OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Data Mining For Fraud Detection

Sushma pal *Aafreen Jamal
Department of computer Science &Engineering Al-Falah School of Engineering and Technology,Dhauj - 121004(Haryana), India
** Assistant Professor, Department of Computer Science &Engineering Al-Falah School of Engineering and Technology,Dhauj -121004(Haryana), India*

ABSTRACT: *This review paper define the way of fraud detection with the help of data mining techniques which summaries from different type of known fraud. Fraud detection includes monitoring of the behavior of user. Fraud is million dollar business and which increase every year very rapidly. This paper defines the techniques used for fraud detection.*

Key Word: *Data mining element, data mining task, techniques, decision tree method.*

I. INTRODUCTION

Fraud means obtaining goods, services and money by illegal way. In a competitive environment fraud can become a business. Data mining combine with data analysis techniques with high-end technology for use with in a process. The primary goal of Data mining is to collect information and process them to get meaning full information.

Data mining consists of five major elements:

• Extract, transform, and load transaction data onto the data warehouse system
• Store and manage the data in a multidimensional database system.
• Provide data access to business analysts and information technology professionals.
• Analyze the data by application software.
• Present the data in a useful format.

II. DATA MINING TASK

The data mining tasks are of different types depending on the use of data mining result the data mining tasks are classified as:

2.1 Exploratory Data Analysis:

In the repositories vast amount of information's are available .This data mining task will serve the two purposes

- Without the knowledge for what the customer is searching, then
- It analyzes the data. These techniques are interactive and visual to the customer.

2.2 Descriptive Modeling:

It describe all the data, it includes models for overall probability distribution of the data, partitioning of the p-dimensional space into groups and models describing the relationships between the variables.

2.3 Predictive Modeling:

This model permits the value of one variable to be predicted from the known values of other variables.

2.4. Discovering Patterns and Rules:

This task is primarily used to find the hidden pattern as well as to discover the pattern in the cluster. In a cluster a number of patterns of different size and clusters are available .The aim of this task is "how best we will detect the patterns" .This can be accomplished by using rule induction and many more techniques in the data mining algorithm like(K-Means /K-Medoids) .These are called the clustering algorithm.

2.5 Retrieval by Content:

The primary objective of this task is to find the data sets of frequently used in the for audio/video as well as images It is finding pattern similar to the pattern of interest in the data set

III. TECHNIQUES

Techniques used for fraud detection fall into two primary classes: statistical techniques and artificial intelligence. Examples of statistical data analysis techniques are:

- Data preprocessing techniques for detection, validation, error correction, and filling up of missing or incorrect data.
- Calculation of various statistical parameters such as averages, quantiles, performance metrics, probability distributions, and so on. For example, the averages may include average length of call, average number of calls per month and average delays in bill payment.
- Models and probability distributions of various business activities either in terms of various parameters or probability distributions.
- Computing user profiles.
- Time-series analysis of time-dependent data.
- Clustering and classification to find patterns and associations among groups of data.
- Matching algorithms to detect anomalies in the behavior of transactions or users as compared to previously known models and profiles. Techniques are also needed to eliminate false alarms, estimate risks, and predict future of current transactions or users.

Fraud management is a knowledge-intensive activity. The main AI techniques used for fraud management include:

- Data mining to classify, cluster, and segment the data and automatically find associations and rules in the data that may signify interesting patterns, including those related to fraud.
- Expert systems to encode expertise for detecting fraud in the form of rules.
- Pattern recognition to detect approximate classes, clusters, or patterns of suspicious behavior either automatically (unsupervised) or to match given inputs.
- Machine learning techniques to automatically identify characteristics of fraud.
- Neural networks that can learn suspicious patterns from samples and used later to detect the

IV. METHOD OF FRAUD DETECTION

4.1 Decision Tree Induction algorithm:

A Decision tree algorithm is a method for approaching discrete-valued target functions, in which the learned function is denoted by a decision tree. These types of algorithms are famous in inductive learning and have been successfully applied to abroad range of tasks.

The decision tree is a structure that contains root node, branch and leaf node. Every internal node indicates a test on attribute, every branch indicates the outcome of test and each leaf node holds the class tag. The uppermost node in the tree is the root node .A Decision trees organize circumstances by sorting them down the tree from the root to some leaf node, which delivers the classification of the instance. Each node in the tree specifies a test of some attribute of the instance and each branch descending from that node links to one of the possible values for this attribute.

4.2 Hunt's Algorithm

Step 1: Let xt be the set of training record from node t.

Step 2: Let y = {y1, y2 . . . . . yn} be the class labels.

Step 3: If all records in xt belong to the same class yt is a leaf node labeled at yt

Step 4: If xt contain records that belongs more than one class

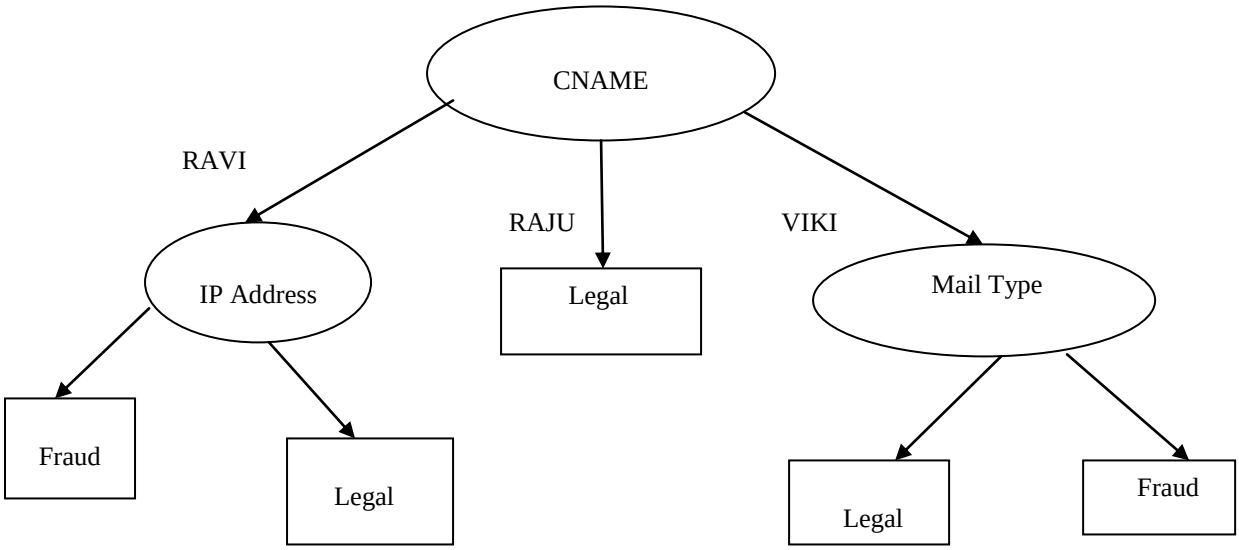
- Select attribute test conditions to partition the records in the smaller subset.
- Create child node for each outcome of the test.

4.3 Hunt's Algorithm using Credit Card Fraud Transaction Data

Table 1 for Credit Card Fraud Transaction Data

TID	CNa me	Mail Type	IP Address	Tran sAmt	Tran type
T1	Ravi	Customer	117.204.23.162	Low	Fraud
T2	Ravi	Customer	117.204.23.162	High	Fraud
T3	Raju	Customer	117.204.23.162	Low	Legal
T4	Viki	Customer	117.204.23.162	Low	Legal
T5	Viki	Merchant	61.16.173.243	Low	Legal
T6	Viki	Merchant	117.204.23.162	Low	Fraud
T7	Raju	Merchant	61.16.173.243	High	Legal
T8	Ravi	Merchant	117.204.23.162	Low	Fraud
T9	Ravi	Merchant	61.16.173.243	Low	Legal
T10	Viki	Customer	61.16.173.243	Low	Legal
T11	Ezhil	Customer	117.204.23.162	High	Legal
T12	Raju	Customer	117.204.23.162	High	Legal

Let x contains five attributes such as x1, x2, x3, x4, x5 and value x1= TID ,x2= CNAME ,x3= Mail Type ,x4=IPAddress ,x5=TransAmt and y be the Class labels contains the attribute TransType and values are Either Legal or Fraud. So Legal contains y3,y4,y5,y7,y9,y10,y11,y12and Fraud Contains y1,y2,y6,y8 class labels. Raju belongs to the same class (Legal).so Raju comes for Leaf Node or Terminal Node and RaviI and Viki have more than one class Legal and Fraud.We canTest the condition and partition the records in to smaller subsets and create child node for each outcome of the test.



Decision tree look like the above

V. CONCLUSION

This Paper contains method and techniques for fraud detection by using data mining. In this we talk about two primary techniques that are statistical techniques and artificial intelligence. There is method for Credit Card Fraud Detection using effective algorithm, Decision Tree Learning. In this Technique we simply find out the Fraudulent Customer/Merchant through Tracing Fake Mail and IP Address. Customer /merchant are suspicious if the mail is fake they are traced all information about the owner/sender through IP Address. It can find out the Location of the customer and Trace all details. Decision Tree is Most Powerful Technique in Data Mining Decision Tree is vital part of Credit card Fraud Detection.

VI. FUTURE WORK

As a future work, other data mining algorithms such as different versions of Artificial Neural Networks (ANN) and logistic regression will be used to build new classification models on the same real world dataset and the performance of the new models will be compared with the performance of the models given in this paper. Also, instead of making performance comparisons just over the prediction accuracy, these comparisons will be extended to include the comparisons over other performance metrics.

REFERENCES

[1] G k Palshikar , The hidden Truth Fraud and their control. A critical application for Business Intelligence.
[2] Green B & Chori, j. Assessing the Risk Management Fraud through Neural Network Technology.
[3] Jon T.S. Quah, M. Sriganesh, "Real-time credit card fraud detection using computational intelligence", Expert Systems with Applications, 35(4), pp.1721-1732, 2008.
[4] Suvasini Panigrahi, Amlan Kundu, Shamik Sural, A.K. Majumdar, "Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning", Information Fusion, 10(4), pp. 3630-3640, 2009.
[5] Quinlan, J.R. 1986. Induction of Decision trees. Machine Learning.
[6] http://www.ijcsi.org/papers/IJCSI-9-5-2-406-412.pdf
[7] https://en.wikipedia.org/wiki/Data_analysis_techniques_for_fraud_detection