I.J. Information Technology and Computer Science, **2016, 5, 37-43**
Published Online May 2016 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2016.05.04

A Hybrid Approach for Detecting Suspicious Accounts in Money Laundering Using Data Mining Techniques

Ch. Suresh
Department Of Computer Science and Engineering, GITAM University Visakhapatnam, Andhra Pradesh, India
E-mail: sureshchalumuru@gmail.com

Dr. K. Thammi Reddy
Department Of Computer Science and Engineering, GITAM University Visakhapatnam, Andhra Pradesh, India,
E-mail: thammireddy@gitam.edu

N. Sweta
Department Of Computer Science and Engineering, GITAM University Visakhapatnam, Andhra Pradesh,India,
E-mail: n.sweta179@gmail.com

***Abstract*—**Money laundering is a criminal activity to disguise black money as white money. It is a process by which illegal funds and assets are converted into legitimate funds and assets. Money Laundering occurs in three stages: Placement, Layering, and Integration. It leads to various criminal activities like Political corruption, smuggling, financial frauds, etc. In India there is no successful Anti Money laundering techniques which are available. The Reserve Bank of India (RBI), has issued guidelines to identify the suspicious transactions and send it to Financial Intelligence Unit (FIU). FIU verifies if the transaction is actually suspicious or not. This process is time consuming and not suitable to identify the illegal transactions that occurs in the system. To overcome this problem we propose an efficient Anti Money Laundering technique which can able to identify the traversal path of the Laundered money using Hash based Association approach and successful in identifying agent and integrator in the layering stage of Money Laundering by Graph Theoretic Approach.

***Index Terms*—**Data mining, Anti Money Laundering, FIU, Hash Based Mining, and Traversal Path.

I. Introduction

Money laundering is a process of converting unaccountable money in to accountable money. Day to day the technology is getting updated and in this fast changing technology many merits as well as demerits are associated. With the advent of E-Commerce the world has been so globalized and further the technology has made everything so user friendly that with a single click of a button, many transactions can be performed. Fraud Detection is mandatory since it affects not only to the financial institution but also to the entire nation. This criminal activity is appearing more and more sophisticated and perhaps this might be the major reason for the difficulty in fraud detection. This criminal activity leads to various adverse effects ranging from drug trafficking to financial terrorism. Traditional investigative techniques consume numerous man-hours. Data Mining is an area in which huge amounts of data are analyzed in different dimensions and angles and further categorized and then eventually summarized in to useful information. Data Mining is the process of finding correlation or patterns among dozens of fields in large databases. The governing bodies like Reserve Bank of India, Securities and Exchange Board of India have listed out various guidelines to the financial institutions. All the banks collect the list of transactions which is not in accordance with the Reserve Bank of India (RBI) and then submit it to Financial Investigation Unit (FIU) for further investigation. The FIU identifies the money laundering process from the statistical information obtained from various banks. This process is becoming more and more complicated since the count of suspicious transactions is increasing substantially and the rules imposed by RBI alone is not sufficient to monitor this criminal activity. The three stages of money laundering include Placement, Layering and Integration. The placement stage is the stage where in the actual criminal person disposes all the illegal cash to a broker. This broker or agent is responsible for distributing money. In the layering stage the cash is spread into multiple intermediaries that can include banks and other financial institution. The major issue lies in this layering stage of money laundering because here the transfer of money may be from one to one or one-to-many .The difficulty arises in tracing out all the chaining of transactions. In the integration stage all the cash is transferred to a beneficiary often called as Integrator. At this stage all the transactions are made legal. To trace out the dirty proceeds immediately this proposed framework aims at developing an efficient tool for identifying the accounts, transactions and the amount involved in the layering stage of money laundering. The rest of the paper is organized as follows; the literature survey is presented in section-2 of the paper, section3 of the paper deals with the proposed method. Section 4 the experimental analysis and in section 5 the conclusion and Further Enhancement has been explained.

II. Literature Review

Gordon has given the detailed analysis of money laundering. Emerging markets have loose regulations with respect to anti money laundering. Money launderers often set up trade companies and have it acquire highly marketable goods and resell it well below market prices. This gives an unfair advantage to the fraudulent companies because they are not concerned about profits as legitimate business is. This destroys competition in the free market. Also fraudulent companies can obtain much cheaper financing from illegal sources than legitimate businesses that needs financing from free markets. Governments are worried about two implications of money laundering one is, money laundering acts as mechanism to aid terrorist financing and second is, money laundering reduces government tax revenues in a wide variety of ways. The underlying concern is that anti-money laundering efforts, in nature, is a detective mechanism and it will never be able to detect all criminal acts (Killick& Parody, 2007).Anti money laundering detection theories like Know-your-customer, Customer due diligence, monitoring client activities etc were identified by financial industry regulators.

R.cory Watkins et al. [6] has mentioned that from the time of layering stage criminals tries to pretend that laundered money looks like as funds from legal activities and that cannot be differentiated normally. Traditional investigative approaches use to uncover money laundering patterns can be broken down into one of three categories: Identification, detection avoidance and supervision.

Nhien An Le Khac et al. [1][2][3] constructed a data mining based solutions for examining transactions to detect money laundering and suggested an investigating process based on different data mining techniques such as Decision tree, genetic algorithm and fuzzy clustering. By merging natural computing techniques and data mining techniques; knowledge based solution were proposed to detect money laundering. Different approaches were proposed for quick identification of customers for the purpose of application of Anti money laundering. In their paper implemented an approach where in they determined the important factors for investigating money laundering in the investment activities and then proposed an investigating process based on clustering and neural network to detect suspicious cases in the context of money laundering. In order to improve running time heuristics such as suspicious screening were applied.

Yang Qifeng et al. [4] in their paper mentioned that online payment becomes a convenient way to launder money with development of e-commerce. They constructed an anti-money laundering system as a service function of union bank center. This system can monitor and analyze the transaction data dynamically, and provide auxiliary judgment and the decision support for anti-money laundering

Jong Soo Park et al. [8] in their paper examined the issue of mining association rules among items in a large database of sale transactions. The problem of discovering large item sets can be solved by construction of candidate set of item sets first and then identifying within this item set those item sets that need the large item set requirement. The generation of smaller candidate sets enables us to effectively trim the transaction database size at a much earlier stage of the iteration thereby reducing the computational cost.

PankajRichhariya et al.[7] views on fraud detection is owing to levitate and rapid escalation of E-commerce, cases of financial fraud allied with it are also intensifying which results in trouncing of billions of dollars worldwide each year. They provided a comprehensive and review of different techniques like credit card fraud detection, online auction fraud, telecommunication fraud detection, and computer intrusion technique. The disadvantage with the intrusion detection system has poor portability because the system and its rule set must be specific to the environment being monitored.

Jiawei Han et al. [5] in their book focused on improving the efficiency of apriori algorithm and then explained Hash based technique to reduce the size of candidate k –tem set, Ck, for k>1.

Liu khan.et.al [9] proposed a model for identification of suspicious financial transactions using support vector machine. In support vector machine classification the random selection of parameters affects the results and he proposed a method to select appropriate parameters using cross validation. Srrekumar.et.al [10] has reviewed various data mining techniques that are used to detect money laundering which consists of huge amount of banking transactions data from day to day activities. He provided an insight into it.

G.Krishna priya,Dr.M.Prabakaran[11] proposed a time variant approach using the behavioral patterns where the transaction logs are split are various timing windows and depend upon it they generated the behavioral patterns of the customer. By the proposed approach it not only identifies the suspicious accounts but also identifies the group accounts which are involved in money laundering

Denys A. Flores, Olga Angelopoulos [14] projected an useful system for Anti Money Laundering which observe and checks the transactions depending on various techniques. The link analysis is the important technique which is used to make stronger the analyst belief. By combing the rule based approach and risk based approach the risk based approach can achieve the customer profile and transaction risky score. The authors used the clustering module to decrease the false positive alarms that may fatigue the Money laundering investigators

In India the scenario is: at individual level, based on the guidelines given by Reserve Bank of India, Banks determine few transactions which seem to be suspicious and send it to Financial Intelligence Unit (FIU). FIU verifies if the transaction is actually suspicious or not. This process is very time consuming and not suitable to tackle money obtained illegally. Hence it is very important to construct an efficient anti money laundering which goes very helpful for banks to report suspicious transactions. Hence this paper aims to improve the efficiency of the existing anti money laundering techniques. The suspicious accounts of the layering stage of the money laundering process are identified by generating frequent transactional datasets using Hash based Association mining. The generated frequent datasets will then be used in the graph theoretic approach to identify the traversal path of the suspicious transactions.

III. Proposed System

Identifying Money Laundering is very difficult task due to vast number of transactions were involved. To overcome this problem we propose a method which makes use of Hash based association mining for generating frequent transactional datasets and a graph theoretic approach for identifying the traversal path of the suspicious transactions, using which all possible paths between agent and integrator are identified. This Graph Theoretic approach seems to be interesting, because they can detect complex dependencies between transactions. It is also possible to take into account properties and relations of entities involved in sending and receiving the transfers.

The proposed system uses Hashing Technique to generate frequent accounts. A synthetic transactional database was used to experiment the proposed method The same scenario which is similar to the present banking system is considered each individual bank's data that is stored in Databases say Data base1, Database 2...and so on are taken together and combined to form a single large database. Now the data of this large database has to be pre-processed in order to obtain data which is free from all null and missing or incorrect values. A hash based technique is applied on the transactional dataset to obtain a candidate set of reduced size. From this reduced size of candidate set we obtain Frequent-2 item set. This frequent-2 item sets further forms the edges and nodes of the graph. On applying the 'Longest Path in a directed acyclic graph' algorithm we obtain the path in which large amount has been transferred. On the basis of in-degree and out-degree of each node, we determine agent and integrator.

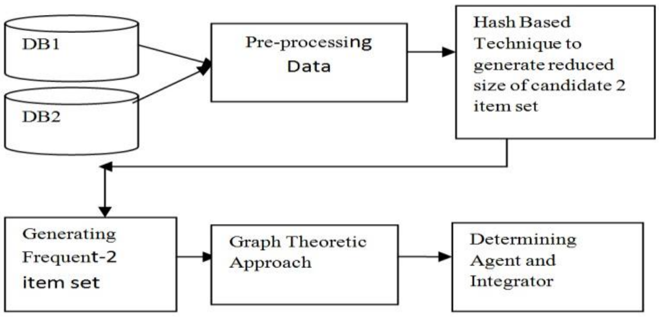
DB1
DB2
Pre-processing Data
Hash Based Technique to generate reduced size of candidate 2 item set
Generating Frequent-2 item set
Graph Theoretic Approach
Determining Agent and Integrator

Fig.1. Proposed System Model for detecting money laundering using Hash Based Association Mining.

The proposed system is represented in two steps

Step1: Applying hash based technique to generate frequent 2 item set. This technique is used to reduce the candidate k-items, Ck, for k>1.

The formula for hash function used here for creating Hash Table is

$$h(x,y) = \big((order\ of\ x * 10) + order\ of\ y\big) mod\ 7 \quad (1)$$

Step 2: Identifying suspicious transactions path using graph theoretic approach

- Linking all the transactions sequentially and generating a graph by considering each account in the frequent item set as a node.
- For each link between the transaction, assign weights to reflect the multiplicity of the occurrence and hence the strength of the path.
- Finding the in-degree and out-degree of each node and determining agent and integrator.

The Hashing Technique which is adopted but the same iterative level wise approach of apriori algorithm is followed. This means that K-item sets are used to explore (k+1) item sets.

A. Hash Based Technique overAprioriAlgorithm:

A hash based technique can be used to reduce the size of the candidate k-item sets, Ck, for k>1. This is because in this technique we apply a hash function to each of the item set of the transaction.

Suppose for equation (1), we have an item set {A1, A4}, then x=1 and y=4.Hence h (1, 4) = ((1*10) +4) mod 7=14 mod 7=0.Now we place {A1, A4} in bucket address 0.

Likewise we fill the hash table and record the bucket count. If any bucket is having count less than the minimum support count, then that whole bucket (i.e. its entire contents) is discarded)All the undeleted bucket counts now form elements of candidate set.

Thus now we have a candidate item set which is smaller in size and hence we need to scan the database less number of times to find the frequent item sets thereby improving the efficiency of apriori algorithm.

Candidate 2-item set generation: All the contents of the undeleted hash table contents are copied and then the duplicate transactions are eliminated. Then we obtain candidate 2 item set.

Transitivity relation: As at a time only 2 accounts are involved in a transaction, to find the chaining of accounts, we have used the mathematical transitivity relation, i.e., if A->B and B->C, then A->B->C.

Frequent 3 item sets: From the transitivity relation we obtain 3 item sets. These item sets have the amount associated with it.

Generating a sequential traversal path: From the frequent accounts, we can create the edges of the graph and also the weight of each edge is equal to the amount transferred between those two accounts.

Longest path in a directed acyclic graph: There are many paths in the graph. Now to find the most suspicious path, we are applying this algorithm and getting the path with the total amount. The entire implementation can be understood by considering a small example of twenty two transactions.

- Step 1: Generating frequent accounts using hashing: Consider a small transaction dataset of 22 transactions. On this set of 22 transactions hash formula is applied equ (1) is applied.

Here x= from_acc_id and y=to_acc_id

Now all these 22 transactions are grouped in to different indexes in hash table. Now the bucket count is calculated for each bucket.

Table 1. Generation of 2 item set using hash based approach

Trans. ID	From-to transaction	2-item set	Trans. ID	From-to transaction	2-item set	Trans. ID	From-to transaction	2-item set
1	A1->A2	{1,2}	9	A4->A5	{4,5}	17	A4->A5	{4,5}
2	A2->A3	{2,3}	10	A1->A2	{1,2}	18	A1->A2	{1,2}
3	A3->A4	{3,4}	11	A5->A6	{5,6}	19	A3->A5	{3,5}
4	A1->A4	{1,4}	12	A3->A5	{3,5}	20	A4->A5	{4,5}
5	A4->A6	{4,6}	13	A3->A6	{3,6}	21	A3->A4	{3,4}
6	A5->A6	{5,6}	14	A1->A2	{1,2}	22	A2->A3	{2,3}
7	A3->A5	{3,5}	15	A3->A5	{3,5}			
8	A3->A6	{3,6}	16	A3->A6	{3,6}			

Now all these 22 transactions are grouped in to different indexes in hash table. Now the bucket count is calculated for each bucket.

Table 2. Bucket table with bucket counts

Bucket Address	0	1	2	3	4	5	6
Bucket contents	1,4 5,6 5,6 3,5 3,5 3,5 3,5	3,6 3,6 3,6	2,3 2,3	4,5 4,5 4,5	4,6	1,2 1,2 1,2 1,2	3,4 3,4
Bucket count	7	3	2	3	1	4	2

Enter the minimum bucket count (say 2).

Then all the bucket whose total count is less than the minimum bucket will be deleted with all its contents. Here bucket 4 is deleted.

Minimum Bucket Count=2

Table 3. Bucket table for item sets and minimum support count

Item set	Bucket Count
1,4	7
5,6	7
3,5	7
3,6	3
2,3	2
4,5	3
4,6	1 (*discarded)
1,2	4
3,4	2

Now the left over transactions in the buckets are taken and then their actual count in database is recorded.

Table 4. The bucket count and actual count are recorded

Item sets	Bucket Count	Actual Count
1,4	7	1 (*discarded)
5,6	7	2
3,5	7	4
3,6	3	3
2,3	2	2
4,5	3	3
1,2	4	4
3,4	2	2

Minimum Support Count =2

Now all the transactions which have occurred 2 or more no of times are taken in to Frequent -2 item sets. Thus the obtained frequent-2 Transactions are:

Table 5.Frequent 2 accounts with their support count

Frequent-2 Item set	Support Count
5,6	2
3,5	4
3,6	3
2,3	2
4,5	3
1,2	4
3,4	2

- Step 2: *Finding the traversal path*

Various paths are identified by connecting all the frequent accounts as nodes.

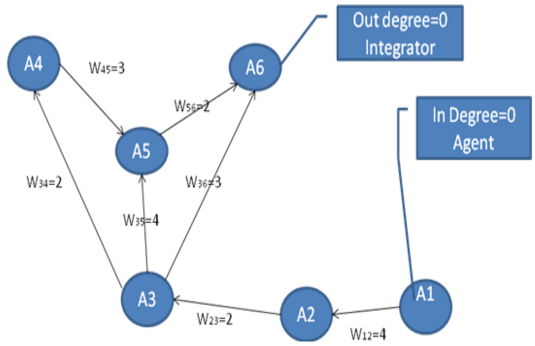

Fig.2. Identifying agent and integrator using graph theoretic approach

IV. Experimental Analysis

The proposed system is accessed using a 17000. We have taken a synthetic transaction dataset from multiple banks over a period of 120 days. We consider here transaction datasets up to seventeen thousand sizes. For different dataset we select different tables and enter different values of support/threshold count. We first apply hashing technique to generate frequent 2 item sets.

Table 6. Experiments carried over different Data sets

S.no	Size of dataset	Minimum bucket count	Minimum support count	No of frequent transactions
1	22	2	2	7
2	5000	700	3	77
3	10000	1400	4	103
4	17221	2500	6	36

This is the result obtained after applying Hashing Technique. We observe that when the no of transactions in the data set increases, the no of frequent account also increases. However if we increase the support count value to a large value then a few no of frequent transactions are obtained as in the last case of our experiment.

A graph can be plotted as Size of dataset v/s No of frequent transactions.

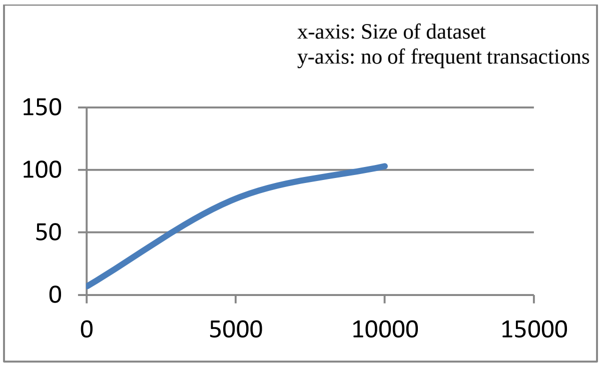

Fig.3. GraphofsizeofDatasetsv/s No of frequent transactions

The Algorithm to find the longest path in a directed acyclic graph is applied on these frequent transactions in each of the above four cases. We get the following results.

Table 7.Identifying the longest path by varying the size of datasets and support count

S.No	Size of Dataset	Min support count	Longest Path	Total Amount
1	22	2	1->2->3->5->6	Rs.2,41,616
2	5000	3	98->63->44->27	Rs.2,96,428
3	10000	4	8->18->20->52->97->3->64->5->26->42->56->99->77	Rs.1,08,86,274
4	17221	6	107->54->33	Rs. 4,95,721

Similarly we can find a path between any two frequent accounts using this algorithm.

For Eg: we can find the path between 31 and 49 in 10000 dataset transaction giving minimum bucket count as 1400 and minimum support count as 4.

31->96->86->36->49.

V. Conclusions & Future Work

The proposed system improve the efficiency of the existing anti money laundering techniques by identifying the suspicious accounts in the layering stage of money laundering process by generating frequent transactional datasets using Hash based Association mining. The generated frequent datasets will then be used in the graph theoretic approach to identify the traversal path of the suspicious transactions. We were successful in finding the agent and integrator in the transaction path. In our solution, we have considered the frequent accounts as the parameter and have obtained a chaining of accounts. These accounts have the highest possibility of being suspicious as there are involved in huge amount of transactions frequently. The solution proposed here is highly advantageous over the existing anti-money laundering rules.

Further enhancement: With the chaining of accounts, we can further develop a system which identifies the sure relation between these identified suspicious accounts using concepts like ontology. The relation between these accounts can give us additional information like whether the involved criminal people are belonging to same occupation or to the same location etc.

The frequent accounts should not be the only criteria for finding out the suspicious transaction as there may be a case when the transaction does not occur frequently but even then they are illegal. To trace out such cases additional parameters have to be considered.

Acknowledgment

I acknowledge the management, principal and coordinator of TEQIP for providing the Research assistantship to carry out the above work under TEQIP phase-II, sub component 1.2 from M.H.R.D Government of India.

References

[1] NhienAn Le Khac, SammerMarkos, M. O'Neill, A. Brabazon and M-TaharKechadi. An investigation into Data Mining approaches for Anti Money Laundering. In International conference on Computer Engineering & Applications 2009.

[2] Nhien An Le Khac, M.Teharkechadi. Application of Data mining for Anti-money Detection: A case study. IEEE International conference on Data mining workshops 2010.

[3] Nhien An Le Khac, SammerMarkos,M.Teharkechadi,. A data mining based solution for detecting suspicious money laundering cases in an investment bank. IEEE Computer society 2010.

[4] Yang Qifeng, Feng Bin, Song Ping. Study on Anti Money Laundering Service System of Online Payment based on Union-Bank mode. IEEE Computer Society 2007.

[5] J.Han and M. Kamber, Data Mining: Concepts and Techniques. Morgan Kaufmann publishers, 2nd Eds., Nov 2005.

[6] R.Corywatkins, K.Michaelreynolds, Ron Demara. Tracking Dirty Proceeds: Exploring Data Mining Techniques as to Investigate Money Laundering. In police practice and research 2003.

[7] PankajRichhariya,PrashantK.Singh,EnduDuneja. A Survey on financial fraud detection methodologies. In International Journal of commerce business and management 2012.

[8] J.S.Park, M.S.Chen, and P.S.Yu. An effective hash-based algorithm for mining association rules. In Proc. 1995 ACM-SIGMOD Int.Conf.Management of Data (SIGMOD'95), pages 175-186, San Jose, CA, May 1995.

[9] Liu Keyan and Yu Tingting,"An improved Support vector Network Model for Anti-Money Laundering, International conference on Management of e-commerce ande-Government.

[10] SreekumarPulakkazhy and R.V.S.Balan,"Data Mining in Banking and its applications –A Review", Journal of computer science 2013.G.

[11] G.Krishna priya,Dr.M.Prabakaran"Money laundering analysis based on Time variant Behavioral transaction patterns using Data mining"Journal of Theoretical and Applied Information Technology 2014.

[12] Xingrong Luo,"Suspicious transaction detection for Anti Money Laundering", International Journal of Security and Its Applications 2014.

[13] ch suresh,Prof.K.Thammi Reddy,"A Graph based approach to identify suspicious accounts in the layering stage of Money laundering",Global Journal of computer science and Information Technology 2014.

[14] Denys A.Flores, Olga Angelopoulou, Richard J. Self," Design of a Monitor for Detecting Money Laundering and Terrorist Financing", International Journal of Computer Networks and Applications 2014.

[15] Anu and Dr. Rajan Vohra," Identifying Suspicious Transactions in Financial Intelligence Service", International Journal of Computer Science & Management Studies July 2014.

[16] Angela Samantha Maitland Irwin and Kim-Kwang Raymond Choo," Modelling of money laundering and terrorism financing typologies",Journal of Money laundering control 2012.

[17] Pamela Castellón González,,Juan D. Velásquez ," Characterization and detection of taxpayers with false invoices using data mining techniques",Expert Systems with Applications ,Elsevier 2013.

[18] Rafal Drezeswski,Jan sepielak,Wojciech Filip Kowsiki,"System supporting Money Laundering detection", Elsevier 2012.

[19] Quratulain Rajput, Nida Sadaf Khan, Asma Larik, Sajjad Haider, "Ontology Based Expert-System for Suspicious Transactions Detection", Canadian Center of Science and Education, Computer and Information Science; Vol. 7, No. 1, 2014.

[20] Mahesh Kharote, V. P. Kshirsagar, "Data Mining Model for Money Laundering Detection in Financial Domain", International Journal of Computer Applications (0975 – 8887), Volume 85 – No 16, 2014.

[21] Harmeet Kaur Khanuja, Dattatraya S. Adane, "Forensic Analysis for Monitoring Database Transactions", Springer, Computer and Information Science Volume 467, pp 201-210, 2014.

[22] Pradnya Kanhere, H. K. Khanuja," A Survey on Outlier Detection in Financial Transactions,International Journal of computer Applications, December2014.

Authors' Profiles

CH.Suresh currently pursuing full time Ph.D. in the area of "Data mining" in the Dept. of CSE, GIT, GITAM University**.** His research includes Data warehousing and Mining, Database management system, operating system etc.

Dr. K. Thammi Reddy, currently working as the Director of Internal Quality Assurance Cell (IQAC) and Professor of CSE. At Gandhi Institute of Technology (GITAM) University, Visakhapatnam. He is having vast experience in teaching, Research, Curriculum Design and consultancy. His research areas include Data warehousing and Mining, Distributed computing, etc

Sweta.N, currently a B.Tech graduate from GITAM University in 2015. Her area of interest includes Problem Solving using Coding, i.e., understanding the problem and finding a code which could run and generate the output as the solution.

How to cite this paper: Ch.Suresh, K.Thammi Reddy, N. Sweta,"A Hybrid Approach for Detecting Suspicious Accounts in Money Laundering Using Data Mining Techniques", International Journal of Information Technology and Computer Science(IJITCS), Vol.8, No.5, pp.37-43, 2016. DOI: 10.5815/ijitcs.2016.05.04