Hong Kong International Journal of Research Studies, ISSN: 3078-4018
Volume 1, Issue 2, July December, 2023
Available online at: https://octopuspublication.com/index.php/hkijrs

A Method for Identifying Financial Transaction System Fraud Using Hybrid Data Mining

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

Article history: Received: 10 September 2023, Accepted: 26 September 2023, Published online: 07 October. 2023

ABSTRACT

As a consequence of the exponential growth of online financial transactions and the consequent increase in the danger of fraudulent conduct, the identification of fraudulent activity has emerged as a significant concern for financial institutions. For traditional fraud detection algorithms, one of the most prevalent challenges is dealing with massive amounts of transaction data that are multi-dimensional and skewed. The purpose of this research is to provide a technique that utilizes hybrid data mining methodologies in order to effectively identify fraudulent information in financial transaction systems. Some of the data mining methods that are used by the recommended approach include anomaly detection, clustering, and classification. These techniques are utilized to improve detection accuracy while simultaneously reducing the number of false positives. Researchers use a variety of supervised and unsupervised machine learning techniques, including as neural networks, decision trees, and support vector machines, in order to identify both frequent and atypical types of fraudulent activity. Preprocessing the data and selecting features are two steps that are taken in order to enhance the effectiveness and performance of the model. According to the data, the hybrid technique is more robust, accurate, exact, and recall-friendly than single-model tactics. This is the case when compared to conventional strategies. Through the provision of an effective and scalable solution for real-time fraud detection, the technique that has been proposed assists financial institutions in lowering losses and enhancing the security of transactions.

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

INTRODUCTION

The phrase "data mining" describes a process for discovering previously unknown information inside massive data sets. There is reason to be hopeful about this new approach to retrieving crucial data from data warehouses. With the use of data mining technologies, businesses can become more strategic and knowledge-driven by predicting what's to come. The original intention behind using data mining to provide inaccurate predictions about the future was to supplement the event analysis produced by decision support system demonstration devices. Data mining methods might ultimately provide solutions to business concerns that would have taken a lot longer to find any other way. This streamlines the process of searching records for analytical data and unexpected trends. Finding reliable patterns in data that have gone unnoticed is a major goal of data mining. Finding commonalities in massive datasets is a key focus of data mining research. Discovering previously unseen patterns within the data is the primary objective. There is a vast array of industries that make use of data mining techniques. Grouping, prediction, classification, sophisticated neural networks, and regression models are a few instances of these methods.

Due to the importance of detecting financial fraud (FFD), data mining has the ability to reveal hidden truths inside massive data sets. Data mining is the practice of searching for meaningful patterns in large datasets in order to extract useful information for making decisions. To extract valuable insights from large datasets by using statistical, mathematical, and machine learning techniques is known as "data mining." The goal of data mining is to find useful, previously unknown information by sifting through massive data sets. One of data mining's several uses is in the creation of new models that may detect emerging threats before humans can. One of the most significant applications of data mining, which may be seen in both public and private sectors, is the detection of fraudulent activities. Within its data mining framework, the FFD employs a broad range of techniques.

Finding and stopping schemes that include money laundering is the main goal of data mining. The most crucial aspect of fraud detection is the discovery of evidence of fraudulent activity in bank accounts via the use of a data mining tool. Where fraud monitoring falls short It would be really appreciated if you could provide some light on the realistic requirements that

banks have about the transfer. The method for combating money laundering is tailored to the specific user's account information. A large portion of the population uses their bank accounts while doing business, whether it's at a mall or on an internet platform.

The national government should take notice of this issue since money laundering is both criminal and poses a serious threat to financial institutions. Most financial institutions have safeguards in place to prevent damage, but these procedures have been deemed inadequate by regulatory agencies. Message padding and other security measures have been integrated to enhance protection; nonetheless, money laundering-based failure detection has not been put into place. Modeling, data collecting, management, and success assessment are the typical processes in data mining that follow feature selection. Data mining has recently uncovered many instances of fraud and money laundering.

A. SPOTTING FRAUD IN THE PROCESS OF MONEY LAUNDERING

Telling a lie with the intention of obtaining financial gain, causing harm to the reputation of another individual, or engaging in criminal action is the definition of the act of committing fraud. The two most important instruments in this battle are systemic approaches to recognizing and preventing fraudulent activity and losses. The avoidance of fraudulent activity is a prudent strategy for avoiding the possibility of fraudulent transactions. Fraud monitoring tools assess each and every transaction in order to identify the possibility of illegal behavior occurring that may have occurred. For the purpose of preventing fraud, this is irrelevant. In addition to this, it detects the ones that are fake before the con artist can finish creating the illusion.

Fraud identification is the process of recognizing fraudulent behavior in situations when there are no previous indications of such behavior. In the first step, we examine the legitimacy of the data trend. A technique known as supervised learning is used when dealing with early scam data. Unsupervised learning is all about data that leads to fraud or crime but isn't truly scam data. This is the data that drives unsupervised learning. There are a lot of terms that need to be employed for the work at hand, and those phrases are meant to be used as signs of fraudulent or illegal activities.

The improvement of information networks and information technology has led to a growth in the number of fraudulent schemes, which has resulted in the loss of a significant amount of money. However, there are many other channels via which fraud might occur, such as the telephone and the internet. Anywhere on the Internet is susceptible to fraud because to the fact that it is accessible to everyone, users have the ability to conceal their location, and transactions are kept anonymous. As a result of the expansion of fast internetworking channels, con artists are able to more easily organize their schemes. These channels enable them to exchange information and engage with persons located all over the world.

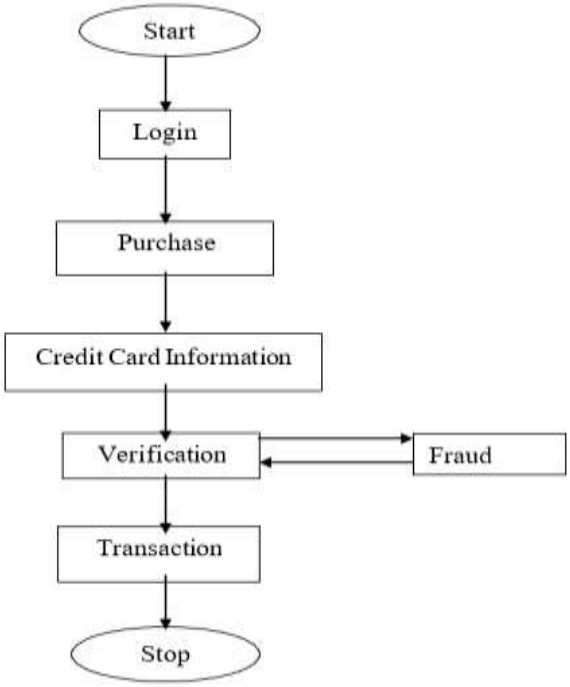


Figure: Flowchart for Credit Card Fraudulent Detection

LITERATURE REVIEW

Data mining techniques have the potential to improve the detection and thwarting of money laundering (ML) operations. Researching account user attributes is an integral part of the ML technique. The bank account has been acting strangely, which suggests that something is amiss. The technique of discovering frauds is not about implementing practical ideas of machine learning banking. This study recommends a method called Probabilistic Relational Model and Audit Sequential Pattern Mining (PRM-ASP) with the aim of identifying money transferers.

Association mapping (AM) files are a kind of data set construction tool for many-to-one and one-to-many file conversions. However, when it fails to provide scalability and flexibility in the process of offender detection, the trait supposedly becomes obvious. The Bitmap Index-based Decision Tree (BIDT) is a method for assessing the adaptation risk associated with money laundering.

Knowledge trees, possible machine learning hazards, and development facilitation are all possible with the BIDT learning technique. To breach the security of major financial institutions, a BIDT may be used. A BIDT bitmap index uses a collection of bits called a bitmap instead of a list of rowids. For every key value (such an account number) stored in a database, a unique sequential number is assigned within the framework of this index type.

The idea behind the Efficient Association Rule Pattern based Money Laundering Detection (EARM-MLD) system came from the difficulty of dealing with high-dimensional data that is organized in many clusters. The association rule pattern mining system that is a part of the EARM-MLD architecture consists of three critical components. The first stage in backing up and protecting the values several times is to identify the large item groups that are common in banking legislation. Eventually, we may be able to utilize these massive datasets to build spatiotemporal model-based association rules. When coupled with a multi-clustering method, this will streamline discovery while also reducing the amount of false positives. Finally, a large number of qualified money transfer groups are used in the multi-clustering technique.

PREVENTION OF MONEY LAUNDERING Using PRASP

A major obstacle for data mining methods aimed at simplifying transactions is the detection of money laundering (ML). Data mining has many uses in financial accounting, one of which is the detection of potential cases of wrongdoing. The technique used to detect fraud may not give sufficient weight to critical machine learning banking considerations.

The "K" financial database agreement is outside ML's commenting capabilities. Log data stored on the user's account is used by the ML. Bank accounts are used by the majority of individuals that participate in commercial activity, whether it offline or online. A major factor in the issues plaguing the financial system is the inner workings of machine learning.

This strategy for detecting financial fraud sorts data mining tasks into categories based on their shared characteristics and takes on problems specific to fraud detection. However, it fails to record the suggestions and answers offered by ML banks. Financial systems reliant on kernel functions are similarly governed by the Joint Threshold Administration (JTA) Model key. In order to sidestep machine learning, transactions and answers are produced from irrelevant database information.

We provide PRM-ASP, or the Probabilistic Relational Model with Audit Sequential Pattern Mining, to improve the efficacy of machine learning discovery. The Association Mapping (AM) method allows you to partition the tasks into many-to-one and one-to-many accounts.

Finding novel machine learning techniques to apply to time series data is a common goal of PRM-ASP mining. In order to guarantee that weak accounts are discovered, it identifies several types of connection accounts between activities. Both the assembly of machine learning accounts and the detection of potentially vulnerable bank accounts require the PRM. In order to identify patterns in accounts that are susceptible to hacking, the PRM-ASP utilizes both relationship thinking and ASP.

A. PROBABILISTIC RELATIONAL MODEL

The money transport amongst dissimilar bank clients are addressed in PRM-ASP mining and relational logic is analyzed. In the data mining step, AM are used for discovering the PRM logic. The PRM is represented in figure 3.3.

Figure 3.3 explains the money transaction to the different banking accounts with the different time frame. Transaction relational logic of ACC_15, ACC_12 and ACC_54 are tested using the PRM. Data mining step predicts susceptible account and amount of data transferred is calculated.

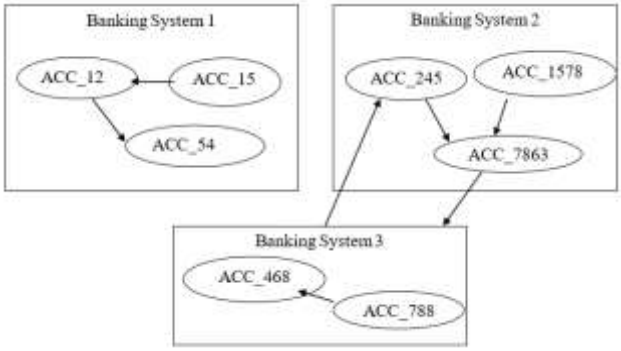
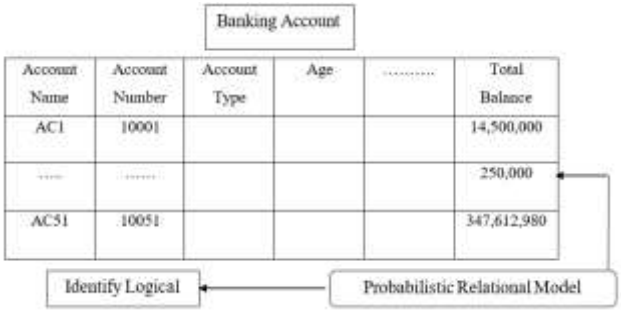


Figure : Probabilistic Relational on Different Time Frame

B. PRM USING THE ASP MINING ON MONEY LAUNDERING DETECTION



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. MONEY LAUNDERING RISK ESTIMATION

Money laundering detection (ML) uses time series data to find many-to-one and one-to-many relationships between transactions. This enables the detection of accounts that might be exploited. The audit sequential pattern (ASP) may find potentially vulnerable account transactions by using related reasoning. In addition, ASP effectively employs a Probabilistic Relational Model to provide a reasonable machine learning identification in this particular case. Criminals' use of ML poses a serious risk to banks and other businesses that deal with money. Although most banks and other financial organizations have security measures in place, many do not conform to the standards imposed by regulators.

While methods like message padding are made more secure by implementing security measures, machine learning may still detect security flaws. A poor financial structure is the price you pay for a reasonable trade-off between performance and security. Because it lacks the scalability and flexibility necessary for ML crime detection, this attribute is seen as being susceptible to attack. The term "ML" is used to describe the process of turning illicitly obtained funds into what seems to be real wealth. Though originally meant to describe financial system abuse, the word "ML" has since grown to include a broad variety of financial crimes that are dealt with by various legal and judicial systems. The vast amount of data available online has greatly enhanced the precision of crime scene identification, which has given offenders more opportunities to hide their true identities.

The BIDT method is suggested for assessing the adaptation risk associated with money laundering. Building an information tree is a cornerstone of BIDT learning; it may help mitigate machine learning risks and boost scalability. Bitmap indexing allows BIDT to efficiently retrieve massive amounts of financial data. Here are the table descriptions found in a BIDT:

Hong Kong International Journal of Research Studies, ISSN: 3078-4018
Volume 1, Issue 2, July December, 2023
Available online at: https://octopuspublication.com/index.php/hkijrs

A bitmap, which is an array of bits, is used in lieu of a catalog for each key value (the account number) and row IDs, and the numbers are numbered sequentially. After that, the BIDT method applies count and bit-wise logical operations to AND variables using the "select" query performance. With the query answers coming together to build a decision tree, the adaptation risk in machine learning techniques may be more accurately assessed. To get the population frequencies for the BIDT root node, all you have to do is include the total number of "1" into the bitmap structure. Using this capability, one may gauge the risk factor rate and forecast instances of money laundering.

A. BITMAP INDEX-BASED DECISION TREE FOR RISK EVALUATION ON FINANCIAL MONEY LAUNDERING

The risk factors are estimated on financial organizations ML using the indexing scheme. The indexing scheme uses the rows and columns to store the information that improves the scalability rate. ML is the illegal amount transacted between different users, which are evaluated using the BIDT technique. The risk related to larger amount of illegal transaction is controlled in a financial organization by constructing a decision tree with mapping of the bit in fuzzy form '0' and '1'. The decision tree contains the root and sub co-ordinate nodes to create the determination rules in the BIDT technique. The purpose of indexing in BIDT technique is to make available pointers to rows in a table consisting of given key values.

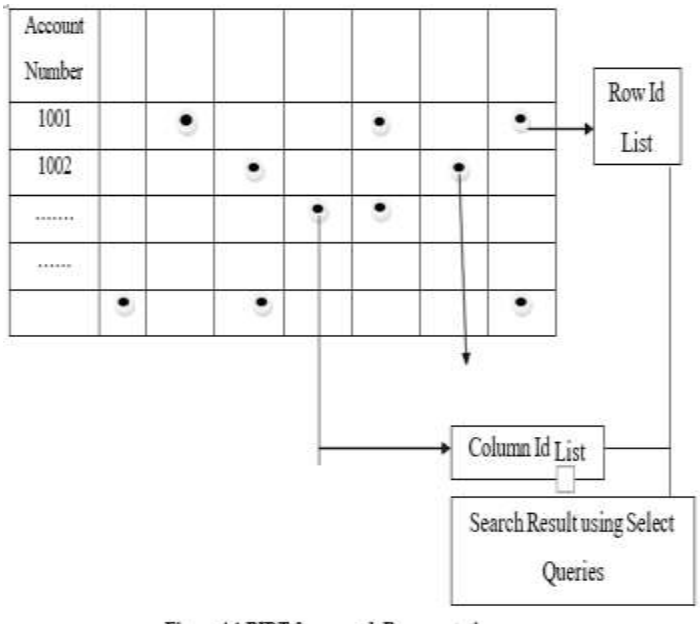


Figure: BIDT framework Representation

B. PERFORMANCE ANALYSIS OF ADAPTABILITY RATE (AR)

AR on crime using BIDT technique is the capability of service provider to alter changes in services based on customers' requests during ML operation. Adaptability of offense measures the time taken for ML changes or updates the service in higher level at less interval of time. Higher the adaptability rate, more quickly, the anti ML system is and therefore is said to be more efficient in handling the ML operations. It is measured in terms of percentage (%).

Table: Tabulation for Adaptability Rate

Hong Kong International Journal of Research Studies, ISSN: 3078-4018
Volume 1, Issue 2, July December, 2023
Available online at: https://octopuspublication.com/index.php/hkijrs

Methods	Adaptability Rate (%)
Proposed BIDT	78.13
Existing SCSF	69.35
Existing MDS	61.25

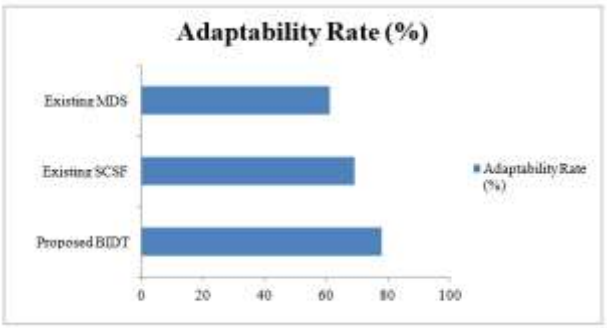


Figure 4.7 Measure of adaptability

C. EFFICIENT ASSOCIATION RULE PATTERN

The first step in the design of Efficient Association Rule Pattern based Money Laundering Detection (EARM-MLD) framework is the construction of efficient Association Rule Pattern. The association rule pattern for detecting ML identifies the frequent large itemsets having support and confidence values more than threshold number of times.

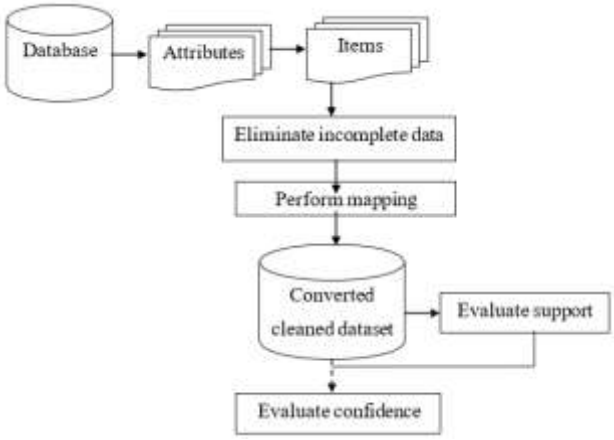


Figure : Block diagram of Efficient Association Rule Pattern

RESULTS

The purpose of developing the PRM-ASP Mining model is to identify the accounts involved in money laundering (ML). In order to separate the transactions from the various types of accounts, the preprocessed data set is subjected to the AM algorithm. Machine learning identification uses time series data to find vulnerable accounts by recognizing various types of accounts in transactions. It gathers ML accounts and uses PRM to classify transactions to find vulnerable bank accounts. This situation also makes good use of PRM, which ASP uses to provide a logical ML identity.

The Bitmap Index-based Decision Tree (BIDT) method is used to assess the adaptation risk in ML. Before inducing a

knowledge tree, BIDT learning ascertains the ML risk to the organization and enhances its scalability. To efficiently access massive financial datasets, BIDT makes use of a bitmap index. As an alternative to a list of rowdies, a BIDT uses a bitmap (array of bits) to represent each key value (i.e., account number) in a table's description. After that, the BIDT method applies count and bit-wise logical operations on AND using the "select" query performance. To more accurately assess the adaptation risk in ML operation, the decision tree may be built using the results of queries. Bitmap architecture allows the root node, or primary account of the decision tree, to compute the whole amount of "1" and use population frequencies to forecast money laundering and assess the risk factor rate.

To deal with high-dimensional data with a multi-clustering structure, we built an efficient money-laundering detection system based on association rule patterns. There are primarily three components to the EARM-MLD framework's association rule pattern mining. At first, we look for big itemsets that appear often in banking regulations and have confidence and support values that are greater than a certain threshold. As a result, the time required to identify ML is decreased. The next step is to use the spatio-temporal model to build an association rule from those massive item sets. With the goal of lowering the false positive rate, it combines with a multi-clustering algorithm and effortlessly conducts the detection operation. Finally, the collection of money transfer groups that meet the requirements is used in the multi clustering procedure. When doing ML detection work, the multi-cluster components coupled with the EARM-MLD framework are treated as a suspicious activity.

CONCLUSION

Successfully determining money laundering accounts with a minimal false positive rate is achieved using the PRM-ASP Mining model. The first step is to use the AM algorithm to partition the transaction process. The mapping technique effectively identifies the transactions involving many-to-one and one-to-many accounts. The Probabilistic Relational Model is a collection of relational logic transactions used to classify accounts as susceptible. The PRM-ASP Mining model improves the audit sequential pattern to classify the route of monetary transfers. The PRM-ASP mining model also provides a logical structure for use in a wide variety of practical settings. Improving the accuracy of fraud detection with little time is achieved by performance analysis of the PRM-ASP mining model. Last but not least, the PRM-ASP Mining model improves account monitoring processing time while decreasing false positive rates.

To find the flexibility risk in money laundering, the Bitmap Index-based Decision Tree (BIDT) method is suggested. The anti-money-laundering strategy now revolves on preserving regulatory risk rate and protecting financial institutions. Time spent detecting money laundering risks is decreased if the level of the real positive rate (i.e., regulatory risk rate) is improved. To assess the impact of regulatory risk rate on performance, Bitmap Index-based Decision Trees are used. By classifying the rows and columns according to the customer's account data, the bitmap indexing approach effectively decreases the time it takes to identify risks and substantially increases its adaptability rate. Bitmap Indexing was first used to enhance the regulatory risk rate; it easily handles big money laundering accounts and delivers results in fuzzy form. After that, in order to reduce the number of false positives, a Select Query Structure is created using many key-value databases that collaborate using a bitwise logical operator. Bitmap Index Frequency, which includes the identifiers for the rows and columns, was subsequently included as well.

Use Statlog's German Credit Data to improve the genuine positive rate using a low cardinality column.

To deal with high-dimensional data with a multi-clustering structure, we built an efficient money-laundering detection system based on association rule patterns. There are primarily three components to the EARM-MLD framework's association rule pattern mining. At first, we look for big itemsets that appear often in banking regulations and have confidence and support values that are greater than a certain threshold. As a result, the time needed to identify instances of money laundering is decreased. The next step is to use the spatio-temporal model to build an association rule from those massive item sets. With the goal of lowering the false positive rate, it combines with a multi-clustering algorithm and effortlessly conducts the detection operation. At last, the multi-clustering method incorporates a collection of transfer groups that meet the row criterion, collecting funds for a specific account with a minimal set size.

The suggested PRM-ASP mining model improves processing time for user account monitoring by 8% and decreases the false positive rate by 22% when utilizing Statlog German Credit Data for money laundering detection. The bitmap indexing approach significantly enhances the adaptation rate and decreases the risk identification time by 21%. Finally, when compared to state-of-the-art approaches for identifying money laundering, the EARM-MLD framework's use of a mapping algorithm improves performance, leading to a 15% increase in the system efficiency ratio and a 9% improvement in fraud detection accuracy.

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 Volume 9, Issue 4, July/August 2012, Pages 610-625
[8]. Rui Liu., Xiao-long Qian., Shu Mao., Shuai- ch on anti-money Conference (CCDC), 2011, Pages 4322 4325
[9]. Sutapat Thiprungsri, and Miklos A. Vasarhelyi Accounting Research, Volume 11, 2011, Pages 69-84
[10]. 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.
[11]. Mohammad Reza Keyvanpour, Mostafa Javideh and Mohammad Reza Ebrahimi, d investigating crime by means of data mining: a general crime matching 880.