

Data Mining in Fraud Detection and Audit

Authors: Alwin Peter, Amala C Mathew, Prachi Kothari, Nidhin Ouseph, Divya M, Pushpitha S, Sharanya V, Sachin M N, Anto Mathachan - Research Scholar's, CHRIST (Deemed to be University)

Guided by : Dr. Kavitha Jayakumar, Associate Professor, CHRIST (Deemed to be University)

Abstract: *The increased technological base and development has given rise to new possibilities as well as challenges. The biggest challenge that any organization and businesses faces is fraud because of new technologies and is difficult to recognize than the conventional form of this crime. Therefore, it becomes very important to develop the techniques for detecting fraud to protect the organization from such a crime. Data mining is the most effective technology that has been developed and implemented for the purpose of detecting fraud. This paper focuses on the different tools and techniques of data mining which are employed for detecting the fraud.*

Key words: *Data mining, Innovation, Fraud detection.*

Introduction

The headway in innovation and communication has made new open doors for submitting deceitful acts. These demonstrations force genuine risk to associations on the budgetary, operational and mental dimensions. Notwithstanding the money related misfortunes, fraud can staggeringly affect the association's reputation as a whole. In this way, associations attempt to implement various techniques to recognize and avert fraud. Among those strategies is data mining. This paper investigates a portion of the data mining techniques utilized for detecting the fraud in various fields

Definition of data mining

Agreeing to Kamber and Han (Kamber, 2006), data mining, which is accepted as the utmost significant phase in the course of knowledge detection, means mining or exploring information from vast amounts of data.

Data mining is determining fresh, dependable, and traceable information through artificial intelligence and the statistical tools in great volume of data (Ma, 2009). The methods used in ascertaining knowledge through databases are (Deshpande & ThakarePu, 2010):

Data Selection: Data linked to analysis and decision-making are divided from further data.

Data pre-processing: cleaning, processing, and integrating the data are executed.

Data conversion: The nominated data are transformed in a manner appropriate for data mining.

Data mining: At this phase, smart methods and decision making tactics are used to mine possibly useful forms.

Interpretation and evaluation: At this point, curious designs which symbolize information are recognized based on the measures taken and the recently exposed knowledge is made open to the handler. It is essential to use imagining in order to support the handler on this point.

Objective

The objective of this paper is to identify the techniques of data mining that supports in detecting the fraud and audit.

Review of Literature

Data mining is the process of seeking interesting or valuable information within large databases. The novelty, and the reason that a new term has been coined to describe the activity, has its origin primarily in the large sizes of modern databases. Data mining has promise, but there are many difficulties associated with it. It is not to be entered into lightly or in ignorance of the obstacles. Perhaps there are similarities to meta-analysis, in that it is easy to carry out a poor analysis, and very hard to carry out a good one (Hand, Blunt, Kelly, & Adams, 2000).

Data mining is the process in which there is analysis of data from different angle and perspectives and summarizing the same data into the relevant information. This kind of information could be utilized to increase the revenue, cutting the costs or both. Data mining uses artificial intelligence techniques, neural networks, and advanced statistical tools to reveal trends, patterns, and relationships, which might otherwise have remained undetected. In contrast to an expert system data mining attempts to discover hidden rules underlying the data. Advanced data mining techniques and the structure of neural networks are combined successfully to gain a record of the hidden forgery with the low wrong warning rate. Data mining to reduce the internal fraud. (Akbarpour & Mohammad Pourzarandi, 2015).

Now a days auditing is one of the demanding and important task for the detection of fraudulent financial statements. Many researchers have said that FFS have increased these days and there is the need to curb this. Management fraud refers to those frauds which has been done deliberately to causes damage to investors, creditors through deceptive financial statements.

This paper mainly focuses on the usefulness of Data Mining (DM) arrangement techniques in finding out those firms that issue fraudulent financial statements (FFS) and identification of factors leading to FFS. Data mining is the process of identifying the hidden designs of data from different viewpoints and arranging them into information, which are then collected and assembled in common areas. Detection of financial fraud is quite difficult only with the help of normal auditing techniques. Hence in this paper the researchers have considered the above three technique and have applied these for 76 Greek manufacturing firms. Auditor have checked all the firms

The factors identified for the FFS are: Financial distress (Cogger.K, 1998) (Stice, 1991) (Kinney .W & McDaniel, 1989), , Companies which are not able to achieve comparable results to past performances are engaged in unethical activities to preserve previous styles (Stice J. A., 1991), many researchers also suggest that management may manipulate inventories (Stice J. A., 1991).

When it comes to these three techniques, base of their performance Bayesian Belief Networks achieved the best performance followed by Neural Networks and Decision Trees. Therefore we can conclude that Bayesian Belief Networks achieved the outstanding accuracy, Neural Networks achieved satisfactory while Decision Trees achieved low accuracy in identifying the FFS.

Prior studies explained about data mining fraud detection in the area of health care. These types of fraud happen in health care industry is due to the inappropriate payment of third party and insurance companies. The auditors feel the

difficulty in finding the fraud detection through traditional method. Knowledge Discovery from Databases process is the basis of data mining which facilitates the auditors in fraud detection. Supervised and unsupervised data approaches are some data mining techniques which helps in detecting fraud and abuse in health care. Studies also suggest a research on the applicability of data mining techniques in low and middle income countries. This is because of poor IT based auditing system that leads more problems in data mining. (Joudaki, et al., 2015)

Fraud related to finance is very common and it is a widespread issue that had influenced the economy around the world. Data mining being a standout aspect and the most successful and integral asset for identifying budgetary misrepresentation has been utilized broadly by the business investigators and analysts. This review paper formalizes distinctive sorts of financial frauds, outlines the compelling properties for recognizing each kind of misrepresentation and present the most recent improvements on the utilization of data mining as a source for money related frauds. The present study examines all distributed research work in the field of money related aspects for the period of 7 years beginning from 2009. The point is to help analysts in distinguishing and identify the factors and data mining systems by giving the overall view of research stages for identification of finance related misrepresentation. (Bhardwaj, 2016)

Data Mining: An Improved Approach for Fraud Detection

Data mining is considered as widely referred concept which helps in data pattern analysis. This paper focuses on various types of frauds and helps in their detection by using various techniques in a simple way. Data mining is a way in which it guides in understanding patterns of a data from huge sets. There a number of techniques involved in data mining like classification, clustering, neural networks, and regression models. Another important area to be discussed is fraud detection. Fraud detection involves in identification of fraud where no previous disbelief that exists.

The primary objective of the paper is to identify the patterns of fraud using the various techniques of data mining.

Generally there are four types of task in Data mining:

- 1) Classification – the data is arranged into specific groups with the help of different algorithms.
- 2) Clustering – It is similar to classification but specific classes do not appear here, but they bring together similar objects in one class of group.
- 3) Regression – helps in obtaining data with less error.
- 4) Association rule – helps in identifying relationship within the objects of data.

When we speak about cellular networks in data mining it involves inspection of the call details, billing data of users which helps in building modules and also uses data mining techniques to detect fraudulent nature. Under Swap card fraud detection it recognizes the transactions that are fake and unrelated to the databases and classifies them into real and fraud transactions. They are mostly based on statistical practices and they are considered to have theoretical background. Data mining with the help of clustering brings together those data which are same in nature and behaviour. It identifies the odd behaviour of the group of clusters. Most of the organisations these days prefer use data mining techniques for fraud detection and also for tracking the claims pertaining to fraud. For this purpose the organisations obtain information from other parties engaged and then look from the organisation point of view. The techniques usually used are decision trees, regression and neural networks.

Selected fraud detection techniques

This section discuss in detail about the selection of data mining techniques for mobile telecommunication, credit card and medical insurance fraud detection.

A. Mobile Telecommunication Fraud Detection

Telecommunication fraud has caused huge losses for telecommunication industries. According to the Communication Fraud Control Association (CFCA). There are many forms of this fraud that can be classified under two types. The first type is subscription fraud in which fraudsters obtain telecommunication accounts without paying for the service. The second type is called superimposed fraud, in which the fraudster takes over the account of a legitimate subscriber and all the call charges are billed to that legitimate subscriber. It includes cellular cloning, calling card theft and cellular handset theft . This form of telecommunication fraud is considered to be the most challenging type for telecommunication companies. Therefore, most telecommunication companies, have dedicated research to develop effective detection methods for this type of telecommunication fraud .

B. Credit Card Fraud Detection

Credit card fraud detection is the process of monitoring the behavior of the customers transaction level through a period of time.

1) Types of Credit Card Fraud:

The first type which is the most common is the application fraud. The individual will falsify an application to acquire a credit card. The individual will give false information about his/her financial status in order to receive a credit card .

The second type is assumed identity. Assuming someone's identity has been in the long-run form for credit card fraud. The individual will falsify a name with a temporary address .

The third type is financial fraud which happens when an individual wishes to gain more credit than he/she currently has. They will apply for a credit card under their own name, but the information regarding their financial status will be false .

The fourth is skimming technology. Magnetic card skimming is a small handheld device with the sole purpose of collecting and storing the information on any credit card.

The fifth type is never received issue. This type of credit card fraud involves the theft of the card while still in transit. This involves the theft of the card from the holder's mail.

2) Data Mining Techniques for Credit Card Fraud:

The first technique is the Peer Group Analysis. This type of analysis is an unsupervised method for monitoring customer behaviors over a period of time. For each individual that has a credit card account, a Peer Group" of accounts is created that exhibit similar behavior. As time goes by, the behavior of an account is tracked by those

accounts in its peer group. If an account has subsequent behavior which deviates strongly from its peer group is thus considered to have behaved anomalously and is flagged as a potential fraudulent.

The second technique is the Break-point Analysis. This technique distinguishes spending activities supported from transaction information in a single account. Current transactions are matched up with prior spending activities to spot features, such as rapid spending and an increase in the level of spending, which would not essentially be captured by outlier detection.

C. Medical Insurance Fraud Detection:

Data mining empowers a variety of insurance providers with the ability to predict which claims are fraudulent so they can effectively target their resources and recoup significant amounts of money. Data mining helps medical insurance company to focus, for example, on claims with high percentage of recoverable fraud, isolate factors which indicates a payment request has a high probability of fraudulence, develop rules to use them to flag only claims likely to be fraudulent, and ensure adjusters could review claims that are not only likely to be fraudulent but also have the greatest adjustment potential.

Benefits of Data Mining in the Audit Process

The impact of these technological advances has been significant. Access to whole data base or demand of certain key components from the database can be made by an auditor. The structure of data base is not changed by data mining; however it allows the advancement of investigation by changing over data from various systems to common platform without much manual intervention. Examination of information is made after which the relevant information is replicated to electronic document. To a degree, investigations are normally time obliged or limited to budgets. It is much easier to determine accounting errors earlier, hence the focus was narrower. Technological advancement can be well sensed as several analyses which were done in several weeks can now be completed in hours. Rather than reviewing summaries software comprising data mining can penetrate down to transactional level and select single transactions from million. Presently with advances in data mining program there is 100% confidence among auditors as such instances can be identified in minutes. Clearly, these are essential tests however more complex ones.

The capacity to identify certain patterns turns out to be lot simpler when database can be completely analyzed. Calculations such as accounts reconstruction, early accounting, and statistical information can now be calculated easily. Data mining would most likely focus on certain files, where reserves had been increased in a specific date. In this way, more relevant files can be selected among the rest of the files. There is lot more prominent ability to detect fraudulent activity. For instances, insurers can identify whether certain cost billings are excessive or billed multiple times. Data mining enables certain functions to be automated upgrades in data mining have enabled the scope of audits to become much wider. Nowadays, auditors can review their own data to identify things such as claims aging , duplicate payments and claims overload. Audit work has developed because of enhancements in data mining software and computer hardware and applications to insurance industry continues to grow. This has helped users to new era of reinsurance auditing.

Conclusion

Fraud remains a test for organizations and associations in numerous fields. Data mining is a successful strategy for identifying different sorts of misrepresentation including portable media transmission, debit and credit card and insurance frauds just as distinguishing interruption to PC frameworks. This paper has displayed just a choice of the different data mining fraud discovery systems utilized in various fields. A portion of those strategies have continued and ended up being fruitful, while others are currently improvement and upgrade to all the more likely apply to new fraudulent acts. All things considered, it is the organization alone who experiences the outcomes of misrepresentation, yet every one of the people and partners identified with that association will be exploited people. In this manner, organization are altogether responsible for learning the prescribed procedures and picking the best technique that coordinates their requirements so as to defend against misrepresentation and fraud.

References

- Akbarpour, G., & Mohammad Pourzarandi, M. E. (2015). Applying the Data Mining Technique for Improving Internal Controls in Financial and Credit Institutes. *Journal of technologies*.
- Bhardwaj, D. (2016). Analysis of data mining trends, applications, benefits and issues. *International Journal of computer science and communication*.
- Cogger, K. F. & (1998). Neural network detection of management fraud using published financial data. . *international journal of intelligent systems in accounting, finance & management* , 21-24.
- Deshpande, S. P., & ThakarePu, V. M. (2010). Data Mining System and Applications: A Review. *International Journal of Distributed and Parallel systems*, 32-44.
- Hand, D. J., Blunt, G., Kelly, M. G., & Adams, N. M. (2000). Data Mining for Fun and Profit. *Institute of Mathematical Statistics*, 111-126.
- Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., & Arab, M. (2015). Using Data mining to detect health care fraud and abuse: a review of literature. *Global Journal Of Social science*, 194-202.
- Kamber, J. H. (2006). *Data Mining: Concepts and Technique*.
- Kinney .W & McDaniel, L. (1989). Characteristics of firms corecting previously reported quarterly earnings . *journal of accounting and economics*, 71-93.
- Ma, B. F. (2009). Data Mining Technology and Its Application in CRM of Commercial Banks. 243-246.
- Stice. (1991). Using financial and market information to identify pre-engagement market factors associated with lawsuits against auditors. *The Accoutning Review*, 516-533.
- Stice, J. A. (1991). Lessons to be learned-ZZZZBEST Regina and Lincoln savings . *The CPA Journal(April)*, 52-53.