Data Mining Techniques for Financial Fraud Detection

[1]Aditi Satpute, [2]Anuj Shenoy

[1]Department of Computer Science, Fergusson College, Pune, Maharashtra, India
[2]Department of Computer Science , Kaveri College of Arts,Science,Commerce, Pune, Maharahtra, India

ABSTRACT

This presented article mainly circumspect the idea of the increasing number of frauds in recent times. Frauds can deliberately cause accident for payout or intentional losses. With all the different methods of fraud, detection still becomes an upheaval task. In this article, we shed light on the various frauds that take account and the fraud detection techniques used with the help of data mining. The broad-based definition of data mining is 'processes and activities designed to obtain and evaluate data to extract useful information'. It is definitely something very important when it comes to detection, which can result in taking immediate action to minimize cost.

Keywords: Data mining, Financial Fraud, Fraud detection techniques

I. INTRODUCTION

Fraud implies intentional deception to gain impetus through unfair means .According to the Kroll Annual Global Fraud & Risk report 92% of the companies experienced a fraud incident in 2017.It also reveals that financial sector is one of the top sectors that are vulnerable to fraud causing a loss of billions every year.

Fraud detection means identifying the fraud or predicting an expected fraud. There are various ways to detect a fraud, data mining being the most popular among them. More and more data is being generated in every aspect we can think of . On most transactions we do there is some sort of data being downloaded into the database. Organizations are storing, processing and analysing data than any time in the history. It is a trend that is going to continue to grow for a very long time.

Every fraud has a specific pattern. Data mining techniques are used to identify those patterns and give results. In terms of data mining it can be classified as fraudulent and non-fraudulent behaviour. Assuming non- fraudulent behaviour as normal and having the rest as exception, we can identify fraudulent behaviour.

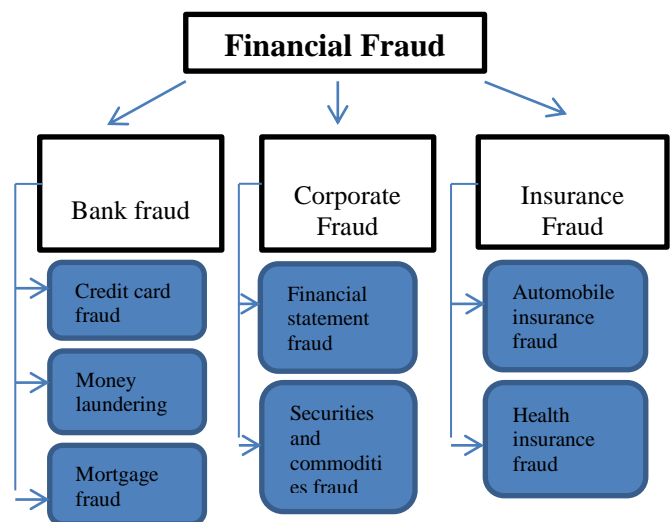Financial fraud can be mainly listed into four three types as showed in figure1.

Figure 1. A flowchart showing different types of fraud under financial fraud

This paper addresses the most common techniques that are used to eliminate the above frauds. The techniques we have discussed are neural networks, text mining, decision tree, Bayesian belief network, logistic model, text mining, Outlier detection, fuzzy logic.

We have also relegated the frauds according to the techniques.

II. TECHNIQUES USED FOR FRAUD DETECTION

The proposed context of financial fraud detection ,we hereby discuss the most efficient techniques-

A. Logistic Model

This model is suitable for classification.
The idea of linear model is to make linear regression produce possibilities. It calculates linear function and then a threshold in order to classify. This model could be binomial, ordinal or multinomial. The result of this model is logistic regression, a popular regression technique.

B. Decision Tree

Decision trees are one of the easiest to use. They are used for prediction and classification. Classification rules are represented on the path from root to leaf. They are best suited for choice model or selection model algorithm. Another use of decision trees is for calculating conditional probabilities. Decision trees extract information in human-understandable form. The rules used for extracting this information are *if* condition1 *and* condition2 *and* condition3 *then* outcome expressions, they explain the decisions that lead to the prediction. Decision trees are effective as they clearly layout the problem so that all options are challenged. Also, they allow us to analyse all the possible consequences of decision.

C. Outlier Detection

Outliers detection is also known as anomaly detection. It is used to identify transactions that are showing abnormal behavior. To do that, outlier detection, such a density based model ,can help us to detect anomalies. Patterns can be used to identify fraudulent behavior by using statistical methods. As each fraud has a pattern they can be identified using this method. e.g. Spending on expensive items .If not typical for an account, it can be considered for outliers detection. Standard statistical methods are used in outlier detection to observe how two variables interact to ascertain normal behavior. Although, it's been observed that such techniques are used where anomalies are less in number.

D. Bayesian Belief Network

Bayesian belief network is probably the most widely used data mining technique as well as the most popular technique used in fraud detection. It is based on conditional probability. It helps in representing a set of variables and their conditional dependencies via Directed Acyclic Graph (DAG).The nodes of the DAG may be observable quantities, latent variables, unknown parameters or hypotheses. Conditional dependencies are shown on edges. It also works on the causality principle i.e. it takes in consideration of the prior event. The variants of Bayesian probability differ mainly in their interpretation and construction of the prior probability. This prior probability is also known as marginal depending on the probability direction.

The probability of event 'A' happening given that event 'B' has already happened, the formula for Bayesian can be stated as:

$P(A|B)=P(B|A)P(A)/P(B)$
where,
- ✓ P(A|B) represents Posterior- Probability of our hypothesis being true, give the data is collected.
- ✓ P(B|A) represents Likelihood- Probability of collecting this data when our hypothesis is true.

- ✓ P(A) represents Prior – Probability of our hypothesis being true before data collection.
- ✓ P(B) represents Marginal – What is the probability of collecting this data under all possible hypothesis?

Example –Getting cards out of a deck
Probability of getting 2 queens out of a deck of cards.

Probability (2 queens)=4/52∗3/51≈0.45%
Prior Probability: 4 queens in the set of 52 cards
Second Probability: Only 3 queens in a set of 51 cards

In spite of its complexity it's one of the most reliable methods used for detection.

E. Text Mining

Generally, text is counted as unstructured data which must be converted into structured data before applying any data mining techniques such as classification or clustering in order to detect fraudulent content. Text mining is capable of studying plain text, which helps give a different approach to the problem. It is typically used in clustering and anomaly detection. The flow chart below explains the proposed text mining approach for financial fraud detection.
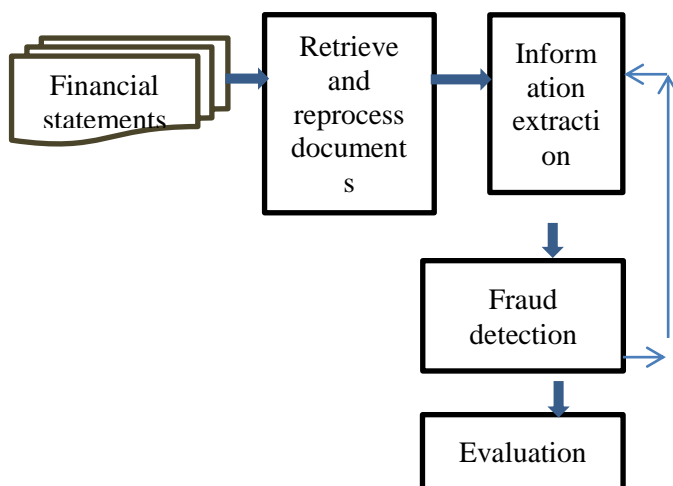

Figure 2. A flowchart showing different types of fraud under financial fraud

F. Neural Network

Neural networks are systems, which are computing systems inspired by the biological neural networks. An assembly of connected units is called as artificial neurons.Fraud detection using neural network is totally

a replica of the working of a human brain. Neural network method gives the computer the technological way to think like a brain. Neural network uses this method to learn experiences and knowledge gained in daily life to take the decisions, which will imply the evaluation for a fraud.
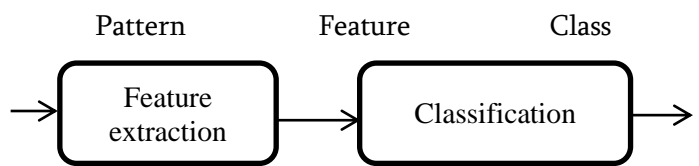

Figure 3. Pattern recognition

In Figure 3, it's shown that when a pattern is detected then it is considered with its features which is further used for the classification. Now, this classification is used for detecting the actual fraud.

G. Fuzzy Logic

Fuzzy logic is quite similar method to human reasoning. In this method it imitates the human way of thinking that involves all possibilities between digital values YES or NO. This approach is based on the degrees of truth. Decision makers always look for some patterns or groups, while studying a particular data, when it does not suggest a clear answer. This logic is also called as cluster analysis which is used for finding groups in a particular data. Fuzzy logic or clustering allows for some ambiguity in the data where the solution might not be clear in that cluster. Fuzzy logic is mainly used to model partial categorizations. Although identifying the number of clusters is very tedious task, with the help of fuzzy-c means algorithm, which is one of the most popular objective function, it becomes easier. As the different methods of fuzzy logic are becoming ubiquitous,

these can be used further in various other applications.

III. CLASSIFICATION

We have further classified all the techniques according to the type of fraud to get a more clear picture.

Table 1. Classification of techniques according to the types of fraud

Sr. no	Technique	Type of fraud
1	Logistic Models	Credit card fraud, Insurance fraud, Financial statement fraud
2	Neural Network	Financial statement fraud
3	Bayesian Belief Network	Insurance fraud, Financial statement fraud, Corporate fraud
4	Outlier Detection	Credit card fraud
5	Text Mining	Financial statement fraud
6	Decision Tree	Credit card fraud, Financial statement fraud
7	Fuzzy Logic	Credit card fraud

IV. CONCLUSION

In this conceptual paper, we discussed the data mining approach for detection of financial fraud. We furthermore talked over the various frauds that are carried out with their fraudulent behavior and some effective methods to detect these frauds. Financial Fraud is a major concern in the world today.

Information received in the annual reports should be carefully investigated by decision makers so as to take measures in identifying any potential threats of a fraud. The only flaw that these methods find while detecting a fraud is that they need some actual data that has been implemented on the organisation before they can actually come to a decisionUnderstanding the problems being the most important aspect her followed by lexical modeling and with the use of technology we can predict or detect a fraud.

V. REFERENCES

[1]. Prof. Gupta Rajan, Gill N.S. 2012 "Data Mining Techniques–A Key for detection of financial statement fraud

[2]. D.WHITLEY,"Genetic Algorithm And Neural Network."2003

[3]. H.C. Koh, C.K. Low, Going concern prediction using data mining technique 2004

[4]. Maes, S., Tuyls, K., and Vanschoenwinkel, B., Machine Learning Techniques For Fraud Detection. 2000

[5]. Hoogs Bethany, Thomas Kiehl, Christina Lacomb and DenizSenturk(2007). A Genetic Algorithm Approach to Detecting Temporal Patterns Indicative Of Financial Statement Fraud

[6]. P.Ravisankar, V. Ravi, G.RaghavaRao, I., Bose, Detection of financial statement fraud and feature selection using data mining techniques,Decision Support System. 2011

[7]. Jenson, F.V., An Introduction To Bayesian Networks. 1998