Data Mining Model for Money Laundering Detection in Financial Domain

Mahesh Kharote
Research Scholar
Government College of Engineering
Aurangabad, India

V. P. Kshirsagar
Head of the Department
Government College of Engineering
Aurangabad, India

ABSTRACT

One of the likeliest problems faced by the banks is the way the nature of money transaction takes place is not always known to the bank. One such problem consisting large amounts of money transferring through various accounts by the same person or entity is Money Laundering. Money laundering system is quite a convoluted process. It takes some understanding of the fund transferring activities at various stages. The system that works against Money laundering is Anti-Money Laundering (AML) system. The existing system for Anti-Money Laundering accepts bulk of data and converts it to large volumes reports that are tedious and topsy-turvy for a person to read without any help of software. Hence, the purpose of making decision within specified time span is de-facilitated. The basic motive behind designing a system is that it should either notify or generate alarm for Money Laundering process within time when the illicit actions are been carried out so that, some action can be taken or an decision could be made before the actual laundering task is completed. The Transaction flow analysis (TFA) is a system that facilitates the basic ideology of Anti-Money Laundering system. TFA consists of various segments those results in clusters of various sizes. However, the existing TFA system allows the data to be taken from a single entity may be bank or a financial institution. Also the cluster generated are large in volume hence the analyst sometime may find it difficult to analyze such huge amount of data. This project uses combination of TFA system with the system designed to extract customer behavior in order to make decision whether the customer is suspicious customer or not. Also this project will allow taking data as input from various sources, integrating and then performing operations on the data.

General Terms
Laundering

Keywords
Transaction analysis, clustering, frequent pattern mining, suspicious transaction, suspicious customer, visualization.

1. INTRODUCTION
This paper deals with Money laundering and its various stages where the illegal sometimes referred as dirty money is passed through a cycle or iterative transaction so as when the money comes out of the cycle it appears to be cleaned money or legal money. This process is referred to as washing. The illegal money is that where the money obtained by the bearer is not notified anywhere; hence the rightful tax that needs to be paid is vanished. According to the income tax act 1961, any individual or organization whose income is above the specified limit must pay the tax that is applicable to their income. When the money is obtained from sources that are illegal the money is said to be illegal. Gambling is illegal in India, and most of the illegal money comes through it. Through gambling huge amount of money can be gained over night. Most of the times the money obtained through gambling are received by the gambler in cash.

The gambler puts the money through cleaning cycle and when the money is cleaned he then uses it for other things. Money Laundering basically is carried out in three stages [2].

1. **Placement: -** The first and the riskiest stage. The gambler when receives the huge amount he places it in a bank account or over several accounts.

2. **Layering: -** The amount then is wired to several overseas bank accounts where the banks follow strict secrecy codes, or may be transferred to a single account.

3. **Integration: -** Now once the money is in the cycle it is used to make purchase of luxury assets and hence making the money appear to be legal.

The diagrammatical representation of all the stages is shown below.

Fig 1: Generic process of Money Laundering

Our focus would be on these transactions where such huge amount of money is deposited by cash or over various accounts by the same person; hence detecting he may be a suspicious customer. Transaction Importer part will be used for transforming data in the required format. Various Data mining tools such as WEKA, IBM SPSS would be used for applying data mining methods. At the later stage the clusters would be examined using Extracting Customer Behavior [2] which will be helpful for the analysis for the analyzers and for the

researchers as human analyst may lack an understanding of the hidden patterns in business data.

2. LITERATURE SURVEY

This section consists of brief description of few data mining techniques.

2.1 Money Laundering Detection Using TFA System

This method was proposed P. Umadevi and E.Divya. In this paper the Transaction Flow Analysis (TFA) method was proposed for detecting money laundering. The method prior to TFA simply transformed vast quantity of data into vast numbers of reports that do not facilitate timely detection or effective interdiction [2].

2.2 PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth

This method was proposed by Jian Pei Jiawei Han Behzad Mortazavi-Asl Helen Pinto. PrefixSpan mines the complete set of patterns but greatly reduces the efforts of candidate subsequence generation. Adjoining that, prefix-projection heavily reduces the size of the projected data and leads to efficient processing. Its performance study shows that PrefixSpan outperforms both the Apriori-based GSP algorithm [4].

2.3 A Financial Data Mining Model for Extracting Customer Behavior

This method was proposed by Mark K.Y.Mak, George T.S. Ho and S.L.Ting. This paper aims at developing an intelligent Financial Data Mining Model (FDMM) for extracting customer behavior in the financial industry, so as to increase the availability of decision support data and hence increase customer satisfaction. The proposed system clusters the records into several sectors, and then finds the correlation among these sectors. It is noted that better customer segmentation can increase the ability to identify targeted customers; therefore extracting useful rules for specific clusters can provide an insight into customers buying behaviour and marketing implications. [1]

2.4 A Decision Support Based on Data Mining in e-Banking

This method was proposed by Irina Ionita and Liviu Ionita. This paper describe briefly the use of data mining techniques in banking domain is suitable due to the nature and sensitivity of bank data and due to the real time complex decision processes. The main concern for a bank is to take good decisions in order to minimize the risks level associated to bank. It is impeccable that the bank should have knowledge of causes which generate the financial crises or imbalances. [3]

2.5 Role of Data Mining in Banking Sector

This concept was proposed by Vivek Bhambri. Banking Sector all over the world is going through updates almost everyday. The computerization of financial institutions, connected via internet and the support of automated softwares has completely changed the work culture of banking. The banks are now getting aware of the fact that their biggest asset is the knowledge and not the financial resources and are tending to apply the techniques of data mining for assisting them in their activities [10].

3. PROPOSED SYSTEM

In this section, the main idea used in the proposed technique is described. The proposed system consists of seven modules: Preprocessing, Data Importer, Clustering and Suspected sequence/sets, Data visualization, Learning from the user decision, company/organization profile generation, extracting behavior. In preprocessing the gathering of data from various data feeds is accomplished and performing preprocessing operation so that data is available in required format. K-means cluster are used in clustering and finding suspected sequence/sets. Frequent Pattern Mining is used for learning purpose. And finally, extraction of customer behavior is done by combination of clustering and association rule mining.

The proposed system consists of two algorithms and the architecture of the system is as follows, the flow of the proposed system is divided in seven phases as depicted in Fig 2.
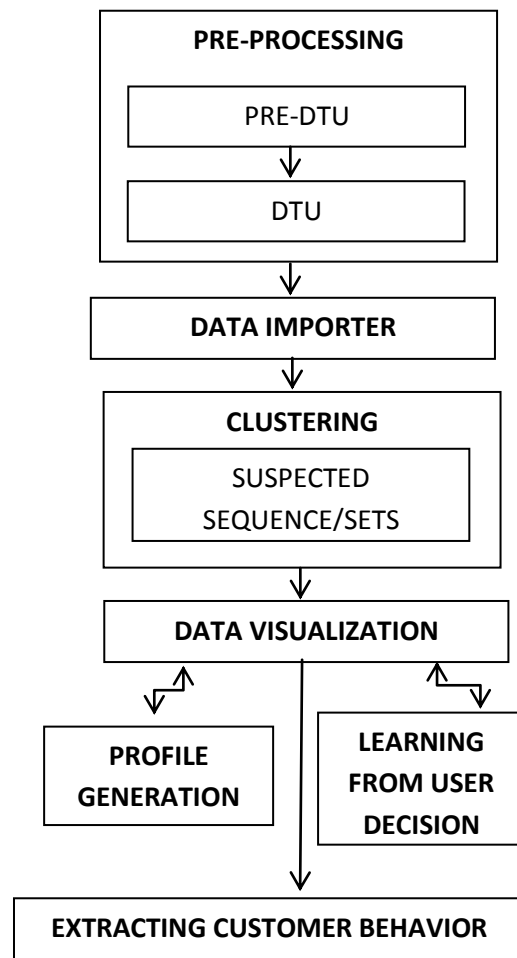

Fig 2: Architecture of the proposed system

3.1 MODULES

The following sections explain the various model of the proposed system.

3.1.1 Preprocessing

The preprocessing module consists of two sub-modules Pre-DTU (Data Transform Unit) and DTU. This module takes raw input data and makes the data available for the Data importer. The main purpose of this module is to integrate data from various sources and transform data in required format and make available to Data Importer.

3.1.1.1 Pre-DTU

The data required for any data mining purpose is not always available on a single source or in the specified format. At this stage gathering of data from various sources and integrating them will be performed. Now, the integrated data may consist of duplicate records hence, after combining data, cleansing operation is performed to remove duplicate records from data if any and make it available to DTU.

3.1.1.2 DTU

The DTU takes input from Pre-DTU and the data is transformed in the required format.

3.1.2 Data Importer

The Data importer takes input data from preprocessing module. This is the stage is considered as entry point into the system. The Data Importer reads the data and transforms the data in system specific format.

3.1.3 Clustering

This module analyzes imported data using clustering algorithm. On the output it provides clusters which are input of frequent pattern mining algorithms.

3.1.3.1 Suspected Sequence/Sets

This module analyzes the clusters with the use of frequent pattern mining algorithms. On the output it provides frequent patterns that are found (sets or sequences).

3.1.4 Data Visualization

This module visualizes the results on dedicated schema and time line diagrams.

3.1.5 Profile Generation

This module performs automatic generation of profiles of suspected companies/organizations typical activities, includes time/date of making money transfers, headers of money transfer, amount usually being transferred, etc. On the basis of such data, the profile of the company may be created and used during further filtering of suspected sequences of money transfers. On the output it provides a profile of typical activities of suspected companies/organizations/individuals.

3.1.6 Learning from user decision

This module deals with the user interacting with the system and learns its decisions during the analysis of suspected sequences of transfers. Then the constructed user profile may be used by the system traditionally filter suspected sequences on the basis of the previous user's decisions. On the output it provides a model with the user's decisions.

3.1.7 Extracting customer behavior

Once the user profile is constructed precise analyses of the user's decision is performed and later track the deviation of decisions from normal profile and hence, whether the suspicious companies/organization/individual is indulging in any kind of illicit operation that with time may turn into some Money Laundering Process is learned.

3.2 ALGORITHMS AND DATA MINING TECHNIQUES

This section describes algorithms and data mining techniques that would be used in the proposed system.

3.2.1 K-means clustering

One of the unsupervised learning algorithms for clusters is K-means. Grouping the records according to some characteristics is called clustering the data. Initially, the numbers of clusters k in the k-means algorithm are defined. Before the clustering, the amounts from the data may be recorded in different currencies and hence are modified to single currency. This operation is carried out at Data Importer module. The algorithm for clustering is as shown below. [2]

Algorithm:
Input
  * K: the number of cluster
  * D: a data set contains n objects
Output
  A set of K cluster
Method

1. Arbitrarily choose k objects from D as the initial cluster centres.

2. Repeat.

3. (Re) assign each object to the cluster to which the object is the most similar, based on the mean value of the object in the cluster.

4. Update the cluster means, i.e., calculates the mean value of the object for each cluster.

5. Repeat steps 3 and 4 until no change.

3.2.2 Frequent Pattern Mining

Frequent pattern mining was first proposed by Agrawal et al. (1993) [8] for market basket analysis in the form of association rule mining. To mine frequent sequences, following algorithms are implemented: Sequence Miner [4] and BIDE. Algorithm sequence miner is similar to Prefix Span (peietal., 2001) algorithm and allows mining all of the frequent sequence. The algorithm was elaborated by J. Wang and J. Han (Wang and Han, 2004). It does not contain any search space pruning method. To speed up support computing for the sequences that are found, pseudo projection is applied. The general idea of the algorithm relies on examining only prefixes of sequences. Prefixes grow using sets of its locally frequent items. Algorithm shows Sequence Miner algorithm. Procedure Frequent Sequences is called recursively.

Algorithm

1 SequenceMiner(SDB, min_sup, FS)
  Data: SDB – an input sequence database, min_sup – a minimum support threshold
  Result: The complete set of frequent sequences, FS

2 FS = ϕ

3 Frequent Sequences(SDB,ϕ, min_sup,FS);

4 return FS;

5 FrequentSequences(S_p_SDB, S_p,min_sup,FS)
  Data: S_p _SDB – a projected sequence database, S_p – a prefix sequence, min_sup – a minimum support. Result: the current set of frequent sequences, FS

6 if S_p is non-empty then

7 FS = FS U S_p;

8 end

9 if LF_S_p= locally frequent items(S_p_SDB, S_p, min_sup);

10 if LF_S_p is empty then

11 return;

12 end

13 foreach locally frequent item i do

14    $S_p{}^i = S_p \cup i$;

15    $S_p{}^i$_SDB = pseudo projected database($S_p{}^i$, S_p_SDB);

16   FrequentSequence($S_p{}^i$_SDB, $S_p{}^i$, min_sup, FS);

17   end

4. CONCLUSION

The money laundering system involves transfer of huge amount of money to some accounts. Preprocessing makes sure the data provided to the system is always the correct one and reduced as possible. The less the data the faster are the operations performed. At the output of the clustering module along with graph structure which will be having all the records involving huge amounts of money transfer the rest further process is to learn exactly which account has been indulging in money laundering practices. To cluster the records K-means algorithm is sufficient. And to extract pattern from records and learns user decision pattern, Frequent Pattern Mining and the Association Rules [1] would be used. A minimum support threshold for the transfers from an account or by individual over several accounts is predefined and then measures the deviation from the normal profile. The software intervention for the analysis purpose will provide more accuracy and save more time.

5. ACKNOWLEDGMENT

6. REFERENCES

[1] Mark K.Y. Mak, George T.S. Ho, S.L.Ting "A Financial Data Mining Model for extracting Customer Behavior" 2011.

[2] P. Umadevi, E.Divya "MONEY LAUNDERING DETECTION USING TFA SYSTEM" ICSEMA, International Conference on Digital Object Identifier, 2012.

[3] Irina Ionita, Liviu Ionita "A Decision Support Based on Data Mining in e-Banking" 2011.

[4] Jian Pei, Jiawei Han, Behzad Mortazavi-Asl, Helen Pinto "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern" 2004.

[5] Rujuta Shinde, Priya Vaghurdekar, Santaji Shinde "Deliberation of Data Mining in Banking" International Journal of Engineering Reaserch & Technology. ISSN: 2278-0181 Vol 1 Issue8, 2012.

[6] Lv LT, Ji N, Zhang JL "A RBF neural network model for anti-money laundering". ICWAPR Internation Conference, 2008.

[7] Kazi Imran Moin, Dr. Qazi Baseer Ahmed " Use of Data Mining in Banking" International Journal of Engineering Research and Applications (IJERA) Vol. 2, Issue 2, Mar-Apr 2012.

[8] Agrawal, R., Imielinski, T., Swami, A.N. "Data Mining: A performance perspective", IEEE Transactions on Knowledge and Data Engineering, vol, 5, no. 6, 1993.

[9] Nenad, J., Svetlozar, N. "Comprehensive data warehouse exploration with qualified association-rule mining, Decision Support Systems", vol. 42, no. 2, 2006.

[10] Vivek Bhambri "Role of Data Mining in Banking Sector", International Indexed & Reffered Research Journal Vol. 3 Issue-33, 2012.

[11] Dr. Madan Lal Bhasin, "Data Mining: A Competitive Tool in the Banking and Retail Industries", The Chartered Accountant October 2006.