

INTERNATIONAL JOURNAL FOR ENGINEERING APPLICATIONS AND TECHNOLOGY

DATA MINING TECHNIQUES FOR FRAUD DETECTION

Akshay S. Nakshane¹, Pranoti R. Patil², Dipak S. Shirsode³, Dhiraj D. Shirbhate⁴

¹Student, C.S.E.Dept, J.D.I.E.T.Yavatmal, Maharashtra, India, Akshay.nakshane123@gmail.com

²Student, C.S.E.Dept, J.D.I.E.T.Yavatmal, Maharashtra, India, pranoti261@gmail.com

³Student, C.S.E.Dept, J.D.I.E.T.Yavatmal, Maharashtra, India, shirsode50@gmail.com

⁴Asst. Prof. C.S.E.Dept, J.D.I.E.T.Yavatmal, Maharashtra, India, shirbhate.dhiraj@gmail.com

Abstract:

Due to the increasing of fraud which results in loss of billions of dollars in worldwide each year several modern techniques detecting fraud are continually evolved and applied to many business fields. Fraud detection involves monitoring the behavior of populations of users with their financial activity in order to detect, estimate, or avoid undesirable behavior. Undesirable behavior is a broad term including misbehavior; fraud intrusion, and account defaulting. Although prevention is the best way to decrease frauds, fraudsters are adaptive and will usually find ways to circumvent such measures. Detecting fraud is essential once prevention mechanism has failed. Several data mining algorithms have been developed that allow one to extract knowledge from a large amount of data like fraudulent financial statements to detect FSF. This paper present the concept of data mining and current techniques used in credit card fraud detection, telecommunication fraud detection, money laundering fraud and computer intrusion detection. The goal of this paper is to provide encyclopedic review of data mining and different techniques to detect frauds.

Keywords: Fraud detection, data mining, computer intrusion, neural network, knowledge discovery.

I. INTRODUCTION

Nowadays, it is major problem for financial institution, stolen credit cards, falsifying, misleading cheques, accounts and more to prevent credit card fraud activity and many of them are formed frauded activity so after storage of data the data mining technique are utilize for extracting useful information from large amount of data

After long research and development data mining many modern techniques based on machine learning sequence alignment genetic programming, artificial intelligence, neural network, data mining etc. Has evolved in detecting various credit card fraud transactions. For ambiguous credit card transaction detection Data mining is a technique which take data and construct representation of reality in form of market. "The word mining refers to a process which we can take interesting knowledge of data and pattern from large amount of data." Data mining is the process sorting through large datasets to identify pattern and establish relationship to solve problem via. Data analysis.

II. What is Data Mining?

The word "Mining" refers to the extraction of precious things like minerals from the earth. However, data mining is the process by which we can extract interesting patterns and also knowledge from large amounts of data. The data mining is a relatively new field of study and research and has to be generated large interests among business communities. Data mining is the important part of business intelligence that deals with how an organization uses analyses, manages and stores data it collects from various sources to make better

decisions. Data mining has the answers to all these questions. Data mining can help the organizations to useful insights into its business from the data it has collected over the years and take better decisions.

A. Step involved in data mining: The following Steps involved process of data mining and various stages through which we extract data from large amount of data.

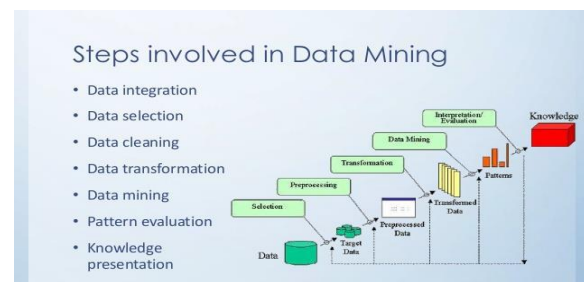


Fig.1 Steps involved in data mining

1. Data Integration:

In this all the data are collected and integrated from all the different sources.

2. Data Selection:

In the first step we may not use all the data we have collected. So in this step we select only those data which we think useful for data mining.

3. Data Cleaning:

We have collected the data that are not clean and it may contain errors, missing values, noisy or inconsistent

data. So we need to apply different techniques to get rid of such anomalies.

4. Data Transformation:

After cleaning the data that are not ready for mining as we need to transform them into forms appropriate for mining. Smoothing, aggregation, normalization etc. are the techniques that used to accomplish these data.

5. Data Mining:

We are ready for apply the data mining techniques on the data to discover the interesting patterns. Techniques such as clustering and association.

B. Major tasks of Data mining:

□ **Summarization:** It is the abstraction of data. A set of relevant data is abstracted and summarized resulting in a smaller set which gives a general overview of data. For example, the calls from long distance customer can be summarized in to total minutes, total calls, total spending etc instead of detailed calls. Similarly the calls can be summarized in to such as local calls, STD calls, ISD calls etc.

□ **Clustering:** Clustering is identifying similar groups from unstructured data. Once the clusters are decided, the objects are labeled their corresponding clusters, and common features of the objects in a cluster are summarized to form a class description.

□ **Classification:** In the classification, the pre-processing of data, designing modeling, learning/feature selection, and validation /evaluation. Classification predicts categorical continuous valued functions. For a training set, an object is given in which every object is represented by attributes along with the class. By analyzing the relationship between attributes and class of the objects in the training set, the classification model can be constructed. Such classification model can be used to classify future objects and develop a better understanding of the classes of the objects in the database.

□ **Regression:** Regression is a process of finding function with minimal error to model data. It is a statistical methodology that is most often used for numeric prediction. Regression analysis is largely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Regression analyses are also used to understand which among the independent variables are related to the dependent variable.

□ **Data Mining Process:** The data mining is defined as a process used to extract usable data from a larger set of any raw data. It implies analyzing data patterns in large batches of data using one or more software. Knowledge Discovery in Data (KDD) is also known as data mining.

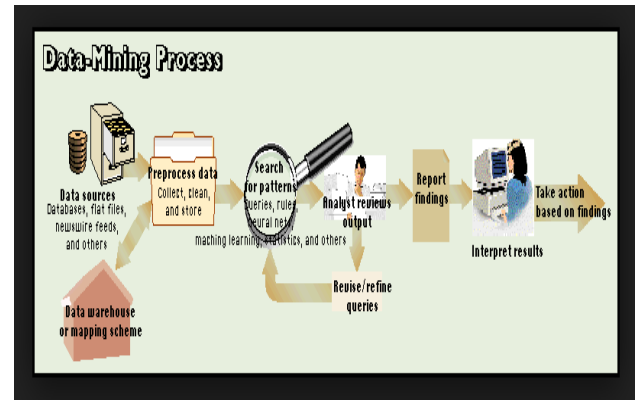


Fig.2 Process of Data Mining Techniques.

III. Types of fraud and its Detection Techniques

Fraud detection requires different techniques that are unique to this application area. Data mining helps to quickly detect fraud and take immediate action to minimize costs. An important step in fraud detection is to identify factors that can cause fraud.

The ACFE (Association of Certified Fraud Examiner) defines fraud as "the use for personal enhance through the conscious misuse or application of the employing organization's resources or assets " In the technological systems, the activities which are fraudulent have occurred in many areas of daily life such as in telecommunication networks, mobile communications, online banking, and E-commerce. Fraud detection involves identifying a fraud as early as possible once it has been performed.

Fraud detection methods are deliberately developed to defend criminals in adapting to their strategies. In fraud detection methods the development is made, methods are more difficult due to the severe limitation of the exchange of ideas in fraud detection at present, fraud detection has implemented by a number of methods such as data mining, statistics, and artificial intelligence.

❖ The main AI techniques used for fraud management include

- Data mining techniques are used to classify, cluster, and segment the data and automatically finds the associations and rules in the data. That may find interesting patterns, including those related to fraud.
- Expert systems to encode facility for detecting fraud in the form of rules.
- The approximate classes, clusters, or patterns of suspicious behavior detects pattern recognition either automatically (unsupervised) or to match given inputs.
- Machine learning techniques are automatically identifying characteristics of the fraud.

1) Supervised Learning for Fraud Detection:

This technique is used supervised learning in which all the available group of records is distinguished as "fraudulent" and "non-fraudulent". Examples: classification,

regression. Then machines are a self-propelled or we can say that trained to identify records according to this classification.

However, these techniques are only capable of identifying frauds that has already occurred and about which the system has been trained. In this the Data mining task of inferring a function from labeled training data. A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. Data of training consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a bearing) and a desired output value it also called as the supervisory signal.

2) Unsupervised Learning for Fraud Detection:

This technique only identifies the conceivability of some records to be more fraudulent than others without statistical analysis assurance. Examples: clustering. Unsupervised learning is closer to the exploratory spirit of Data Mining as stressed in the definitions given above. In these learning situations all the variables are treated in the same way, there is no distinction between explanatory and dependent variables. In Data mining, the problem of unsupervised learning is that of trying to find hidden structure in unlabeled data.

IV. Types of fraud

There are following types of fraud that can be detected by using data mining techniques which are as follows.

Types of Credit Card Fraud

- Application fraud.
- Skimming technology.
- Financial fraud.

1. Credit Card Fraud detection techniques:

Credit card fraud detection is the process of monitoring the behavior of the customer's transaction level through a period of time. Its process to stop the unauthorized activity by hackers. Or we can say intruders.

i. 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. Means this is technique used when there is transaction activity done an Outsider are which not before used by the particular person at that area so peer gap is checked that activity along with their transaction

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. it Means the particular persons withdraw money at a time so it will not happen before so depend on that activity the suspicious activity detected by break point analysis it will break a process ,something is wrong with account .

2. Telecommunication Fraud:

Types of Telecommunication Fraud:

Some common varieties of fraud in the telecommunications world.

- **Subscription fraud:** It happens when someone signs up for a particular service (e.g., extra lines with no intent to pay for service. In this case all calls associated with the given fraudulent activity line are fraudulent but, are stable with the profile of the user.
- **Intrusion fraud:** This occurs when an existing system , otherwise legitimate account, typically a financial or any business, is compromised in some way by an intruder, who subsequently makes or sells calls on this account. In contrast to subscription calls, the legitimate calls may be interspersed with fraudulent calls, calling for an anomaly detection algorithm.

ii. Data Mining Technique for Telecommunication Fraud:

Signature-based Methods: These methods are considered supervised/unsupervised learning hybrid. Signatures are simply telecommunication accounts summaries or can we say history that computer user including frequency of use, length of use, type if use and location of use. A signature is a "vector of feature" whose variables are obtained from the coded fields of a collection of Call Detail Records (CDRs). Each Records is in turn a vector of features that can be discrete, such as the, or continuous calling city, such as the calling duration. Examples of those fields from the call details records include following:

- Origin of call
- Destination of call
- Time of call

3. Money Laundering Fraud:

Money laundering is defined as the process of concealing or disguising the proceeds of a crime or converting those proceeds into goods and services. It allows criminals to infuse money obtained illegally into the stream of commerce, thus corrupting financial institutions and the money supply. Money laundering is usually associated with crimes that provide a financial gain. This includes, but is not limited to, bank fraud, insurance fraud, mortgage fraud, health care fraud, securities/commodities frauds, advanced fee schemes, high yield and prime bank note schemes, government fraud, corporate and occupational frauds, public corruption, cyber-crimes, organized crime, and the financing of terrorism.

Data Mining Techniques for AML:**Rule-based Approach:**

A forensic methodology for private banks, which included the monitoring of transaction audit logs as per Reserve Bank of India (RBI) guidelines to mark the suspicious transactions activities if any and the of rule based approaches Evidence to generate reports. They proposed an ontology based system for detecting suspicious transactions based on a set of Semantic Web Rule Language rule.

Classification-based Approach:

Financial transactions and aimed to detect anomalies inside a data set of mobile money financial transactions by using the classification techniques to group transaction as suspicious or non-suspicious. A novel algorithm is proposed to detect money laundering using an improved minimum spanning tree clustering, and an analysis of similarity measure distance metric.

- **ADVANTAGES OF DATA MINING**

1. Marking/Retailing
2. Banking/Crediting
3. Law enforcement
4. Researchers

- **DISADVANTAGES OF DATA MINING**

1. Privacy Issues
2. Security issues
3. Misuse of information/inaccurate information

CONCLUSION

This seminar describes overview of the concepts of data mining and fraud detection, followed by a discussion of evolution, techniques and characteristics. It also includes fraud detection in three areas, credit card fraud detection, computer intrusion detection, and telecommunication fraud detection .It presents the characteristics of fraud types, the need of fraud detection systems, and several current fraud detection techniques. These systems are effective against several kinds of fraud.

Fraud remains a challenge for businesses and organizations in many fields. Data mining is an effective method for detecting various types of fraud including telecommunication, credit card and medical insurance fraud as well as detecting intrusion to computer systems.

ACKNOWLEDGEMENT

The making of the seminar needed co-operation and guidance of a number of people. I therefore consider it my prime duty to thank all those who had helped me through their venture. It is my immense pleasure to express my gratitude to Prof. D.D.SHIRBHATE as guide who provided me constructive and positive feedback during the preparation of this seminar I express my sincere thank to the head of department prof. J.H.SATURWAR and all other staff members of CSE department for their kind co-operation. I would like to thank Dr. A.W.KOLHATKAR, Principal of our institution for providing necessary facility during the period of working on this report. I am thankful to my friends and library staff members whose encouragement and suggestion helped me to complete my seminar.

REFERENCES

- [1] In ACTS Mobile Sum- met, Grenada, Spain, 1997. [18] S. Patel. Location identity and wireless fraud detection. In ICPWC'97 Technical Program, Lucent technologies, Wire- less Secure Communications Lab, 1997.
- [2] Moeller, R. A. (2001), "Distributed Data Warehousing Using Web Technology", AMACOM, New York.
- [3] Canadian Securities Administrator (CSA) 2010.report, last access March 7, 2012 at: http://www.osc.gov.on.ca/documents/en/About/csa_20110222_csa-rpt-enf-2010.pdf
- [4] Shashidhar, H.V. and S. Varadarajan, 2011. Customer segmentation of bank based on data mining-security value based heuristic approach as a replacement to k-means segmentation. Int. J. Compute. Apply, 19:13-18.
- [5] HarmeetKaurKhanuja, Dattatraya S. Adane,"Forensic Analysis for Monitoring Database Transactions", Springer, Computer and Information Science Volume 467, 201-210,
- [6] S.D.Gheware, Int. J. of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 10, October 2014,"Data Mining: Task, Tools, Techniques and Applications"