.

Early fraud detection research basically focused on statistical models and neural networks. In past years, fraud detection was mainly based on statistical methods and neural networks, however these methods are still used in some certain data mining problems for fraud detection. Many of the researchers used neural network to investigate financial statement fraud (Ravisankar et al., 2011; Bose and Wang, 2007), credit card fraud (Dorronsoro et al., 1997; Yeh and Lien, 2007); audited motor insurance claims (Viaene et al., 2007). There are some other work where logistic regression-based approaches are reported for credit card fraud detection (Bhattacharyya et al., 2011), insurance frauds modelling (Artis et al., 1999), insurance frauds detection (Bermudez et al., 2008) and financial statement fraud detection (Huang et al., 2014). Similarly, Bayesian belief network is also used to detect financial statement fraud detection (Kirkos et al., 2007), credit card fraud detection (Bhattacharyya et al., 2011; Chan et al., 1999). Compared to these algorithms, use of other techniques in fraud detection research are

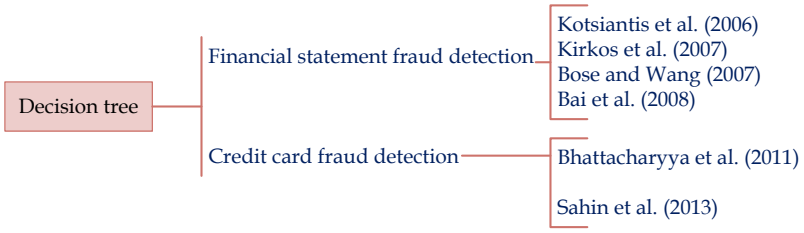
limited in number. Though decision trees (Chan et al., 1999; Yeh and Lien, 2007; Viaene et al., 2007; Kirkos et al., 2007) and naïve Bayes (Yeh and Lien, 2007; Viaene et al., 2007; Holton, 2009) are getting some attention among the researchers.

Table 2 Classification based on type of fraud

Fraud type	*Techniques used*	*Reference*
Credit card fraud	Discriminant analysis, logistic regression, Bayes classifier, nearest neighbour, artificial neural network, classification tree	Yeh and Lien (2007)
	Genetic algorithm, scatter search	Duman and Ozcelik (2011)
	Artificial immune system	Gadi et al. (2008)
	Ada cost algorithm (a variant of Ada boost algorithm)	(Chan et al., 1999)
	Neural network	Dorronsoro et al. (1997)
	Hidden Markov model	Srivastava et al. (2008)
	Self-organising map	Zaslavsky and Strizhak (2006
Financial statement farud	Neural network, Genetic programming, group method of data handling, logistic regression	Ravisankar et al. (2011)
	Beneish M-score model, Benford's law	Aris et al. (2013)
	Beneish M-score model, Altman's Z-score model	Muntari Mahama
	Neural network, Bayesian belief network, decision tree	Kirkos et al. (2007)
	CART	Bai et al. (2008)
Insurance fraud	Social network analysis	Subelj et al. (2011)
	Bayesian belief network, logistic regression	Bermudez et al. (2008)
	Fuzzy logic	Pathak et al. (2005)
Money laundering	Network analysis	Gao and Ye (2007)

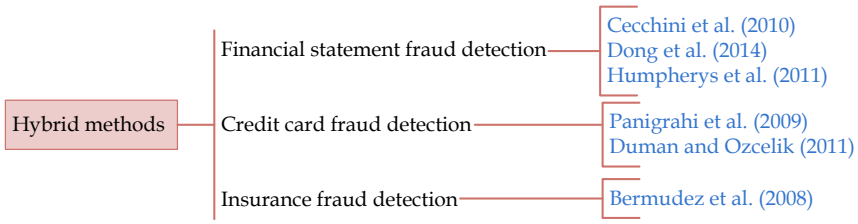
Kirkos et al. (2007) studied financial statement fraud detection using decision tree of the financial records from selected Greek manufacturing firms. According to the classification rules derived from the decision tree, Kirkos et al. (2007) reported 73.6% accuracy of fraud detection. Similarly, Humphreys et al. (2011) investigated financial frauds from managerial statement of US companies using hybrid approach of decision tree alongwith textmining and reported an accuracy of 67.3%. In Viaene et al. (2007), decision tree is used to investigate frauds in auto-mobile insurance claims of Spanish insurance companies. Application of decision trees to investigate different financial frauds by different researchers are shown in Figure 6.

Figure 6 Different application of decision tree for financial fraud detection (see online version for colours)



Now according to popularity of neural networks and logistic regression, these are basically often used to detect fraud, also used when other techniques already tested on that database. In other researcher's attention on a single form of fraud detection which they pleader above others, look like studying text mining with the singular validation decomposition vector (Glancy and Yadav, 2011), self-organising maps (Quah and Sriganesh, 2008; Olszewski, 2014), logistic regression (Viaene et al., 2007; Pinquet et al., 2007), and fuzzy logic (Sanchez et al., 2009). Additionally, some researchers focused solely on classification and regression trees (Bai et al., 2008; Sahin et al., 2013), Bayesian belief networks (Holton, 2009), individual statistical techniques (Panigrahi, 2009), artificial immune systems (Wong et al., 2013; Halvaiee and Akbari, 2014), or their own hybrid methods (Duman and Ozcelik, 2011). But it is difficult to understand the relative play of the technique without comparing it to other methods or techniques. Some extra factors such as the fraud type researched and the specific dataset used can influence the results of the experiment. Many researchers used fuzzy logic to introduce variation to their samples, attempting to transform it to resemble real world data before deploying a different technique to actually detect the presence of frauds (Jans et al., 2011).

Figure 7 Hybrid methods used in financial fraud detection (see online version for colours)



Moreover, many researchers use hybrid methods which utilise the strengths of multiple algorithms to classify samples and result a better fraud detection accuracy (Duman and Ozcelik, 2011; Panigrahi, 2009). Duman and Ozcelik (2011) have used a combination of Scatter search and Genetic algorithm, based on the latter but targeting attributes of scatter search such as the smaller populations and recombination as the reproduction method. Panigrahi (2009) used two methods sequentially, beginning with the Depster-Schaefer method to combine rules and then using a Bayesian learner to detect the existence of fraud. Some others researchers combined traditional computational intelligence methods with text mining to analyse financial statements for the presence of fraud (Cecchini et al., 2010; Humpherys et al., 2011; Dong et al., 2014). The applications of hybrid method in different types fraud detection is given in Figure 7.

The different fraud detection techniques are compared with their strong point and drawbacks in Table 3.

Table 3 Comparison of different data mining techniques for financial fraud detection

Techniques	*Strengths*	*Limitations*
Logistic model	Easy to implement and mainly used for credit card and financial statement fraud detection, provide better accuracy for credit card fraud detection	Higher computational complexity but lower performance
Decision trees	Simple and easy to implement, requires very low computational power for training and operation which makes it suitable for real-time operation, used for financial statement fraud detection	Initial setup requires high computational power and demands training set with high representation of problem domain
Genetic algorithm	Simple method and suitable for non-algorithmic and binary classification; mainly used in financial statement fraud detection	Suffering from local minima/ local maxima problem; computational cost is high for training and operation
Bayesian belief network	Suitable for non-algorithmic and binary classification problems; used for financial statement fraud detection with high efficiency	Complete knowledge about typical and abnormal behaviour of the fraud type is a primary requirement
Artificial immune system	Suitable to classify imbalanced data like credit card transactions	Unsuitable for real-time function due to high computational cost for operation
Neural network	Suitable for non-algorithmic and binary classification problems; applied for investigating credit card and financial statement fraud detection	High computational power is required for training
Support vector machine	Capable to solve nonlinear classification problem and requires low computational power for training and operation; used for credit card, insurance, financial statement fraud detection with high accuracy for credit card fraud detection	Difficult to process results due to transformed input set
Self-organising map	Provides easy understandable results for auditors and this technique is simple to implement; better accuracy is reported for credit card fraud detection	Visualisation needs consistent observation of auditors as the technique is not fully automated in operation
Fuzzy logic	Simple and saves computing power with respect to other techniques, suitable to design a solution for the dataset where attributes are described in linguistics and/ or numerical terms	Fuzzy logic does not fit to every problem

4.3 Performance measures of different techniques

In general, several factors are being considered for measurement of performance. But, most commonly used matrices are accuracy, sensitivity and specificity. Accuracy is the ratio of successful classification number to unsuccessful classification number. Sensitivity measure amount of samples correctly identified as fraud to the amount of incorrectly identified fraud. Specificity compare between true positive and false positive (West and Bhattacharya, 2016).

In economic world where, financial fraud is a real concerned whereas the accuracy measurements are generally being taken under consideration for performance measure. In addition to this performance measures several other measures have been used by many researchers, few of them used software-based success rate to determine the success rate of fraud detection algorithm (Sanchez et al., 2009).

Analysing various techniques applied in fraud detection and their results, it is clear that computational methods have better success rate compared to statistical method. But in some cases statistical methods were found to be more accurate than computational method. For example, Bayesian belief network produces more accurate result than decision tree and neural network. According to the feature selection of a particular fraud type, the result of computational and statistical method may differ.

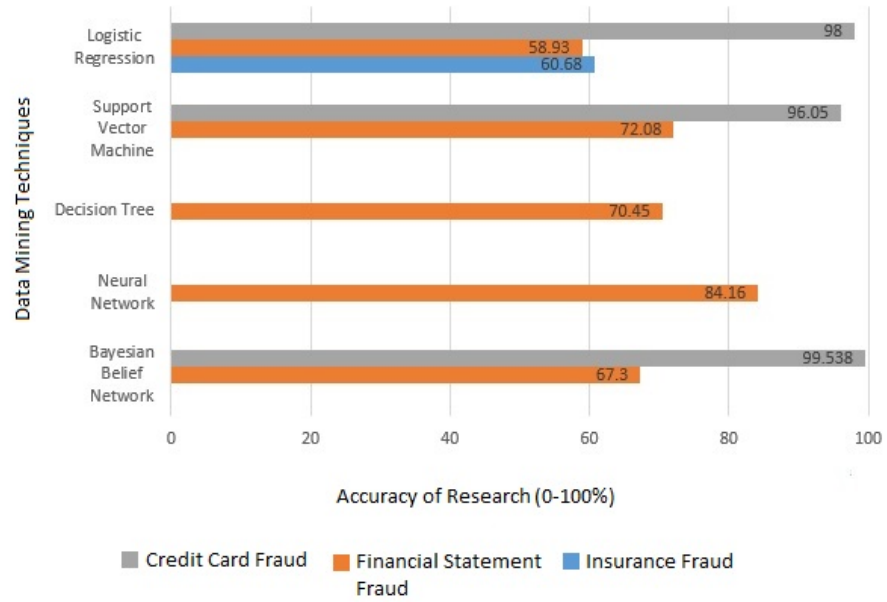
Table 4 Classification based on accuracy and techniques

Technique	*Type investigated*	*Accuracy*	*Reference*
Logistic model	Credit card fraud	96.6%–99.4%	Bhattacharyya et al. (2011)
	Insurance fraud	60.680%	Bermudez et al. (2008)
	Financial statement fraud	66.86%–70.86%	Ravisankar et al. (2011)
	Financial statement fraud	19%–79%	Huang (2014)
Support vector machine	Credit card fraud	95.5%–96.6%	Bhattacharyya et al. (2011)
	Insurance fraud	70.41%–73.41%	Ravisankar et al. (2011)
	Financial statement fraud	50%–81.97%	Cecchini (2010)
	Financial statement fraud	65.8%	Humpherys (2011)
Self-organising map	Credit card fraud	100%	Olszewski (2014)
Genetic programming	Financial statement fraud	89.27%–94.14%	Ravisankar et al. (2011)
Neural network	Financial statement fraud	80%	Kirkos et al. (2007)
	Financial statement fraud	77.14%	Bose and Wang (2007)
	Financial statement fraud	95.64%–98.09%	Ravisankar et al. (2011)
Artificial immune system	Credit card fraud	94.6%–96.4%	Halvaiee and Akbari (2014)

The accuracy of the financial fraud detection techniques with respect to the different types of frauds as reported by the researchers are shown in Table 4. The analysis of the accuracy of different techniques shows that the logistic model, support vector machine, self-organising map, genetic programming, neural network, artificial immune system applied with proper fraud type and intelligent feature selection produce results with better accuracy. We have also analysed the accuracy of a specific fraud detection technique with respect to different financial frauds and the observation is plotted in

Figure 8. We observed that the result produced by a technique applied in two different types of fraud can be different.

Figure 8 Accuracy comparison between fraud types and applied techniques (see online version for colours)



5 Financial fraud challenges

Financial fraud detection is an emerging topic in recent year. Here researchers need to think above the fraudsters. Also there are many frauds exist which have not been investigated. In this section we will point out few important challenges for researches.

- *Data inadequacy:* Financial fraud is a delicate issue, as stakeholders are cautious to share information related to financial fraud, this has often led to problem for research due to data inadequacy.
- *Fraud type, and applied method:* Financial fraud is a varying field. It demands continuous study and upgradation of different techniques according to the nature of frauds to minimise the effect of frauds. But, there is a disproportion in both fraud types and detection methods studied. Some have studied extensively while others have been looked in very few cases. It has been found that the research in credit card fraud is more frequent compared to research in money laundering.
- *Computational performance:* Often it is necessary to detect frauds as soon as possible to prevent catastrophic effect on economy, a little amount of research have been conducted on computational performance of fraud detection technique for real-time use.

- *Emerging problem:* Fraudsters are continually evolving and modifying their technique to remain undetected. So, researchers also need to constantly adapt to the situation and find a way to prevent the fraudsters.

6 Conclusions and future implementation

In the modern economic industry, financial fraud detection is an essential part. This literature review looks into many emerging techniques and their performance on different frauds occurred in past. This review presents a comparison of accuracy of fraud detection techniques used on financial frauds and their result and effects on economic industry are also discussed. Both computational and statistical methods are used in financial fraud detection, and the efficiency and suitability of the techniques are discussed in this paper. Support vector machine, neural network are to be popular among researchers due to their adaptability to new techniques, and evolving tactics of committing frauds. A handful number of papers published on financial fraud detection in last decade. But it has been noticed that the researchers considered very few types of frauds and techniques. Mainly, credit card fraud, Insurance fraud and financial statement fraud got more attention but very few number of paper was published on money laundering. Many of researchers conducted their research based on established techniques in past, but they need to adapt new technique with improved feature selection and variation in parameter selection.

References

Agyemang, M., Barker, K. and Alhajj, R. (2005) 'A comprehensive survey of numeric and symbolic outlier mining techniques', *Intelligent Data Analysis*, Vol. 10, No. 6, pp.521–538.

Alden, M.E., Bryan, D.M., Lessley, B.J. and Tripathy, A. (2012) 'Detection of financial statement fraud using evolutionary algorithms', *Journal of emerging technologies in accounting*, Vol. 9, No. 1, pp.71–94.

Alruily, M., Ayesh, A. and Zedan, H. (2014) 'Crime profiling for the arabic language using computational linguistic techniques', *Inf. Process. Manage.*, Vol. 50, No. 2, pp.315–341.

Argyrou, A. (2012) *Auditing Journal Entries using Self-organizing Map*, Association for Information Systems, Atlanta, Georgia, ISBN 978-0-615-66346-3.

Aris, N.A., Othman, R., Arif, S.M.M., Malek, M.A.A. and Omar, N. (2013) 'Fraud detection: Benford's law vs Beneish model', *IEEE Symposium on Humanities, Science and Engineering Research*, pp.726–731.

Artís, M., Ayuso, M. and Guillen, M. (1999) 'Modelling different types of automobile insurance fraud behaviour in the Spanish market', *Insurance: Mathematics and Economics*, Vol. 24, Nos. 1–2, pp.67–81.

Bai, B., Yen, J. and Yang, X. (2008) 'False financial statements: characteristics of China's listed companies and cart detecting approach', *International Journal of Information Technology & Decision Making*, Vol. 7, No. 2, pp.339–359.

Bell, T.B. and Carcello, J.V. (2000) 'A decision aid for assessing the likelihood of fraudulent financial reporting', *Auditing: A Journal of Practice & Theory*, Vol. 19, pp.169–184.

Benitez, I., Quijano, A., Diez, J.L. and Delgado, I. (2014) 'Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers', *International Journal of Electrical Power & Energy Systems*, Vol. 55, pp.437–448.

Bermudez, L.l. and Perez, J.M., Ayuso, M., Gomez, E. and Vazquez, F.J. (2008) 'A Bayesian dichotomous model with asymmetric link for fraud in insurance', *Insurance: Mathematics and Economics*, Vol. 42, No. 2, pp.779–786.

Bhattacharyya, S., Jha, S., Tharakunnel, K. and Westland, J.C. (2011) 'Data mining for credit card fraud: a comparative study', *Decision Support System*, Vol. 50, pp.602–613.

Bloomfield, R. (2012) 'Discussion of detecting deceptive discussions in conference calls', *Journal of Accounting Research*, Vol. 50, No., pp.541–552.

Bose, I. and Wang, J. (2007) 'Data mining for detection of financial statement fraud in Chinese companies', *International Conference on Electronic Commerce*, pp.15–17.

Caron, F., Vanthienen, J. and Baesens, B. (2013) 'Comprehensive rule-based compliance checking and risk management with process mining', *Decision Support Systems*, Vol. 54, No. 3, pp.1357–1369.

Cecchini, M., Aytug, H., Koehler, G.J. and Pathak, P. (2010) 'Making words work using financial text as a predictor of financial events', *Decision Support Systems*, Vol. 50, pp.164–175.

Chan, P.K., Fan, W., Prodromidis, A.L. and Stolfo, S.J. (1999) 'Distributed data mining in credit card fraud detection', *IEEE intelligent systems and their applications*, Vol. 14, No. 6, pp.67–74.

Chen, R-C., Chen, T.-S. and Lin, C-C. (2006) 'A new binary support vector system for increasing detection rate of credit card fraud', *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 20, No. 2, pp.227–239.

Colantonio, A., Pietro R.D., Ocello, A. and Verde, N.V. (2011) 'A new role mining framework to elicit business roles and to mitigate enterprise risk', *Decision Support Systems*, Vol. 50, No. 4, pp.715–731.

Debreceny, R.S. and Gray, G.L. (2011) 'Data mining of electronic mail and auditing : a research agenda', *The journal of Information Systems: JIS*, Vol. 25, No. 2, pp.195–226.

Dilla, W.N. and Raschke, R.N. (2015) 'Data visualization for fraud detection: practice implications and a call for future research', *International Journal of Accounting Information Systems*, Vol. 16, pp.1–22.

Dong, W., Liao, S.S., Fang, B., Cheng, X., Chen, Z. and Fan, W. (2014) 'The detection of fraudulent financial statements: an integrated language model', *PACIS Proceeding*, p.383.

Dorronsoro, J.R., Ginel, F., Sgnchez, C. and Cruz, C.S. (1997) 'Neural fraud detection in credit card operations', *IEEE Transactions on Neural Networks*, Vol. 8, No. 4, pp.827–834.

Duman, E. and Ozcelik, H.M. (2011) 'Detecting credit card fraud by genetic algorithm and scatter search', *Expert Syst. Appl.*, Vol. 38, pp.13057–13063.

Gadi, M.F.A., Wang, X. and Lago, A.P.d.(2008) 'Comparison with parametric optimization in credit card fraud detection', *Machine Learning and Applications, ICMLA '08, Seventh International Conference on*, pp.279–285.

Gao, Z. and Ye, M. (2007) 'A framework for data mining–based anti–money laundering research', *Journal of Money Laundering Control*, Vol. 10, No. 2, pp.170–179.

Genzman, L. (1997) 'Responding to organized crime: laws and law enforcement. organized crime', in H. Abadinsky (Ed.), p.342, Belmont.

Glancy, F.H. and Yadav, S.B. (2011) 'A computational model for financial reporting fraud detection', *Decision Support Systems*, Vol. 50, No. 3, pp.595–601.

Goode, S. and Lacey, D. (2011) 'David Lacey, detecting complex account fraud in the enterprise: The role of technical and non-technical controls', *Decision Support Systems*, Vol. 50, No. 4, pp.702–714.

Gray, G.L. and Debreceny, R.S. (2014) 'A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits', *International Journal of Accounting Information Systems*, Vol. 15, No. 4, pp.357–380.

Halvaiee, N.S. and Akbari, M.K. (2014) 'A novel model for credit card fraud detection using artificial immune systems', *Applied Soft Computing*, Vol. 24, pp.40–49.

Han, J. and Kamber, M. (2006) 'Data mining: concepts and techniques, 2nd ed., Burlington, Massachusetts, USA, ISBN: 1-55860-901-6.

Holton, C. (2009) 'Identifying disgruntled employee systems fraud risk through text mining: a simple solution for a multi-billion dollar problem', *Decision Support Systems*, Vol. 46, pp.853–864.

Huang, S.Y., Tsaih, R.H. and Yu, F. (2014) 'Topological pattern discovery and feature extraction for fraudulent financial reporting', *Expert Systems with Applications*, Vol. 41, No. 9, pp.4360–4372.

Humpherys, S.L., Moffitt, K.C., Burns, M.B., Burgoon, J.K. and Felix, W.F. (2011) 'Identification of fraudulent financial statements using linguistic credibility analysis', *Decision Support Systems*, Vol. 50, No. 3, pp.585–594.

Jans, M., Werf, J.M.V.D., Lybaert, N. and Vanhoof, K. (2011) 'A business process mining application for internal transaction fraud mitigation', *Expert Systems with Applications*, Vol. 38, pp.13351–14009.

Kanapickiene, R. and Grundiene, Z. (2002) 'The model of fraud detection in financial statements by means of financial ratios', *Procedia – Social and Behavioral Sciences*, Vol. 213, pp.321–327.

Kirkos, E., Spathis, C. and Manolopoulos, Y. (2002) 'Data mining techniques for the detection of fraudulent financial statements', *Expert Systems with Applications*, Vol. 32, No. 4, pp.995–1003.

Koh, H.C. and Low, C.K. (2004) 'Going concern prediction using data mining techniques', *Managerial Auditing Journal*, Vol. 19, No. 3, pp.462–476.

Kotsiantis, S., Koumanakos, E., Tzelepis, D. and Tampakas, V. (2006) 'Forecasting fraudulent financial statements using data mining', *International Journal of Computational Intelligence*, Vol. 3, No. 2, pp.104–110.

Kpodoh, B. (2009) *Bankrupty and Financial Distress Prediction in the Mobile Telecom Industry the Case of MTN-Ghana Millicom-Ghana and Ghana telecom*, School of Management, Blekinge Institute of Technology.

Kummer, T.F., Singh, K. and Best, P. (2015) 'The effectiveness of fraud detection instruments in not-for-profit organizations', *Managerial Auditing Journal*, Vol. 30, Nos. 4–5, pp.435–455.

Kwon, O. and Sim, J.M. (2013) 'Effects of data set features on the performances of classification algorithms', *Expert Systems with Applications*, Vol. 40, No. 5, pp.1847–1857.

Lin, C.C., Chiu, A.A., Huang, S. Y. and Yen, D.C. (2015) 'Detecting the financial statement fraud: the analysis of the differences between data mining techniques and expert's judgments', *Knowledge-based Systems*, Vol. 89, pp.459–470.

Lisic, L.L., Silveri, S.D., Song, Y. and Wang, K. (2015) 'Accounting fraud, auditing, and the role of government sanctions in China', *Journal of Business Research*, Vol. 68, No. 6, pp.1186–1195.

Mahama, M. (2015) 'Detecting corporate fraud and financial distress using the Altman and Beneish models the case of Enron Corp', *International Journal of Economics, Commerce and Management*, UK, Vol. 3, No. 1.

Mei, D. and Zhou, L. (2015) 'Anti-money laundering game between banking institutions and employees in the progressing CNY internationalization', *Modern Economy*, Irvine, Vol. 6, No. 4, pp.490–497.

Levi, M. (2002) 'Money laundering and its regulation', *The Annals of the American Academy of Political and Social Science*, Vol. 582, pp.181–194.

Ngai, E.W.T., Hu, Y., Wong Y.H., Chen, Y. and , Sun, X. (2011) 'The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature', *Decision Support System*, Vol. 50, No. 3, pp.559–569.

Olszewski, D. (2014) 'Fraud detection using self-organizing map visualizing the user profiles', *Knowledge-based Systems*, Vol. 70, No. C, pp.324–334.

Pal, M. and Paul M.M. (2003) 'An assessment of the effectiveness of decision tree methods for land cover classification', *Remote Sensing of Environment*, Vol. 86, No. 4, pp.554–565.

Panigrahi, S., Kundu, A., Sural, S. and Majumdar, A.K. (2009) 'Credit card fraud detection: a fusion approach using Dempster-Shafer theory and Bayesian learning', *Information Fusion*, Vol. 10, pp.354–364.

Panik, M. (2009) *Regression Modeling Methods, Theory and Computation with SAS*, 2nd ed., The CRC Press, United States of America, Boca Raton, ISBN 1420091980.

Pathak, J., Vidyarthi, N. and Summers, S.L. (2005) 'A fuzzy–based algorithm for auditors to detect elements of fraud in settled insurance claims', *Managerial Auditing Journal*, Vol. 20, No. 6, pp.632–644.

Pinquet, J., Ayuso, M. and Guillen, M. (2007) 'Selection bias and auditing policies for insurance claims', *Journal of Risk and Insurance*, Vol. 74, pp.425–40.

Quah, J.T. and Sriganesh, M. (2008) 'Real-time credit card fraud detection using computational intelligence', *Expert systems with Applications*, Vol. 35, No. 1, pp.1721–1732.

Ravisankar, P., Ravi, V., Rao, G.R. and Bose, I. (2011) 'Detection of financial statement fraud and feature selection using data mining techniques', *Decision Support Systems*, Vol. 50, No. 2, pp.491–500.

Sahin, Y., Bulkan, S. and Duman, E. (2013) 'A cost-sensitive decision tree approach for fraud detection', *Expert systems with Applications*, Vol. 40, No, 15, pp.5916–5923.

Sánchez, D., Vila, M., Cerda, L. and Serrano, J.M. (2011) 'Association rules applied to credit card fraud detection', *Expert Systems with Applications*, Vol. 36, pp.3630–3640.

Shaw, M.J., Subramaniam, C., Tan, G.W. and Welge, M.E. (2001) 'Knowledge management and data mining for marketing', *Decision Support Systems*, Vol. 31, No. 1, pp.127–137.

Specht, D.F. (1990) 'Probabilistic neural networks', *Neural Networks*, Vol. 3, No. 1, pp.109–118.

Srivastava, A., Kundu, A., Sural, S. and Majumdar, A. (2008) 'Credit card fraud detection using hidden Markov model', *IEEE Transactions on Dependable and Secure Computing*, Vol. 5, No. 1, pp.37–48.

Yang, S.(2014) *5 Years Ago Bernie Madoff Was Sentenced to 150 Years In Prison – Here's How His Scheme Worked*[online] http://www.businessinsider.in (accessed 01/06/2016).

Strohmeier, S. and Piazza, F. (2013) 'Domain driven data mining in human resource management: a review of current research', *Expert Systems with Applications*, Vol. 40, No. 7, pp.2410–2420.

Subelj, L., Furlan, S. and Bajec, M. (2011) 'An expert system for detecting automobile insurance fraud using social network analysis', *Expert Systems with Applications*, Vol. 38, No. 1, pp.1039–1052.

Svozil, D., Kvasnicka, V. and Pospichal, J. (1997) 'Introduction to multi-layer feed-forward neural networks', *Chemometrics and Intelligent Laboratory Systems*, Vol. 39, No. 1, pp.43–62.

Tarjo, Herawati, N. (2015) 'Application of Beneish M-score models and data mining to detect financial fraud', *Procedia – Social and Behavioral Sciences*, Vol. 211, pp.924–930.

Thanathamathee, P. and Lursinsap, C. (2013) 'Handling imbalanced data sets with synthetic boundary data generation using bootstrap re-sampling and AdaBoost techniques', *Pattern Recognition Letters*, Vol. 34, No. 12, pp.1339–1347.

Throckmorton, C.R, Mayew, W.J., Venkatachalam, M. and Collins, L.M. (2015) 'Financial fraud detection using vocal, linguistic and financial cues', *Decision Support Systems*, Vol. 74, pp.78–87.

Sidney Tsang, S., Koh, Y.S., Dobbie, G. and Alam, S. (2014) 'SPAN: finding collaborative frauds in online auctions', *Knowledge-based Systems*, Vol. 71, No. 7, pp.389–408.

Turban, E., Aronson, J.E., Liang, T.P. and Sharda, R. (2007) *Decision Support and Business Intelligence System*, 8th ed., Pearson Education.

Vapnik, C. and Cortes, V. (1995) 'Support-vector networks', *Remote Sensing of Environment*, Vol. 20, No. 3, pp.273–297.

Viaene, S., Ayuso, M., Guillen, M., Gheel, D.V. and Strategies, D.G. (2007) 'For detecting fraudulent claims in the automobile insurance industry', *Eur. J. Oper. Res*.

West, J. and Bhattacharya, M. (2016) 'Intelligent financial fraud detection: a comprehensive review', *Computers & Security*, Vol. 57, pp.47–66.

Wong, N., Ray, P., Stephens, G. and Lewis, L. (2012) 'Artificial immune systems for the detection of credit card fraud: an architecture, prototype and preliminary results', *Information Systems Journal*, Vol. 22, pp.53–76.

Yamanishi, K., Takeuchi, J.I., Williams, G. and Milne, P. (2004) 'On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms', *Data Mining and Knowledge Discovery*, Vol. 8, No. 3, pp.275–300.

Yeh, I-C. and Lien, C-h. (2007) 'The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients', *Expert Systems with Applications*, Vol. 36, No. 2, Part 1, pp.2473–2480.

Yigitbasioglu, O.M. and Velcu, O. (2012) 'A review of dashboards in performance management: implications for design and research', *International Journal of Accounting Information Systems*, Vol. 13, No. 1, pp.41–59.

Yue, D., Wu, X., Wang, Y., Li, Y. and Chu, C.H. (2007) 'A review of data mining-based financial fraud detection research', *International Conference on Wireless Communications, Networking and Mobile Computing*pp.5519–5522.

Zaslavsky, V. and Strizhak, A. (2006) 'Credit card fraud detection using self-organizing maps', *Information & Security: An International Journal*, Vol. 18, pp.48–63.

Zhang, D. and Zhou, L. (2004) 'Discovering golden nuggets: data mining in financial application', *IEEE Transactions on Systems, Man, and Cybernetics, Part c (Applications and Reviews)*, Vol. 34, No. 4, pp.513–522.

Zhou, W. and Kapoor, G. (2011) 'Detecting evolutionary financial statement fraud', *Decision Support Systems*, Vol. 50, No. 3, pp.570–575.