Fraud Detection Using Data Mining Techniques

E. Henrik#1, R. Anitha*2

[1]*Assistant Professor, Department of Computer Application, S.A Engineering College, Chennai, India*
[2]*Scholar, Department of Computer Application, S.A Engineering College, Chennai, India*
[1]*Chrishenrik.ch@gmail.com*

Abstract— The purpose of this study is to develop a data mining model for fraud detection in various fields such as online transactions (i.e) net banking, tax payers, credit cards etc. and securing data from the third parties(intruders). For fraud detection, very large volumes of data are to be mined and needs some complex mechanism in order to extract full information about those integrators and frauds. It is truly based on baseline study. Some base line studies might help in establishing the extent of fraud detection. These baseline studies uses some current technologies such as ID3 algorithm, Hashing algorithms, Apriori algorithm, query outlier analyses, use case diagrams, pattern reorganization, etc. In order to detect fraud, they depend upon some random audits, informants and some undercover operations. To overcome this problem we used a hashing association mining technique. This study might increase the efficiency and accuracy in fraud identification. For fraud detection, there is no need of any electronically signatures and ensures that some process would not be rejected so that it would increase the performance or working of false alarm rate for fraud detection. Not only have credit cards, but also in every fields such as money laundering, taxed payers and so on. The algorithm is based on the traversal path using hashing approach which is a theoretical approach. Not only managing fraud detection in tax payers but also it is applicable for integrators in network areas in changing data. It uses various tools in finding the frauds and also the integrators. This paper is mainly concentrates on fraud detection and integrators. There is a major need to develop the hierarchical tree structure into different constraints for the fraud detection system. The rate efficiency of this system will determine by analysing raw data and compare with previous techniques. This algorithm will build or improve the alertness by alarm rate which may be classified into two types, one is false alarm rate and the other is actual alarm rate. With the help of these alarm rates, frauds and as well as integrators might be easily detected. So there is a necessity of alarm or alertness for the frauds and integrators.

Keywords— *Data Mining; Hashing Technique; False Alarm Rate; Actual Alarm Rate.*

1. Introduction

In today’s world, technologies get updated according to the future trends. The rates of intrusion of frauds are increasing double every year. The fraud detection is mandatory since it affects not only to the financial situations but also the entire nations. Those technologies may be used for good activities and as well as the bad activities depends upon the user mind set who handles those technologies for retrieving information without getting permission to access. Criminal activities are appearing more and more complicated and perhaps this might be the major reason for the difficulties in fraud detecting and integrators. To handle those frauds and securing data, some methodologies are used. Those methodologies might have a chance for enhancing the efficiency and accuracy for fraud detection. Traditional data mining techniques can also be used for fraud detection, but it has no power in today’s trend. Since the amount of data is unstructured and semi structured, it is very difficult to find integrators. Those unstructured data must be monitored without losing some confidential data or any changes made by the intruders. In early days, there was no existence of this technique. Recently the world is improving by gaining some knowledge about this technique. But they are not more effective because of lack of knowledge, experience etc. This hash based association technique is the most powerful fraud detection technique. To satisfy those constraints a hash based association mining technique is used.

In this study, data mining model is used as a fraud detection tool for detecting the frauds and securing the data as well from the intruders. The integrators will silently sneaking into our data store or any database in which we store large amount or volume of data. Since the world is keep on upgrading or transforming themselves, the crime rates also increased. Many criminal activities are mostly done in networks and for this case, neural network is used. But our idea is that for detecting frauds and intruders in network can also be done by using this hash based association mining technique. Regular updates are also possible in this algorithm. Based on the analytical process of the data, it is the most familiar way to reduce the frauds and intruders. It spends some time in spending the published system.

From the above literatures there might be some defects in this proposal study, because [1] there might be some issues such as difficulty in finding fraud activities and securing data. Sometimes it may be [2] less effective in finding frauds without the authorities knowledge. They might have small changes in the values in the database. We are sure that these methodologies and techniques are not enough for the fraud detection. The techniques have to be updated till it permanently detecting frauds and safeguards data from intruders.

1.1 Drawbacks

- It has some limitation in time constraints to find frauds and integrators. Similarly, the data is not properly normalized and also difficult to delete the duplicate tuple.
- Other reason is that it has some drawbacks such as it proposed several techniques in fraud detection. But it is not more efficient and stronger.
- It consumes so much time and human power to identify the frauds.
- If the result is not accurate, this may lead to lack in identifying frauds and integrators.

2. Proposed System

Fraud identification and securing data is very difficult task. Several intruders might sneak our database and doing some fraudulent activities [3]. This might take more time to identify those intruders. In order to overcome such situation, a well known hash based technique is used. The hash based technique is the powerful and more effective techniques used in fraud detection. With the help of this technique there are many possibilities of finding the frauds. The main use of this technique is to reduce the possibilities of fraudulent activities and securing the data. This is a most powerful technique used in detecting frauds. The results generated will be more accurate while comparing to other techniques. This study is not 100% effective in fraud detection, but it ensures that it will reduce intruders sneaking into our databases or gathering information without any authentication from authorized person.

3. Methodology

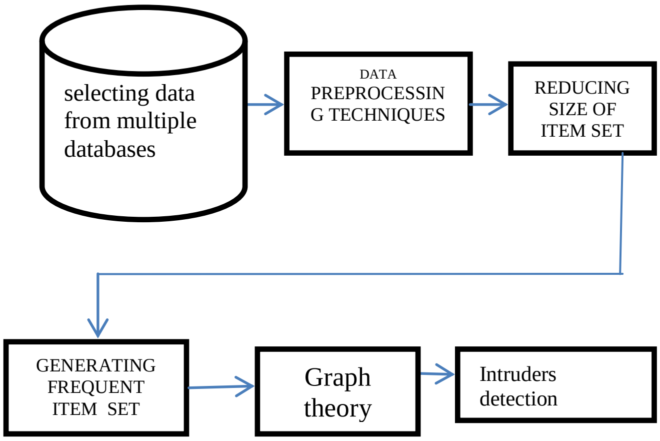
selecting data from multiple databases
DATA PREPROCESSING TECHNIQUES
REDUCING SIZE OF ITEM SET
GENERATING FREQUENT ITEM SET
Graph theory
Intruders detection

Fig.1: Hash based Association Technique

In this hash based technique, all data contained in the database db1, db2 are collected [4]. Data can also be collected from multiple numbers of databases for gathering further information. After collecting the data, they are filtered by using some data preprocessing techniques such as data cleaning, data reduction, data integration, data transformation etc. Data cleaning consist of smoothing noisy data, removing outlier data, removing inconsistence data. Then the data is to be integrated by merging from multiple sources like relational database or data cube. Then it reduced with the help of data reduction technique. Data reduction eliminates redundant features. Then in data transformation, it has some operations such as summarization, normalization and aggregation. In normalization method data is normalized which are scaled as to fall within a small specified weights. min_max method is used and data will be scaled according to the maximum and minimum value [5]. In z-score normalization, the given data is normalized by mean and standard deviation. In decimal scaling we can normalize data by scaling the decimal point of attributes value in the given data. The resulting data is reduced into frequent 2 itemset by identifying the repeated number of accessing to the data by the same person [6]. For finding the frequent itemsets, apriori algorithm or function point is used, which can be more effective with the help of candidate generation. The graph is generated according to the generated frequent itemset. The intruders can be easily detected by analysing the graphical results of the techniques in order to detect the frauds.

4. Implementation and Result

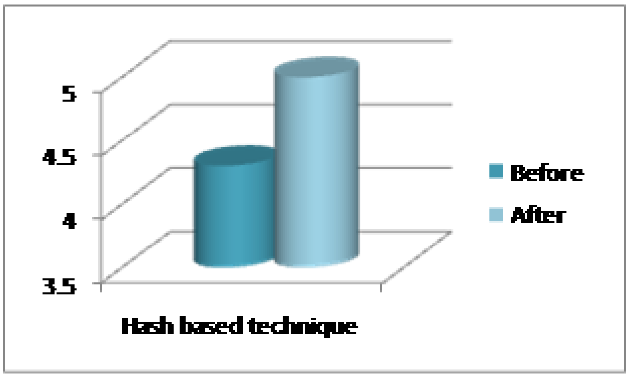

Fig. 2: .Efficiency of Hash based Association

Figure 2 clearly explains the efficiency of the hash based technique used for fraud detection. First bar describes about the comparison of efficiency before adding any enhancement or additional features and the second bar shows the efficiency after adding some additional features to the hash based technique.

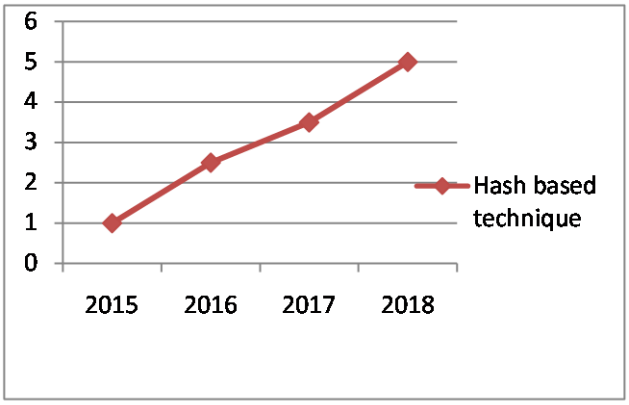

Fig. 3: Year by Year Comparison

Figure 3 describes the effectiveness of the hash based technique from current year to the previous years. This might give some predictions about the effectiveness of this technique in the next year.

Figure 4 describes about the comparison between the techniques used with hash based association techniques. We compare the techniques with each other in order to improve the efficiency. Frequent itemset efficiency is more than data cleaning and frequent 2 itemset with candidate generation is more efficient than previous frequent itemset generation.

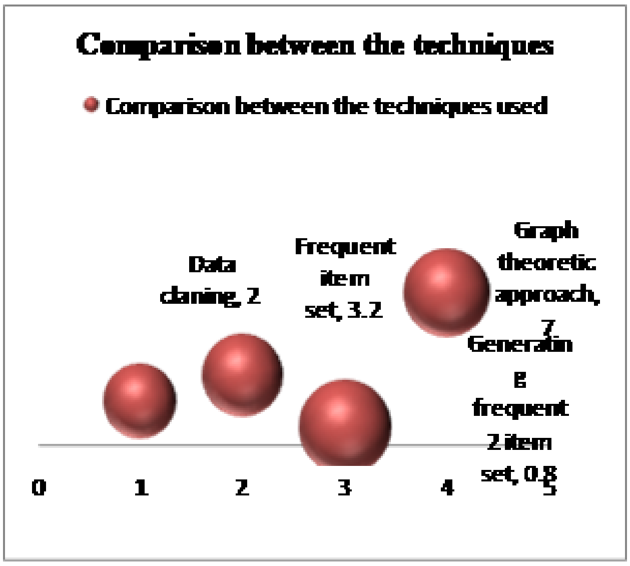

Fig. 4: Comparison among the Techniques

5. Conclusion

The proposed system introduced to improve more efficiency of existing fraud detection techniques by identifying the suspicious members in the layering stages of fraud detection process. It determines frequent access to the databases with hash based association technique .This technique is successful in finding intruder agent and the modifier in the transaction path. This has the highest possibilities of detecting the intruders. The solution proposed here is highly advantageous and necessary for online transactions in banks, and in any other applications related to transaction of data. The main purpose is to completely eliminate frauds or some anomalies by using the results of the tools for baseline study and some data mining technique.

6. Future Enhancement

This technique will be updated and might shows more deeper information. Frequent access should not be the only criteria, there might be the case identifying the non frequent access to the data. These cases will also be included in future works for fraud detection. Also we can slightly integrate in the technique to predict the future outcomes by analysing previous and current year.

References

[1] Memorie Mwanza, Jakson Phiri, Fraud detection on bulk Tax data using Buisness intelligent data mining tool:A case of zambia revenue authority, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 3, March 2016, pp. 793 – 798.

[2] Priyanaka Yadav, Pavan, Wangade, Proposed Distribution of Credit Card Fraud Detection, International Research Journal of Engineering and Technology, Volume: 03 Issue: 04, Apr 2016, pp. 460 – 463.

[3] Suresh, Swetha, Thammi Reddy, Hybrid approach for detecting suspicious account, I.J. Information Technology and Computer Science, Vol.5, 2016, pp. 37-43.

[4] Sandeep kumar, Satbir Jain, Avnish Banger, Improved detection and classification using ID3 algorithm, International Journal of Database Theory and Application, Vol.9, No.5, 2016, pp.241-250

[5] Samudre, Vaishali, Review on intrusion detection technique, International Journal of Engineering Technology, Management and Applied Sciences, Volume 4, Issue 1, January 2016, pp.123-128.

[6] Abdul Khanadar, Bhanupriya Sharma, Shambhavi Srivastava, Data Mining from Smart Card Data using Data Clustering, International Journal of Applied Engineering Research, Volume 11, Number 1, 2016, pp. 347-352.