A Survey of Financial Statement and Credit Card Fraud Prevention and Detection by Using Data Mining Techniques

Moayd Bahumayd
Department of Information Systems, Faculty of Computing and Information Technology, King Abdul-Aziz University, Jeddah, Saudi Arabia

Abstract—Fraud has become widespread in the world and that make governments, companies and people in concern. All countries in the world nowadays concerned for reducing risks from fraud. Fraud has different types and forms but financial statement fraud still the most prevalent. Also in financial fraud, there are various kinds like insurance fraud, bank fraud, credit card fraud, financial statement fraud, and other kinds. Organizations continually try to improve and increase effectiveness and efficiency in prevent and detect fraud. Many steps can used to prevent and detect fraud but Data mining still best technology can help in overcome and decrease risks of fraud. Data mining has also many algorithms but in this paper, we will only discuss five types of theses algorithms that are decision tree, super vector machine, logistic regression, neural network and genetic algorithms.

Keywords—Financial statement fraud, Credit card fraud. Prevention and detection Fraud, Data mining techniques,

I. INTRODUCTION
Nowadays with the expansion in technology and the large amount of data, fraud has become spread and diverse. Increasing in data make the control on it very difficult, that increase chances to fraud. Fraud consider as a popularity problem in the world. Technology expanded domain of fraud in our life, of course anyone exposed for any type of fraud in anytime and anywhere. Fraudsters every day develop their ways to execute their crimes. There are different definitions of fraud, oxford dictionary defined fraud as intentional mistake or criminal deception in order to acquire personal financial gains (1). In detail, we can defined fraud as any work or track of deceit, hiding deliberate or change truth in order to gain illegal benefits or encourage other to share some value without right (2). Fraud detection is the process of prevention the scammers to access and achieve their aims by using different ways or systems. Fraud can be in different fields such as credit card and financial statement fraud. Annually all countries suffer from financial losses because the fraud and this affect on governments, organizations, companies and persons. Based on survey made by Association of Certified Fraud Examiners during period of January 2004 and 2006 depending on only 1134 cases, 45% of companies in the world have been victims to financial crime through period of 2004 and 2005. Sometimes fraud gets through the angel, in ACFE survey found that Frauds cases that occurred by owners or executive managers around $1 million. That is almost five times of the median loss caused by managers, also around 13 times plus on median loss caused by employees. At ACFE survey, in organizations that have fewer of 100 employees the median loss of fraud was $190,000 per scheme (3). Around the world, the average financial damage to companies subjected to the PWC survey, was US$ 1.7 million per company (4). The United Kingdom loss at private and government sectors around £73 billion per year in all economic fields. The cost of fraud only at local government sector is £2.2 billion a year (5). Today data mining consider as a top technology for prevent and detect fraud. Data mining one of the ways that currently used for fraud detection. Data mining for fraud detection has different techniques (such as classification, clustering, prediction and so on). Definition of data mining is the process of carefully accurate previously stored huge data to extract information. This information should be clear and understood to use for decision support (6). Data mining means extracting patterns and structure from large, diverse and unknown databases (7). Data mining consist of different things like statistics, artificial, intelligence, visualization, and database management. Extracted and understand the nature of the stored data are the two primary missions of data mining (8). Data mining means studying, monitoring and analyzing data to take information on relations or patterns in data then gain knowledge that help in make DSS. Not important how the data are collected, which mean data was collected not for purpose of mining. Data mining does not look for special information, but it study all characteristics in data. Data mining has different algorithms everyone has special characteristics. In this paper, we will briefly talk on five famous types of data mining algorithms Decision Tree, Super Vector Machine, Logistic Regression, Neural Network and Genetic Algorithms.

II. TYPES OF FRAUD
Fraud has become more spread term around the world. Diversity and difference shapes of fraud forms now is surrounding of us. Today with new technologies expanded limits of fraud domains either place and time domain. This expend made anyone now is exposed to fraud whether if he was at home or work or in anywhere also this fraud can occur
at any time. Fraud is a broad term so it can cover different illegal benefits. Technical boom contributed to the development different areas. Fraudsters try to use this technology to create new fraud types. Nowadays many types of fraud have emerged recently. Financial fraud is a most famous type and has many different shapes can classify as type of it. Financial crimes and fraud are a type of deception or theft that occur when persons or entities take unauthorized benefits in illegal way this usually occur in order to gain personal interests(9). Financial fraud means deception and fraud in financial transactions either by hiding truth information or by using incorrect information and access to confidential information in order to using them for gain illegal financial benefits. There are too many types of financial fraud like electronic mail messages that targeting to collect personal banking information from victims by introduce offers or goods with low prices then, the purchasing requires credit card and bank account information. Another example of electronic mail fraud is Phishing in electronic mail by sending messages ask to confirm personal's bank's information and using fake links and fake banks sites to delude people that these are a real sites. In addition, there are another ways for fraud like fraud in automated teller machines (ATM) that are available in anywhere, the fraudsters use machines for scanning card's information and password after that they use this information by fake cards. One of examples of banking card's fraud is calls from unknown persons asking to refresh banking card's information then use this information in electronic fraud. The theft of cards or peeking on card's owner during his using ATM also are types of financial fraud. In this paper will focus on two famous types of financial fraud. First is financial statement fraud in organizations and companies. Second is credit cards fraud that we gave some examples of it.

A. Financial statement fraud:-
Financial statement fraud is one of the fraud types. It refers to deliberate use incorrect information for prepare financial statements that may cause misunderstand(10). Inaccurate performance in the financial accounts may occur due to fraud or error. The difference between error and fraud is financial statement fraud come in intentionally way but in error there is no intentional (11). In the modern business world, financial statement fraud is classifying as one of main problems. Financial statement fraud is intention misrepresentation on financial amounts. Companies use financial statement to provide stakeholders an overview of the sales and revenues. In addition, Misinformation in financial statement considered also as fraud(12). Financial statement fraud is intentional deception by hiding peace of truth information and replace it with false information for creating incorrect impression on organization's financial stamen and deception the readers(13). In contemporary businesses, there are many concerns but still financial statement fraud is a one of big concern. Manipulation in financial reports or budgets gives chances to managers in give a mistake overview on current financial statues for company. Financial statement fraud can occur in many forms for example disclosure of false revenues like Reduce the real value of the amount of revenue from high to low in order to gain personal benefit from concealed amounts. Another form of financial statement fraud is Manipulation of expenses such as coordination with suppliers to raise the value of products in invoices comparing with real value in order of taking personal benefits from the difference between two values. Victims in financial statement fraud are the angel, shareholders, financier and investors because they used financial reports to take decision on their investments. Crimes in financial statement fraud occur by some people inside company not all company; the criminals maybe have different levels like managers, employees and others. The main motivations for financial statement fraud are the easy of doing these crimes because criminals always inside company and sometimes maybe managers have top levels that mean there are broad powers to make fraud. Another motivations to make manipulation in financial statements are the high amounts that gained by employees. In addition, sometimes criminals think there are difficult to detect their frauds. Companies always give high importance to prevent fraud. To detect and prevent fraud in management, accountants and auditors companies need to use modern techniques to detect financial statement fraud. Standards auditing procedures sometimes are not enough and insufficient because sometime top managers are involved in fraud crimes. Currently Data mining techniques considered as a good choice with traditional ways for detect and prevent fraud with better and effective results(14).

B. Credit card fraud:-
Although many of the controls set by the banks to protect their customers and their assets, but the credit card fraud still occur. In last period, fraud in banks have become widespread and a lot of money Stolen by crooks through fraud the bankcards. There are some definitions for credit card fraud, we will mention to some of them. Credit card fraud occur if someone trying to gain money, services or goods by using someone's card without take an authorization from card's owner(9). The ways that can used for credit card fraud are different, CCF can occur through a various ways of methods like theft the card or theft card's for the purpose of taking illegal benefits from an account(15). Credit card fraud is a form of theft of card's information and that by taking the person to a bank card information to someone else and then be used for the purpose of purchase or withdraw funds. Fraud in the credit card is to use a card of another person without his permission and for the purpose of use in personal benefits(16). All ways that used forcredit card fraud depend on two main ways, automated teller machine (ATM) and internet. Almost many of fraud operations occur by internet, In the European Union, reports on credit cards in 2008 showed that 50% of frauds were via the Internet. Criminals to do their operations can use different forms. Some professional styles need to expert criminal gangs. One of forms is deceiving and misleading people through delude them earn large amounts of money. Some people Deceived by fake e-mail messages requesting card information. One of the ways when buying from websites, crook enters a secret random numbers may sometimes one of them to be true. Another form is Calls that come for customers to ask them to update their card's information. Sometimes fraudsters put small devices in point of sale (POS) to read card information and then reused it by using forged cards. Finally Traditional methods such as lost or stolen card.
III. IMPACTS OF FRAUD

These days a lot of fraud has spread around the world. All a fraud occur have a negative impact. This impact not only on direct victim but sometimes extend to their family or all society. There are many impacts of fraud first on the individuals and all operations together can effect on economic of countries. Financial losses can effect on economic of countries. We will discuss here two main impacts individual and global.

A. Individual impacts

There are different impacts both tangible and non-tangible. Tangible means the impact that can see as financial loss or theft of bankcard. Non-tangible means the impact that cannot be seen as psychological impact. In a study in England and Wales on fraud victims, the results demonstrate that there is a variation of impacts: some of victims say the impacts were very little. Other victims say the fraud is a crime has devastating impacts more than only loss of money. About 1 in 10 of respondents the financial losses of fraud were more than year's household income. The research found that there are some changes of victim's behavior because of fraud. The services and support provided to victims are still not enough and not comprehensive to their needs (17). The work at unsafe environment leads too threats and negatives affects as loss of confidential information. There are different impacts and costs of fraud operations. Many of measures that talk about impacts of fraud only talks about financial impacts but there are another negatives results and impacts on victims. We cannot specified negative impacts only at loss of money. However, physical, psychological, and social costs considered as negatives impacts of fraud. Sometimes the fraud not only affect on victims but also their families maybe suffer of fraud (18).

B. Global impacts

The losses and costs of financial fraud around the world are very high. The United Kingdom loss at private and government sectors around £73 billion per year in all economic fields. The cost of fraud only at local government sector is £2.2 billion a year (5). Depending on ACFE survey, 45% of companies in the world have been victim to financial crime through period of 2004 and 2005 (3). In news report on BBC at 2007, the United Kingdom losses annually 1.6 billion pound from fraudulent insurance claims. At Enron Broadband Services company (EBS) after his confession that he was involved in fraud operations, the chief executive officer announced the allotment of the nearly 8.7 million $ to the Enron victims (19). Auditor General’s (AG) (2012) Reports in some Malaysian states showed that some of the deceased was receiving welfare payments to them for several years after their death. In the 2012 report, questioned at RM 303,813 demanded by the senior official in the Ministry of Communication and Culture of Malaysia on travel costs for four days to Geneva, auditors showed it is worth only RM 50,000. In many cases, some government departments were paid an inflated price for their purchases. In the same subject, but in the private sector, only it reports in 2006 about existence of 6921 cases of commercial crimes, and in the same year showed reports on the 2892 crime in counterfeiting banknotes and credit card operations (20). In a report in the European countries from credit cards, it found 50% of the lost cards in 2008 used in fraud via the internet (21).

The impact of fraud can be either small or large impact depended on the victim and size of fraud. Domain of impact is not limited on victim only but usually extend to his family and society. AS we need to previous steps to prevent and detect fraud before it occurs, we need to use steps to treatment impacts of fraud on victims. Victims of fraud can be a person, company and banks.

IV. FRAUD PREVENTION AND DETECTION.

Data mining algorithms still a first choice at fraud prevention and detection although their implementation work follow traditional way of information flow as data collection, feature selection, representation and management, preprocessing and performance evaluation. There is a general framework to define the basic steps for using data mining in fraud detection techniques (fig1)

![Figure 1: Framework for DM based on FFD](image-url)
The traditional method of fraud detection and prevention systems are depend on simple comparisons but also depend on main data mining types as clustering, association, and prediction. The one of the main disadvantages at current fraud prevention and detection systems are not good at diverse databases and time complexity. To overcome on the issues there is need to use an automated fraud detection framework as expert system in order to identify data mining techniques, data fusion techniques and intelligent agents (23).

Traditionally, the governments are usually dealing with fraud by the reaction such as the investigation and trial. The one of the efficient ways to prevent fraud is processing it when they occur but the better and crucial way at detection and prevention fraud is a counter fraud before it occur. Exceeded stage of making sure that prevention fraud systems are available and ready to extending in development anti-fraud culture that can help in zero tolerance and prevent criminals. Making better use of information and technology is very importance of fraud prevention that can help and give a major rule at data analytical software and at reporting processes. It is very importance when analyzing the previous fraud cases and study them to make a report about all pre cases to give information on where the fraud is existing. The audit committees must receive continues and updating reports about fraud (5).

V. DATA MINING TECHNIQUES

Data mining plays a main rule in different fields around the world. The organizations constantly works to prevent and detect fraud, and always try to found modern and best technology to avoid risk of fraud. Data mining can help to give safeguard of fraud by different algorithms. In recent years, financial world gave data mining a widespread attention and increasing in popularity. Recent surveys have found that data mining applications have a growth in usage and effectiveness. Furthermore, if there is trick by financial statement fraudsters to deceive the detection programs, data mining able to use an urgent Alternatives way to prevent that risk if there is trick by financial statement fraudsters to deceive the detection programs, data mining able to use urgent alternatives ways to prevent that risks (24).

Many data mining algorithms can use in prevention and detection fraud systems. Every one of data mining algorithms has his characteristics, positives and negatives. In this paper, we will only talk briefly about five famous algorithms in this field.

A. Decision Tree

A Decision Tree (DT) is a tree structure, in this tree there are nodes and branches. Every node has a special test on attributes and the outcome of test representing by branch. In this way, the tree continually attempts to separate nodes to subgroups based on the selection of the attributes and tests to give finally the quality when there is no possibility to give more subsets (25).

B. Super Vector Machine

Support vector machine (SVM) is statistical learning techniques that can give nice results in a variety of classification tasks. Support vector machine has a various unique features of other algorithms to be best algorithms in solve problems that have binary classification like fraud detection. The possibility of work in dimensional space without need to add any computational complexity is one of SVM features. The simplicity of SVM and his features attract to using it for fraud detection tasks (26).

C. Logistic Regression.

In problems that have binary choice models like in fraud, logistic regression is very appropriate because of that logistic regression is widely used in detection and prevention fraud systems. Previous works in same areas has used logistic regression models in financial fraud (26).

D. Neural Network

Neural Networks (NN) contains one group of neurons and connected processing units. Neural network consider as a forward and fixed theory in different application areas. Numerical value of each association between neurons is called “weight”. Because of integrating between neurons and the connected input signal are calculated, every neuron has signals. (25).

E. Genetic Algorithms

Genetic algorithms are development algorithms that aim to found best solutions as time progresses.

Since their first introduction by Holland at (1975), genetic algorithms have successfully used in many fields to solve problems. In a computer field and data mining applications, genetic algorithms also have used for select a variable. Sometimes genetic algorithms integrating with another data mining techniques for give better results (27). In data mining and studying the patterns of data Genetic Algorithms considers as an effective tool. The main strength point of genetic algorithms is the way that of showing patterns of data. In genetic algorithms, there are two different approaches to use GA. First one for studying the patterns of data and other one is using genetic algorithms to resetting parameters in classifiers. Almost many of applications use GA in studying patterns of data.

VI. LITERATURE REVIEW

Problems, risks and impacts of fraud that come as results with every fraud operation, encouraged researchers to write many researches on fraud. This give anyone who want to write in this field open space to write. Researchers in the fraud...
are not same because there are different types in fraud and there are different techniques to prevent and detect fraud. In this paper, we only focus on papers that talk on financial statement fraud and credit cards fraud. For techniques, we also focus papers that talk on data mining algorithms. We have founded that there are various kinds of papers, some paper only talk on one algorithm for special type of fraud. Other papers may take comparison between two or more types of algorithms in specified type of fraud and there are papers wrote as a survey to show which type of algorithms and types of frauds profusely used in previous papers and which did not used.

Financial statement fraud is one of famous types of fraud because it can be available in any organization. Fraudsters in financial statement fraud usually are in the organization's staff, so this type is easy to occur. There are famous cases in financial fraud in biggest organizations around the world. Organizations are concerned from this problem, so they try to overcome on it. Many different ways can be used for prevent and detect financial fraud, but none of them are suitable as data mining techniques. Data mining techniques help in studying previous cases of frauds to prevent them in future also can study previous cases to predict new ways of frauds that can occur in the future. There are different types of data mining techniques are used in prevention and detection fraud. Every one of these techniques has its characteristics, strength points, and weak points. For take a better performance some fraud detection and prevention systems combining between two or more data mining techniques, this give better result. Every type of data mining techniques has positives and negatives. One of these types maybe has a strength in specific domain for detection financial fraud but in conversely has limitations so maybe not suitable to use it in another domain. Unlike other financial statement fraud detection systems that only depend on the historical data of financial fraud. There is a modern module that can suggest new future types of fraud before they occurs (24). In different side, other paper made a study in chines companies, they found that there are different data mining techniques to detecting fraud. They used various kinds of data mining techniques like genetic programming, group method of data handling, logistic regression, probabilistic neural network, and support vector machine tested these kinds on datasets of 202 Chinese companies in two time first with feature selection and another without feature selection. The best one without feature selection was probabilistic neural network, genetic programming and probabilistic neural network were best with feature selection (28). On other hand (25) made comparison between only three models of data mining techniques for fraud detection which are neural networks decision tree, and bayesian belief networks. The comparison deepened on performance. The result is bayesian belief networks achieved best performance between three models with 90.3%, neural networks achieved 80% and decision tree achieved 73.6% (22). Searched about previous papers in data mining techniques for fraud detection to help in define different ways of financial statement fraud and to show different data mining techniques for prevent this fraud. They used in their research 18 papers talks about various kinds of data mining techniques as neural networks, regression, decision tree, statistical tests, bayesian and others, 8 of them in regression, 6 for neural networks and others for rest types. In addition, they give a generic framework for using data mining in financial fraud detection.

Data mining techniques are not same in characteristics, so in some cases are suitable to use one of them but in another case may different techniques are better. Most of fraud detection systems depend on the history of data, but there are some systems can also suggest the future types of fraud before they occurs. Famous data mining techniques for fraud detection are bayesian, decision tree, genetic programming, support vector machines, logistic regression, and neural network. The performance is important to differentiation between them. Bayesian and neural network are the best between them. There are many papers were published in data mining techniques for detect and prevent financial statement fraud. Most of papers talked about regression and neural networks. Some of data mining techniques do not have enough papers, so there is a need to extend domain to help in display them. In financial fraud detection, there are some main points that can build a good framework. Feature selection, data collection, data mining and evaluation performance are main structures in that framework (29).

(26) In their paper, they evaluated three types of data mining techniques, support vector machine, random forests and logistic regression. They found that RF hasin general better performance. Secondly LR does better and then SVM in the last. Table 1

<table>
<thead>
<tr>
<th></th>
<th>acc</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Precision</th>
<th>F</th>
<th>AUC</th>
<th>mAPacc</th>
<th>G-mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>0.847</td>
<td>0.654</td>
<td>0.979</td>
<td>0.778</td>
<td>0.709</td>
<td>0.942</td>
<td>0.772</td>
<td>0.8</td>
</tr>
<tr>
<td>SVM</td>
<td>0.938</td>
<td>0.524</td>
<td>0.994</td>
<td>0.792</td>
<td>0.624</td>
<td>0.908</td>
<td>0.878</td>
<td>0.718</td>
</tr>
<tr>
<td>RF</td>
<td>0.967</td>
<td>0.727</td>
<td>0.987</td>
<td>0.866</td>
<td>0.767</td>
<td>0.953</td>
<td>0.818</td>
<td>0.847</td>
</tr>
</tbody>
</table>

(30) Association rules are one of the best models for data mining. In this article, they used association rules to extract knowledge from transactional credit cards databases. The aim was differentiation between legitimate and illegal transactions in order to detect and prevent fraud. They extracted a set of associations rules from credit cards databases contains illegal and legal transactions. The risk analysts compare these results with results of fraud analysis processes. Fuzzy logic related data mining techniques help in reduces the risks of credit cards fraud. We can improve the performance of current fraud detection system's performance by just playing with values of the parameters. Misclassification cost is one of main problems in fraud detection systems. Standard data mining techniques cannot perform well in this case. The solution is to combine between two algorithms: genetic algorithms and scatter search. This way can improve the performance of misclassification almost 200%. The effectiveness of variables of in detecting fraud is important also, the region of card's owner is very important (27).

(31) Dempster-shafer adder, rule based filter, bayesian learner and transaction history database they are four components of fraud detection system. In this article, in innovative way they proposed new fraud detection system in the credit card by
making integrating between three approaches namely, dempster–shafer theory, bayesian learning and rule-based filtering. To give a first impression about each new transaction, dempster’s rules is applied with the rule-based component. By using bayesian learning for history database for card’s owner and fraudster, we can update the suspicion degree. The appropriate way to dealing with real world problems is fusion of multiple evidences and learning where there are complex of behavior and patterns or there is little knowledge about application domain. Studying and analyzing all previous transactions made by holders of credit cards are very necessary to found kind of patterns for each card. Based on the patterns of previous data of every card help in determining either a new transaction is legal or not. There are some of data mining techniques can help in detecting and analyzing patterns of data. Mukesh and Rajashree in their paper proposed comparison between three types of data mining techniques multi-layer perceptron, chebyshev functional link artificial neural network and decision tree with some measures like elapsed time for credit card fraud detection and classification accuracy. This study was depending on two datasets first one in Australia and another one in Germany. The results of their study in Australian data set was that maximum accuracy obtained by CFLANN is 0.8 where value of order is five. In German data set, CFLANN obtained maximum accuracy by 0.6 where value of order is five. The classification accuracy of MLP had maximum detection. Their results showed that MLP was a best between three types in fraud detection. Results also showed that FLANN has successfully used in prediction (32).

VII. CONCLUSION
This paper discuss spreading the problem of fraud around the world. Especially talk on two famous types of financial fraud are financial statement fraud and credit card fraud. Financial statement fraud is one of diffuse issues these days. Countries, governments and companies suffer from this issue. Annually every country has financial loss and of course this effect on their economic. Credit card fraud also worry banks and people because the modern technology can used by fraudsters to make new ways in fraud. So all organization try to use many steps help them to avoid the risks of fraud. The awareness for people and penalties on fraudsters and prosecution are very important but that are not enough to avoid fraud. There are need to predict new ways that fraudsters can use in the feature before they occurs. Data mining can help in studying patterns and relations in databases to give us an overview on nature of current data that help us in identify and difference odd new data. Data mining has many algorithms, everyone has special characteristics, positives and negatives. There are many data mining algorithms can used in detection and prevention financial fraud. In this paper, we discussed five popularity types of data mining algorithms for fraud prevention and detection decision tree (DT), Super Vector Machine (SVM), Logistic Regression (LR), Neural Network (NN) and Genetic Algorithms (GA). In future work there is need to expand researchers on financial fraud and data mining algorithms determining to determine which algorithms can be appropriate for every type of fraud.

REFERENCES
[23] Dr. R. Jayabrabu, Dr. V., Saravanan and Dr. J., Jebamalar Tamilselvi. A framework for fraud detection system in automated data mining using intelligent agent for better decision making process, p. 8.