Efficient Fraud Detection Using Hybrid Data Mining Techniques in Financial Transaction Systems

Fasi Ahmed Parvez Mohammad[1], Dr. Manisha[2]

[1]Research Scholar, Department of Computer Science & Engineering, JS University, Shikohabad, UP
[2]Assistant Professor Supervisor, Department of Computer Science & Engineering, JS University, Shikohabad, UP

ABSTRACT

The rapid growth of digital financial transactions has significantly increased the risk of fraudulent activities, making fraud detection a critical challenge for financial institutions. Traditional fraud detection methods often struggle with large-scale, high-dimensional, and imbalanced transaction data. This paper presents an efficient fraud detection framework using hybrid data mining techniques in financial transaction systems. The proposed approach integrates multiple data mining methods, including classification, clustering, and anomaly detection, to improve detection accuracy and reduce false positives. Machine learning algorithms such as decision trees, support vector machines, and neural networks are combined with unsupervised techniques to identify both known and previously unseen fraud patterns. Feature selection and data preprocessing techniques are applied to enhance model performance and computational efficiency. Experimental results demonstrate that the hybrid approach outperforms single-model techniques in terms of accuracy, precision, recall, and overall robustness. The proposed system provides a scalable and effective solution for real-time fraud detection, helping financial institutions minimize losses and enhance transaction security.

Keywords: Fraud Detection, Hybrid Data Mining, Financial Transactions, Machine Learning

INTRODUCTION

The term "data mining" refers to a method that is used in order to unearth information that has not been found before inside large data sets. A great deal of optimism is shown by this novel way to gaining access to the essential information that is stored in data warehouses. Companies are able to become knowledge-driven and strategic via the use of data mining tools, which allow them to anticipate future events and developments. The use of data mining to give false future evaluations was intended to serve as an alternative to the event analyses that are generated by devices that demonstrate decision support systems. When it comes to business questions, data mining techniques may be able to supply answers that would otherwise need a longer amount of time to locate. Using this, the process of examining records for unexpected patterns and analytical data is simplified. One of the primary objectives of data mining is to identify trustworthy patterns in data that have not been noticed before. One of the primary goals of the area of data mining is to identify patterns within large data sets. The most important goal is to find new patterns hidden inside the data. A number of different types of data mining methods are used in a wide variety of different businesses. Some examples of these approaches include regression models, complex neural networks, grouping, prediction, and classification.

Data mining has the potential to unearth previously concealed realities inside enormous data sets, which is why the identification of financial fraud (FFD) is so important. Data mining is the process of identifying patterns within datasets with the intention of providing information that may be used to inform decision-making. The process of obtaining useful information from big datasets via the use of mathematical, statistical, and machine learning methods is referred to under the heading of "data mining." Data mining is the process of searching through large data sets in order to discover information that was previously undiscovered and helpful. Among the many benefits of data mining is the capacity to develop new models that are capable of identifying new attacks before individuals are able to do so. The identification of fraudulent activity is one of the most important uses of data mining, which may be found in both the public and commercial sectors. A wide variety of data mining strategies are used by the FFD.

Mining data is done with the intention of identifying and preventing schemes that include money laundering. The fact that evidence of fraudulent activity in bank accounts was discovered via the use of a data mining technique is the most important part of fraud detection. What monitoring for fraud is unable to help with It would be helpful if you could offer some insight on the practical criteria that banks have for the transfer. When it comes to money laundering, the approach is determined by the user's account data. When doing business, whether locally or online, a significant number of individuals utilize their bank accounts.

It is a matter of national concern since the act of money laundering is illegal and it presents a substantial risk to the institutions that deal with financial matters. The majority of banks and other financial institutions have measures in place to avoid harm; but, regulatory bodies have determined that these precautions do not meet the required standards. In order to improve protection, security measures such as message padding have been included; nevertheless, failure detection based on money laundering has not been implemented. Following the selection of characteristics to make use of, the usual phases in data mining include modeling, data collection, management, and measurement of the level of success achieved. A number of cases of money laundering and fraud have been discovered in recent times via the use of data mining tools.

A. DIVERSION OF FUNDS

Financial institutions, such as banks and other financial institutions, are especially susceptible to the damage that is caused by money laundering. When illegal cash are converted into what seems to be genuine currency, a procedure that is known as "money laundering" is used. However, several rules and legislation have established a connection between money laundering and a variety of other types of financial crimes. There is a possibility that the financial system will be utilized for the purpose of abuse when individuals misuse it. Thieves are able to disguise their identities due to the volume of information that is available on the Internet, which makes the identification of illegal activity far more important and widespread.

Figure 1.1 DIVERSION OF FUNDS

LITERATURE REVIEW

Through the use of data mining methods, it is possible that money laundering (ML) schemes may be recognized and stopped more effectively. The ML approach involves doing research on the characteristics of account users. Because of the strange behavior in the bank account, there is something wrong with the situation. The implementation of practical machine learning banking principles is not the goal of the process of finding frauds. For the purpose of locating individuals who transfer money, this research suggests using a technique known as Probabilistic Relational Model and Audit Sequential Pattern Mining (PRM-ASP). Association mapping (AM) files are a kind of file that are used to build data sets that are derived from many-to-one and one-to-many files. On the other hand, it is claimed that the characteristic becomes apparent when it is unable to give flexibility and scale in the process of detecting offenders. An technique known as the Bitmap Index-based Decision Tree (BIDT) is used in order to evaluate the adaptation risk that is connected with money laundering.

The BIDT learning approach allows for the support of knowledge trees, the identification of potential risks associated with machine learning, and the facilitation of development. A BIDT may be used to get access to the systems of large banks. Instead of a list of rowids, a BIDT bitmap index makes use of a collection of bits that is referred to as a bitmap. inside the context of this index type, a sequential number is issued to each key value (such an account number) that is included inside a database. Because it is challenging to work with high-dimensional data that is arranged in a large number of clusters, the Efficient Association Rule Pattern based Money Laundering Detection (EARM-MLD) system was created. There are three key components that make up the association rule pattern mining system that is part of the EARM-MLD architecture. When it comes to backing up and safeguarding the values more than once, the first step is to locate the huge item groupings that are usual in banking laws. It is possible that we may next create spatiotemporal model–based association rules by making use of such huge datasets. This will ease the discovery process while also minimizing the number of false positives since it will be combined with a multi-clustering technique. Finally, the multi-clustering strategy is used, which involves the utilization of a large number of qualifying money transfer groups.

3. PRASP-BASED MONEY LAUNDERING DETECTION

Methods of data mining that are designed to make transactions easier have a significant challenge when it comes to the identification of money laundering (ML). Identifying possible instances of illegal conduct is one of the applications of data mining in the field of financial accounting. It is possible that the method used to identify frauds does not prioritize important machine learning banking factors. It is not possible for ML to make a comment about the agreement regarding the "K" financial database. The ML makes use of the log information that is included inside the user's account. Most persons who engage in commercial activities, whether they do so offline or online, make use of bank accounts. Internal workings of machine learning are a significant contributor to the problems that the financial system is experiencing.

In this method for identifying financial fraud, data mining jobs are categorized according to the similarities they share, and the approach also tackles challenges that are unique to the detection of fraud. On the other hand, it does not keep track of the ideas and solutions that ML banks provide. The Joint Threshold Administration (JTA) Model key is also responsible for controlling the financial systems that are dependent on kernel functions. Transactions and responses are derived from database information that is not very relevant in order to circumvent the use of machine learning.

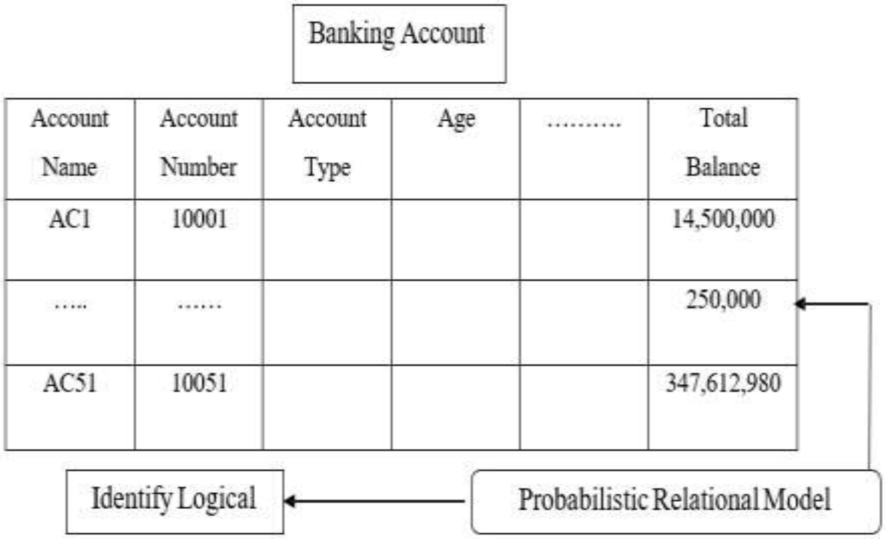
For the purpose of enhancing the effectiveness of machine learning discovery, we provide the Probabilistic Relational Model with Audit Sequential Pattern Mining, often known as PRM-ASP. You are able to divide the activities into many-to-one and one-to-many accounts by using the Association Mapping (AM) approach. PRM-ASP mining is often used for the purpose of discovering new machine learning algorithms that can be applied to time series data. Several different kinds of link accounts between activities are identified by it in order to ensure that weak accounts are identified. The PRM is used for both the process of assembling ML accounts and the identification of bank accounts that may be susceptible to vulnerability. Both relationship thinking and ASP are used by the PRM-ASP in order to recognize trends in accounts that are vulnerable to being hacked.

A. DATA MINING TECHNIQUES FOR MONEYLAUNDERING IDENTIFICATION

Applying data mining methods for machine learning fraud detection relies on the traditional data mining information flow. Functions such as feature selection, representation, data collecting, performance assessment, and management are all part of its established performance. Data mining processes use the capability of machine learning to uncover financial fraud. This is due to the fact that these processes build models that detect fraud by analyzing past thefts.

ML is a major concern for the country's financial institutions and is seen as a significant crime overall. Due to auditors' limited understanding of the organization's management and the fact that detecting fraud is not their primary purpose, auditing firms and processes are unable to detect and prevent ML fraud. Furthermore, conventional auditing methods are insufficient since they rely on sampling techniques and do not check every transaction that uses data mining tactics.

B. PRM USING THE ASP MINING ON MONEY LAUNDERING DETECTION



Account Name	Account Number	Account Type	Age	……….	Total Balance
AC1	10001				14,500,000
…..	……				250,000
AC51	10051				347,612,980

PRM-ASP Mining determines the ML accounts in the bank dataset. ML is an illegal action for financial institutions and hence become a major risk to the entire nation, so that PRM-ASP mining is used to discover the faulty bank accounts.

From the figure 3.1, logical relationships among client information is employed to recognize the ML by using PRM-ASP. PRM-ASP mining is achieved based on the personal information of the clients. The relational logic and ASP are extracted from the client and banking companies. PRM is used to explain the associations among the objects.

4. DECISION TREE–BASED ML RISK

The detection of money laundering (ML) makes use of time series data to identify one-to-many and many-to-one linkages among transactions. This allows for the identification of accounts that are susceptible to being abused. Through the use of related reasoning, the audit sequential pattern (ASP) has the potential to discover account transactions that are susceptible. Furthermore, in order to deliver a rational machine learning identification in this specific scenario, ASP makes good use of a Probabilistic Relational Model. ML is practiced by criminals, which presents a significant threat to financial institutions such as banks and other companies. There are security mechanisms in place at the majority of banks and other financial institutions; nevertheless, these systems are not meeting the requirements that regulators have set out.

Although security does enhance approaches such as message padding, it does not preclude the identification of security failures based on machine learning. It is possible to have a fair trade between security and performance, but it comes at the expense of having a weak financial structure. The consequence of this is that the characteristic is seen as being vulnerable since it does not provide the scalability and flexibility that are essential for machine learning (ML) crime detection. When referring to the process of changing money that have been gained illegally into what seems to be genuine riches, the term "ML" is used. In spite of the fact that it was first used to refer to the misuse of the financial system, the term "ML" has since expanded to include a wide range of financial offenses that fall under a number of legal and judicial systems. The enormous quantity of data that is accessible on the Internet has resulted in a significant improvement in the accuracy of crime scene identification, which has, in turn, made it possible for criminals to disguise their genuine identities.

For the purpose of evaluating the adaptation risk that is linked with money laundering, it is recommended that the BIDT approach be used. A fundamental aspect of BIDT learning is the cultivation of an information tree that has the potential to reduce the risk of machine learning and increase the scalability of a company. Through the use of a bitmap index, BIDT is able to retrieve enormous financial information in an effective manner. In a BIDT, table descriptions are as follows:

Numbered in a sequential fashion, with a bitmap (an array of bits) acting as a replacement for a catalog for row IDs and each key value (that is, account number). Following this, the BIDT approach uses the "select" query performance to apply count and bit-wise logical operations to AND variables. The query results coincide to form a decision tree, which allows for a more accurate assessment of the adaptation risk in machine learning procedures. The only thing that is required of you in order to get the population frequencies for the BIDT root node is to include the total number of "1" into the structure of the bitmaps. In order to measure the risk factor rate and make predictions about money laundering, this feature is used.

A. SECURITY MODEL FOR MONEY LAUNDERING IDENTIFICATION

The process in the security model consists of failure detector framework to investigate the system faults. Here, identification process have different correction scheme to provide the security to communicate the information’s. In real data applications, crime fault is identified and it accesses the unsecured data in the model. Duplicate frame fault is very rarely identify and it avoids the improving the laundering, thus results in higher success rate.

B. SMART CARD-BASED SECURITY FRAMEWORK

The computation of security provided in ML is solved by developing a design named as smart card-based security framework. They consider trusted and un-trusted system for designing the security framework and it solves the fault occurrence in security. Trusted model is designed to solve the failure attacks that are not involved in smart card, buffer attacks and timing attacks. In general, TrustedPals are designed that used to detect the failure in ML. With the help of failure detector model, the process and communication is carried out for providing security in framework. The detection model is based on the well-connected process and consensus algorithm. Here, well-connected process involves both in-connect and out-connect processes.

The process for providing the security in ML is based on hybrid model which is divided into two models namely trusted system and un-trusted system. The trusted system involved in the software applications is possible to provide the system with higher security model. The stored information’s in smart card-security model are transmitted and protected from the tamper-proof trusted process. The applications like e- banking or e-voting are executed with the process involved in un-trusted system.

C. EFFICIENT ASSOCIATION RULE PATTERN BASED MONEY LAUNDERING DETECTION FRAMEWORK

An Efficient Association Rule Pattern based Money Laundering Detection framework that supports the surveillance in detecting the ML is planned. The objective of the framework includes efficient design of association rule pattern that eliminating the incomplete data by performing mapping technique. The cleaned dataset is mapped and analyzed using mapping algorithms based on Spatio Temporal-based Association Rule Pattern Mapping. The results obtained with the use of mapping algorithm helps in reducing the time for detecting ML and improve the fraud identification accuracy by reducing the false positive time. Figure 5.1 shows the block diagram of the Efficient Association Rule Pattern based Money Laundering Detection (EARM-MLD) framework.

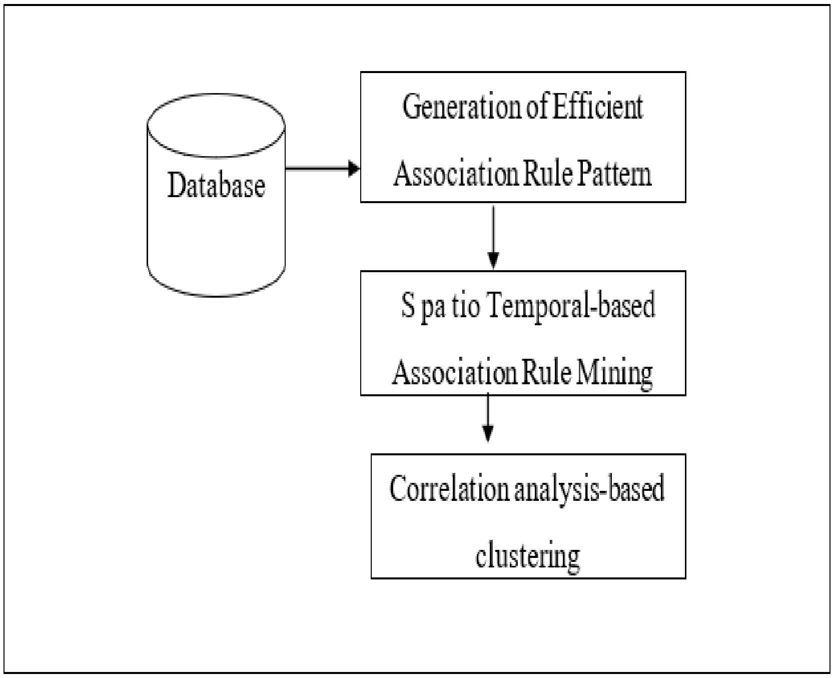

Database
Generation of Efficient Association Rule Pattern
S pa tio Temporal-based Association Rule Mining
Correlation analysis-based clustering

Figure: Block diagram of Efficient Association Rule Pattern based Money Laundering Detection framework

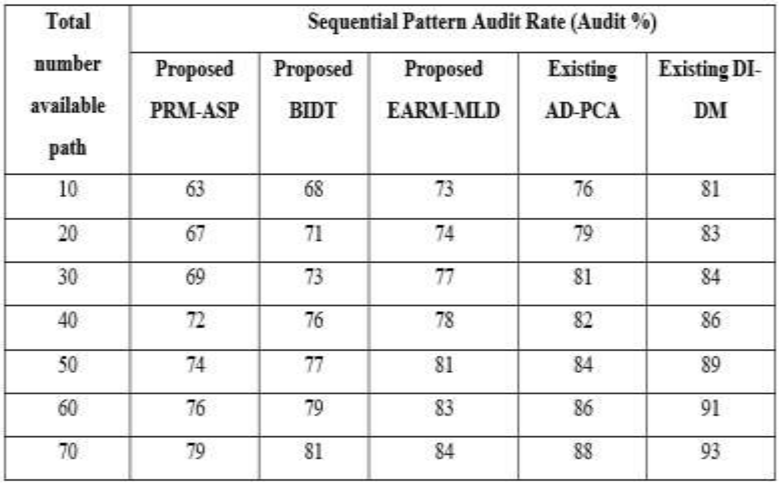
RESULTS

PRM-ASP Mining model is planned for determine the money laundering (ML) accounts. AM algorithm is carried out on the preprocessed data set and to divide the transactions from different type of accounts. ML identification utilizes the time series data for detecting the different type of accounts among transactions to identify the susceptible accounts. The susceptible bank account are identified based on categorize the transaction using PRM and it collects the ML accounts. As well to present a logical ML identification in the current scenario, PRM is efficiently used by ASP.

The adaptability risk in ML is evaluated by introducing Bitmap Index-based Decision Tree (BIDT) technique. Initially, the BIDT learning determines the company's ML risk and improves the scalability which is used to induce knowledge tree. A bitmap index in BIDT is used effectively to access large banking databases. In a BIDT, description in a table is numbered in sequence with each key value (i.e.,) account number and a bitmap (array of bits) used in its place a list of rowdies. Subsequently, BIDT algorithm uses the "select" queries performance to apply count and bit-wise logical operations on AND. The decision tree can be constructed using Queries result more precisely to estimate the adaptability risk in ML operation. The root node (i.e.,) main account of decision tree, from the bitmap construction to calculate the full amount of "1 " an population frequencies is achieved to predict the money launder and evaluate the risk factor rate.

An Efficient Association Rule Pattern based Money Laundering Detection framework is developed to handle high dimensional data with multi clustering structure. The association rule pattern mining in EARM-MLD framework consists of three major parts. Initially, frequent large itemsets are identified from banking rules which have support and confidence values more than a threshold number of times. This in turn reduces the time taken for detecting ML. Next, an association rule is constructed based on spatio temporal model from those large itemsets. It easily performs the detection operation and integrates it with multi clustering algorithm with the objective of reducing the false positive rate. Finally, the multi clustering algorithm involves the set of money transfer group which fulfills the criteria. The multi cluster elements integrated with EARM-MLD framework are handled as a suspected operation which performs the ML detection work.

Table : Tabulation for Sequential Pattern Audit Rate



Total number available path	Sequential Pattern Audit Rate (Audit %)				
	Proposed PRM-ASP	Proposed BIDT	Proposed EARM-MLD	Existing AD-PCA	Existing DI-DM
10	63	68	73	76	81
20	67	71	74	79	83
30	69	73	77	81	84
40	72	76	78	82	86
50	74	77	81	84	89
60	76	79	83	86	91
70	79	81	84	88	93

CONCLUSION

PRM-ASP Mining model is used to determine the accounts of money laundering successfully with minimum false positive rate. Initially, AM algorithm is recruited to divide the transactions process. Using mapping process, the transactions of one-to-many and many-to-one accounts are successfully identified. Relational logic transaction set called as Probabilistic Relational Model that employed to categorize the vulnerable accounts. Audit sequential pattern is enhanced to categorize the money transfer path in PRM-ASP Mining model. In addition, PRM-ASP mining model offers the logical scheme in many number of real-world domains. Performance analysis of PRM-ASP mining model improves the accuracy of fraud detection with minimum time. Finally, PRM-ASP Mining model is reduces the false positive rate and increases the processing time to monitor the user accounts.

Bitmap Index-based Decision Tree (BIDT) technique is proposed to identify the adaptability risk in money laundering. Maintaining regulatory risk rate and providing security for financial organizations has become the key for money laundering. With improving the level of true positive rate (i.e., regulatory risk rate), reduces the time taken for identifying the risk on money laundering. Bitmap Index-based Decision Trees is used for evaluation of the performance effects of regulatory risk rate. Bitmap indexing method efficiently reduces the risk identification time and greatly improves the adaptability rate by categories the rows and columns based on the account details of the customer. Initially, the use of Bitmap Indexing that efficiently handles large money laundering accounts and produces the result with fuzzy form to improve the regulatory risk rate. Next, Select Query Structure is developed with multiple key value databases that work with bitwise logical operator to minimize the false positive rate. Then it also integrated with Bitmap Index Frequency that consists of the rows and columns id using

Low Cardinality column for improving the true positive rate using Statlog German Credit Data.

An Efficient Association Rule Pattern based Money Laundering Detection framework is developed to handle high dimensional data with multi clustering structure. The association rule pattern mining in EARM-MLD framework consists of three major parts. Initially, frequent large itemsets are identified from banking rules which have support and confidence values more than a threshold number of times. This in turn reduces the time taken for detecting money laundering. Next, an association rule is constructed based on spatio temporal model from those large itemsets. It easily performs the detection operation and integrates it with multi clustering algorithm with the objective of reducing the false positive rate. Finally, the multi clustering algorithm involves the set of money transfer group which fulfills the row condition, money collecting to a particular account with minimum set size.

The performance of money laundering identification using Statlog German Credit Data demonstrates that the proposed PRM-ASP mining model reduces 22% of false positive rate and improves processing time by 8% to monitor the user accounts. Bitmap indexing method reduces the risk identification time by 21% and greatly improves the adaptability rate. Finally, EARM-MLD framework applies mapping algorithm to provide better performance with an improvement of fraud identification accuracy by 9% and system efficiency ratio is improved by 15% when compared with state-of-methods for detecting money laundering.

REFERENCES

[1]. AashleshaBhingarde, AvnishBangar, Krutika Gupta and SnigdhaKarambe International Journal of Advanced Research in Computer and Communication Engineering, Volume 4, Issue 3, March 2015, Pages 169 170.

[2]. Andrei Sorin-122 Response System for Relational Databa Engineering, Volume 23, Issue 6, June 2011, Pages 875 888 International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 2, February 2015, Pages 997 1000.

[3]. Clifton Phua, Kate Smith-Miles, Vincent Cheng-Siong Lee, and Ross Gayler, Engineering, Volume 24, Issue. 3, March 2012, Pages 533-546

[4]. Data Mining in Money Laundering Detection Springer, Volume 7197, Pages 207-216

[5]. Gilbert Sebe- Performance of Agricultural Develo Accounting Auditing and Finance Research, Volume 2, Issue 1, March 2014, Pages 1-23

[6]. Mihaela A. Bornea, Vasilis Vassalos, Yannis Kotidis, and Antonios Deligiannakis, Transactions on Knowledge and Data Engineering, Volume 22, Issue 8, August 2010, Pages 1110 1125.

[7]. Roberto Cortinas, Felix C. Freiling, Marjan Ghajar-Azadanlou, Alberto Lafuente, Mikel Larrea, Lucia Draque Pe

[8]. Volume 9, Issue 4, July/August 2012, Pages 610-625

[9]. Rui Liu., Xiao-long Qian., Shu Mao., Shuai- ch on anti-money Conference (CCDC), 2011, Pages 4322 4325

[10]. Sutapat Thiprungsri, and Miklos A. Vasarhelyi Accounting Research, Volume 11, 2011, Pages 69-84

[11]. Tamer Hossam Eldin Helmy , Mohamed zaki Abd-ElMegied, Tarek S. Sobh,Laundering an Applications Volume 1, Issue 1, November - December 2014, for Secure Computations with Two Non-Colluding Servers and Multiple Clients, Security, Volume 10, Issue 3, March 2015, Pages 445 457.

[12]. Mohammad Reza Keyvanpour, Mostafa Javideh and Mohammad Reza Ebrahimi, d investigating crime by means of data mining: a general crime matching 880.