A Review on Present Technologies for Fraud Detection Using Data Mining

Reena G.Bhati
Assistant Professor
Computer Science Department, TMV, PUNE, INDIA

Abstract—In Today's era of Internet lots of data are stored and transferred from one location to another. Data transferred online is most likely vulnerable to attack. As there is striking increase of fraud that is leading to loss of many billions of dollars worldwide every year; various modern ways in detecting fraud are regularly proposed and applied to several business fields. The main task of Fraud detection is to observe the actions of tons of users to detect unwanted behavior. To detect this various kinds of data mining methods have been proposed and implemented to lessen down the attacks. In this paper a deep literature survey is presented on different techniques for fraud detection applying data mining techniques.

Keywords—Fraud, Fraud detection systems (FDSs), Areas of fraud E-commerce systems, Credit card system

I. INTRODUCTION

The Association of Certified Fraud Examiners outlined fraud as “the use of one’s occupation for personal enrichment through the deliberate misuse or application of the using organization’s resources or assets [1].” Within the technological systems, dishonorable actions have happened in several areas of everyday life like telecommunication networks, mobile communications, on-line banking, & Ecommerce.

There is exponential growth in fraud as the modern technology is growing fast, leading to substantial losses to the companies. There is need to explore more the fraud detection issue. It involves distinguishing fraud as early as possible once it's been detected.

Fraud detection strategies are ceaselessly trained to defend criminals in altering to their methods. The event of recent fraud detection strategies is formed tougher because of the severe limitation of the exchange of thoughts in fraud detection. Data sets don't seem to be created available and results are typically not disclosed to the general public.

The systems should be capable of detecting the fraud cases from obtainable large data sets like logged data & user behavior. Now a days various techniques are implemented for fraud detection like data mining, statistics, and computing. Fraud is exposed from anomalies in data & patterns. It is quite hard to be certain about the legitimacy of & intention behind an application or transaction. To keep the electronic commerce system secure alone Fraud prevention systems (FPSs) are not adequate. FDS and FPS together could be efficient solution to fix electronic commerce systems. Still, this systems lacks as there are effects & disputes that block the performance of FDSs, like concept drift, skewed distribution, large amount of data etc. The survey done by IC3 depicts the number of complaints and the losses done till 2017

2017 Overall Statistics

There are two type of Credit card fraud: offline fraud & online fraud. Offline fraud is done by using a stolen physical card. This can be stopped by blocking the card from the banks before they are misused. For doing online fraud web, phone shopping is used. For doing this only card details are needed and manual signature and card imprint are not required at the time of purchase.

The remainder of this paper is organized as follows. We provide an overview of various fraud detection methods in...
Section II, and discussion on this system is given in Section III and conclusion is then given in Section IV.

II. LITERATURE SURVEY

One to prevent credit card fraud, research works were carried with peculiar a focus on data mining & neural networks. Authors in [2] have presented a novel method using neural network for credit card fraud detection. A detection method is developed which is trained and it is checked on a huge sample of labeled credit card account transactions. The study proved that due to its power to discover fraudulent patterns on credit card accounts, it is easy to accomplish a reduction of from 20% to 40% in total fraud losses, at importantly reduced caseload for human review. A database mining model named CARDWAT CH is presented by authors in [3] which can be implemented for credit card fraud detection.

Neural network is used to train the specific historical used data & then neural network model is generated in the system. It was developed to notice fraudulence. A credit card fraud detection system [4] is proposed using Bayesian & neural network methods to find out models of fraudulent credit card transactions. Authors in [5],[6] found that there are two principal causes for the complexity of credit card fraud detection i.e. inclined distribution of data & mix of legitimate & fraudulent transactions. Fraud density of real transactions data is used a confidence values & generate the weighed fraud score to cut down the number of misdetections. All above-mentioned approaches do not concern to convert the training data into confidence value before putting into neural network. Furthermore, a fixed threshold is set to find abnormal and normal spending pattern in the above-mentioned approaches without concerning the cost problem derived from false positive and false negative. The Threshold cannot be adjusted dynamically based on frequency of fraudulent either. NNM combining with confidence and ROC analysis technology for fraud detection is introduced in detail by the authors. Fraud detection system is the next layer of protection; which is also the concern of this paper. Authors in [7] explains the ways in which fraud can be used versus organizations. They also evaluated the limitations and drawbacks of current systems & methods to detect & prevent fraud. Fraud detection tries to discover and identify fraudulent activities as they enter the systems and report them to a system administrator [8]. In previous years, manual fraud audit techniques such as discovery sampling have been used to detect fraud, such as in [9].

Authors in [10] proposed a effective computerized & automated FDS. Because of complicated & time consuming techniques there is need of system which can help the fraud detection. The developed FDS capabilities were limited which was the drawback of the system. Understanding these drawbacks more complex FDS with more accurate and precise data mining methods were developed for efficient fraud detection [11]. The used method involves AI, machine learning and statistical methods. These methods helps to collect and identify the needed information from the huge databases. Implementing this system will give following advantages: capturing fraud patterns from the data, stipulation of fraud likelihood for every case, accordingly that effort in exploring shady cases can be placed, and finding new fraud types which have not been recorded yet[12]. The main categories in data mining are clustering, outlier detection, classification, regression, prediction and visualization. A specific technique is supported by each of these methods. Like SVM & neural network are used for data mining classification and k means for data clustering method.

Various mechanisms are implemented for outlier detection [13], authors in this paper use unsupervised learning approach. Mostly they obtained results of unsupervised learning will lead to improve the future decisions. In these strategies there's no need of antecedent knowledge of fraudulent & non-fraudulent transactions in database. It detect if any uncommon behavior happens. All the models in supervised strategies are aimed to isolate among fraudulent & non-fraudulent manner in order that new readings are often assigned to classes. correct identification of fraudulent transactions in database is needed in supervised strategies and this could solely be used to detect the frauds that have occur earlier that why most of the researchers prefer using unsupervised strategies as the undiscovered forms of fraud are often detected.

In [14] researchers have presented a unsupervised credit card fraud detection which utilizes outlier detection technique. The potential fraud cases which can be named as outliers are Abnormal spending behavior & frequency of transactions. A deep survey is presented in [15] for fraud detection methods. The authors have distinct the frauduster, its types and subtypes and they also posed the nature of data evidence collected from the industries. In [16] authors stated the application of data mining is anomaly detection. Authors in[17] depicted ready made data mining methods which can be imposed to detect intrusion. Authors in [18] surveyed all the network intrusion detection systems; they also provide techniques which are potentially deployed to notice the fraud.

Many data mining methods are explained in [19], [20], [21], [22]. In [23], [24], [25] clustering is the only tool which is used with complex data mining. Authors in [26] proposed clustering visualization techniques for financial fraud detection. K-means & its variations for outlier detection techniques are explained in [27] most cases where standalone clustering methods are being used make use of k-means and its variations for outlier detection. In most cases, Euclidian distance is getting used because the dissimilarity metric. [19] Implements k-means with the intent of identifying fraudulent refunds within a telecommunication company with fraudulent transactions being considered outliers. K-means is implemented in [20] for fraud filtering during an audit. Similar features are sorted together & small clusters are flagged for further processing.
Based on ANOVA analysis k means is implemented on newly added attributes, authors in [21] identified 3 fraud schemes related to purchasing, double payment, changing purchasing order after release. To detect the early symptoms of insider trading [22] employs k-means n option markets before any news release. A technique proposed which mines the transaction data using the text document in monetary vector [28]. Computed monetary vectors are either clustered via k-means or projected to a histogram. Another case group standing out consists of clustering methods used for training classifiers. Due to proliferation of enterprise resource planning systems and an ever growing amount of available data to be studied, manually labeling training data for various classifiers has become unfeasible in many cases. In these situations a clustering technique is first used on the uncategorized data in order to automatically split it into meaningful categories.

Each cluster/category is labeled (usually manually) and then classifiers are being trained on each cluster/category. The majority of papers found are mostly using classifiers implemented on decision trees, neural networks & SVMs. Authors in [29] manually divides the data into several large classes from the dataset and conducted raw class analysis on it and also perform k mean with Euclidian distance as dissimilarity metric on each class. Subclasses are generated by this local clustering process with comparatively stable sizes within each main class, sub-classes used later on for training a SVM classifier. The proposed system is tested on neural networks and decision trees and the results depict this method produces higher prediction on rare classes as compare to other systems.

Authors in [24] generates new composite attributes from transaction data & uses them in k-means clustering to divide transactions into suspicious and unsuspicious, most being unsuspicious. For training of classifiers like neural networks & decision trees on identified cluster the full set attributes are used. [30] Discerns sorts of behavior alters from various fraudsters with the help of k-means clustering method. Later on, C4.5 decision trees are implemented for getting the rules of the labeled clusters. [31] Performs k-means on insurance data and trains a NB classifier on each found cluster. In [32] authors propose a method to detect fraud to look like normal activities in domains with relationships like accounting fraud detection for rating & investment and attacks on corporate networks, health care insurance fraud. To classify k-means classifier is used. There are cases where clustering methods are implemented to group already flagged, possible fraudulent entries by classifiers.

The clustering goal in this situation is to define taxonomy of the already identified fraud entries in order to implement counter measures for each found fraud category. In some situations, some categories may be even found to contain legitimate data, wrongly labeled by the classifier due to insufficient training to such cases. Authors in [33] proposed a system to find the errors in payments in insurance claims by implementing ranked discordant clustering on entities flagged as fraudulent via SVM. In [34] the author calculates financial ratios from the statements from companies. Here a pre defined map neural network is applied with financial ratios as its input vector, along with that k-means is performed on self-organizing map node vector...
III. SUMMARY

Many Data mining approaches are studied and describes in the literature survey for fraud detection in the past few years, this review will help the researcher to gain basic idea of various methods & techniques for fraud detection. Although enormous research has been done using different methods and algorithms, many hybrid methods are mainly used to furnish better results and helps to overcome the drawbacks of other methods. Each day a new attack is witness and hence there is need of modern approaches for detecting such frauds from the datasets. Combination of various methods can help to detect the fraud at early stage.

IV. CONCLUSION

This paper presents a deep literature survey on fraud detection techniques in data mining. Various methods and their advantages & disadvantages are explained & studied. The systems studied are effective for several fraud not still persist some problems for some methods. The developed systems need to stay upgraded where they lack, and they have to keep working now new fraud methods. This is costly.

REFERENCES


