ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 2, Issue 8 , August 2015

Data Mining Techniques for Anti Money Laundering

Manjunath K.V.

Director, Samvardhana Coaching Centre, Bangalore, India

ABSTRACT: Money Laundering is the process of creating the appearance that large amounts of money obtained from serious crimes, such as drug trafficking or terrorist activity, originated from a legitimate source. Through money laundering, the launderer transforms the monetary proceeds derived from criminal activity into funds with an apparently legal source. 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. Data mining approaches have been developed and are considered as well-suited techniques for detecting ML activities.

KEYWORDS: Data mining, Anti Money Laundering, Knowledge Discovery in Databases, Clustering, Association, Regression.

I. INTRODUCTION

A. Money Laundering:

Money laundering is the process of taking cash earned from illicit activities such as drug trafficking, and making the cash appears to be earnings from a legal business activity. The money from the illicit activity is considered dirty and the process "launders" the money to make it look clean. Money laundering is the generic term used to describe the process by which criminals disguise the original ownership and control of the proceeds of criminal conduct by making such proceeds appear to have derived from a legitimate source.

Illegally earned money needs laundering in order for the criminal organization to use it effectively. Dealing in large amounts of illegal cash is inefficient and dangerous. The criminals need a way to deposit the money in financial institutions, yet they can only do so if the money appears to come from legitimate sources. There are many ways to launder money. These methods span from the very simple to the very complex. One of the most common ways is to launder the money through a legitimate cash-based business owned by the criminal organization. For instance, if the organization owns a restaurant, it might inflate the daily cash receipts to funnel its illegal cash through the restaurant and into the bank. Then they can distribute the funds to the owners out of the restaurant's bank account.

There are no firm statistics, but it is estimated that as much as $500 billion dollars in illegal funds is laundered each year.

Money-laundering is a dynamic three-stage process that requires

- **a. Placement:** This is the movement of cash from its source. On occasion the source can be easily disguised or misrepresented. This is followed by placing it into circulation through financial institutions, casinos, shops, bureau de change and other businesses, both local and abroad. The process of placement can be carried out through many processes
- **b. Layering:** The purpose of this stage is to make it more difficult to detect and uncover a laundering activity. It is meant to make the trailing of illegal proceeds difficult for the law enforcement agencies.
- **c. Integration:** This is the movement of previously laundered money into the economy mainly through the banking system and thus such monies appear to be normal business earnings. This is dissimilar to layering, for in the integration process detection and identification of laundered funds is provided through informants.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 2, Issue 8 , August 2015

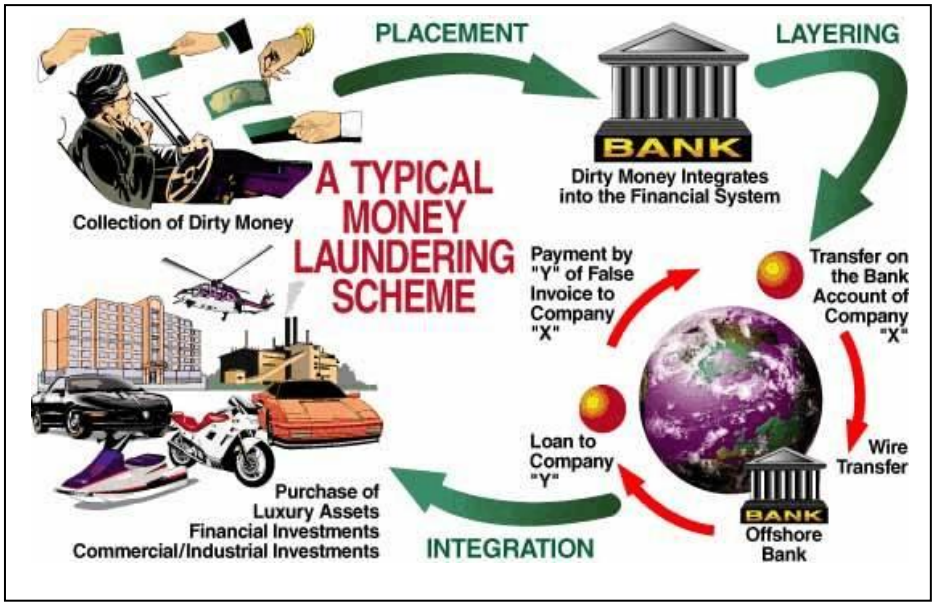

Fig. 1 Process of money laundering

B. Anti Money Laundering (AML):

AML is a set of procedures, laws or regulations designed to stop the practice of generating income through illegal actions. In most cases money launderers hide their actions through a series of steps that make it look like money coming from illegal or unethical sources was earned legitimately. Anti-money laundering software is a type of computer program used by financial institutions to analyze customer data and detect suspicious transactions. Anti-money laundering systems filter customer data, classify it according to level of suspicion and inspect it for anomalies. Such anomalies would include any sudden and substantial increase in funds or a large withdrawal. Smaller transactions that meet certain criteria may be also are flagged as suspicious.
For example, a person who wants to avoid detection will sometimes deposit a large sum as multiple smaller sums within a brief period of time. That practice, known as "structuring," will also lead to flagged transactions.

C. Data mining:

Data mining is the practice of automatically searching large stores of data to discover patterns and trends that go beyond simple analysis. Data mining uses sophisticated mathematical algorithms to segment the data and evaluate the probability of future events.
The key properties of data mining are:

- Automatic discovery of patterns
- Prediction of likely outcomes
- Creation of actionable information
- Focus on large data sets and databases

II. KNOWLEDGE DISCOVERY IN DATABASES (KDD)

Data mining is the core part of the knowledge discovery in databases. The KDD process consists of the following steps: data selection, data cleaning, data transformation, pattern searching (data mining), and finding presentation, finding interpretation and finding evaluation. Data mining and KDD are often used interchangeably because data mining is the key part of KDD process.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 2, Issue 8 , August 2015

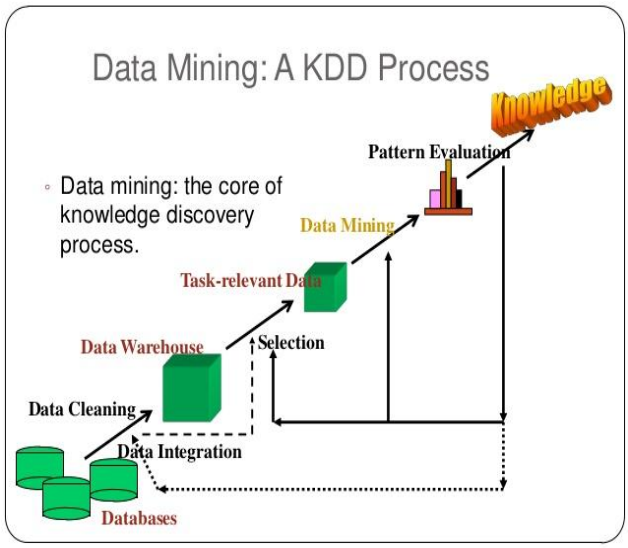

Figure 2: A KDD Process

III. GOALS OF DATA MINING

The two "high-level" primary goals of data mining in practice are *prediction* and *description*.

1. **Prediction** involves using some variables or fields in the database to predict unknown or future values of other variables of interest.
2. **Description** focuses on finding human-interpretable patterns describing the data.

The relative importance of prediction and description for particular data mining applications can vary considerably. However, in the context of KDD, description tends to be more important than prediction. This is in contrast to pattern recognition and machine learning applications where prediction is often the primary goal of the KDD process.

IV. DATA MINING TECHNIQUES FOR AML

A. Clustering:

Clustering is the process of grouping the data into classes so that objects within the same cluster have high similarity and objects within different clusters are very dissimilar. There are different clustering methods in the literature and they have been successfully exploited for scientific datasets, spatial datasets, business datasets, etc.

In AML, clustering is normally used for grouping transactions/accounts into clusters based on their similarities. This technique helps in building patterns of suspicious sequence of transactions and detecting risk patterns of customers/account.

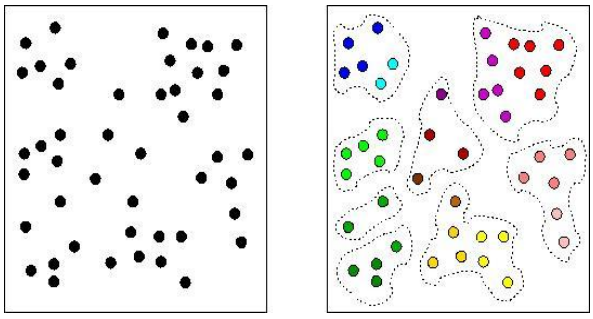

Figure 3: A set of points on a plane (on the left) can be clustered in various ways, one possible clustering is shown (on the right).

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 2, Issue 8 , August 2015

Types of Clustering

- Partitioned Clustering – A division data objects into non-overlapping subsets (clusters) such that each data object is in exactly one subset.
- Hierarchical clustering – A set of nested clusters organized as a hierarchical tree

B. Classification and Prediction:

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and credit risk applications are particularly well suited to this type of analysis. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. For a fraud detection application, this would include complete records of both fraudulent and valid activities determined on a record-by-record basis. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier.

Types of classification models

- Classification by decision tree induction
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

Classification works with discrete and unordered data and helps to identify class labels of the members of the population.
But **prediction models** works with continuous-valued functions. That is, it is used to predict missing or unavailable numerical data values from the sample attribute values. Commonly used technique for prediction is regression analysis. It is a statistical methodology that is used to forecast values from existing numerical values. In predictive models for data mining, we have a set of independent variables whose values are already known and a set of dependent or response variables whose values we want to predict. Regression helps us to express the relationship between these variables as a linear or non-linear function. In many real world problems related to banking such as stock price predictions, or credit scoring follow complex models with many independent variables and requires multidimensional regression analysis and logistic regression.
Stefan Axelsson et al.[8] 2012, analysed the implications of using machine learning techniques for money laundering detection in a data set consisting of synthetic financial transactions and aimed to detect anomalies inside a data set of mobile money financial transactions by using the classification techniques to group transactions as suspicious or non-suspicious. Xingqi Wang et al.[14] 2009, proposed a novel algorithm to detect money laundering using an improved minimum spanning tree clustering, an analysis of similarity measure and distance metric.
In classification approach, risk levels are organized into two categories based on past default history. For example, customers with past default history can be classified into "risky" group, whereas the rest are placed as "safe" group.

V. MONEY LAUNDERING DETECTION

Money Laundering is the process of hiding the illegal origin of “black” money so as to legitimize it. Banks are commonly used as channels to launder money. Therefore governments and financial regulators require banks to implement processes, systems and procedures to detect and prevent money laundering transactions. Failure to detect and prevent such illegal transactions can invite hefty fines both monetarily and operationally which can prove very costly for the bank and even can make its survival difficult. Conventional rule-based transaction analysis based on reports and tools will not be sufficient to detect more complicated transaction patterns. Here data mining techniques can be applied to dig out transaction patterns that can lead to money laundering. Typically such systems take client risk assessment data, transaction risk measurement data and patterns and behavior patterns into consideration for detecting money laundering patterns. Transactions are then grouped into clusters based on their similarities found in these chosen

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 2, Issue 8 , August 2015

attributes. In a large database of banking transactions, it is possible that a huge number of patterns emerge and will be classified as money laundering transactions thereby increasing false positives. Statistical false reduction methods based on decision tree classification are employed to limit the number of false patterns detected.

VI. CONCLUSION

Data mining is a process to extract knowledge from existing data. It is used as a tool in banking and finance in general to discover useful information from the operational and historical data to enable better decision-making. It is an interdisciplinary field, confluence of Statistics, Database technology, Information science, Machine learning and Visualization. It involves steps that include data selection, data integration, data transformation, data mining, pattern evaluation, knowledge presentation. Banks use data mining in various application areas like marketing, fraud detection, risk management, money laundering detection and investment banking. The patterns detected help the bank to forecast future events that can help in its decision-making processes. More and more banks are investing in data mining technologies to be more competitive.

REFERENCES

[1] Harmeet Kaur Khanuja, Dattatraya S. Adane, "Forensic Analysis for Monitoring Database Transactions", Springer, Computer and Information Science Volume 467, pp 201-210, 2014
[2] Pradnya Kanhere and H. K. Khanuja "A Survey on Outlier Detection in Financial Transactions", International Journal of Computer Applications (0975 – 8887) Volume 108 – No 17, P.No 23-25, December 2014
[3] 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
[4] Bhambri, V. "Application of data mining in banking sector". Internat. J. Comput. Sci. Technol., 2: 199-201, 2011
[5] Nida S. Khan, Asma S. Larik, Quratulain Rajput, Sajjad Haider, "A Bayesian approach for suspicious financial activity reporting", International Journal of Computers and Applications, Vol. 35, No. 4, 2013
[6] Chopra, B., V. Bhambri and B. Krishnan, "Implementation of data mining techniques for strategic CRM issues". Int. J. Comput. Technol Appli., 2: 879-883, 2011
[7] Shashidhar, H.V. and S. Varadarajan, "Customer segmentation of bank based on data mining-security value based heuristic approach as a replacement to k-means segmentation". Int. J. Comput. Appli., 19:13-18, 2011
[8] P .Umadevi, E .Divya "Money laundering detection using TFA system", in IET publisher, 2012.
[9] Stefan Axelsson, "Money Laundering Detection using Synthetic Data", The 27th annual workshop of the Swedish Artificial Intelligence Society (SAIS); 14–15 May 2012
[10] Laisheng Xiao and Zhengxia Wang, "Internet of Things: a New Application for Intelligent Traffic Monitoring System", Journal of Networks, vol. 6, no. 6,P.No-887-894,JUNE 2011

AUTHOR'S BIOGRAPHY:

Manjunath K.V received MCA from VTU Belgaum, MBA(HR) from Sikkim Manipal University, Ganktok and M.Phil (CS) from Vianayaka mission University in 2002, 2008 and 2011 respectively. He is currently the Director of Samvardhana Coaching Centre, Bangalore, India. His research interests include data mining, distributed systems, Information Security and cloud computing.