DATA MINING TECHNIQUES FOR ANTI MONEY LAUNDERING

David Chuparkoski

University of agribusiness and rural development, Plovdiv, Bulgaria

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 the 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. To develop a structure to research in datamining, we create a taxonomy that combines research on patterns of observed fraud schemes with an appreciation of areas that benefit from a productive application of data mining. The aim of this study was to review research conducted in the field of fraud detection with an emphasis on detecting honey laundering and examine deficiencies based on data mining techniques. Which include a set of predefined rules and threshold values. In addition to this approach, data mining techniques are very convenient to detest money laundering patterns and detect unusual behavior. Therefore, unsupervised data mining technique will be more effective to detect new patterns of money laundering and can be crucial to enhance learning models based on classification methods. Of course, the development of new methods will be very useful to increase the accuracy of performance.

Keywords: financial fraud, fraud detection, money laundering detection, data mining, anomaly detection.

Introduction

Detecting management fraud is a difficult task when using normal audit procedures. First, there is a shortage of knowledge concerning the characteristics of management fraud. Secondly, given its infrequency, most auditors lack the experience necessary to detect it. Finally, managers deliberately try to deceive auditors. For such managers, who comprehend the limitations of any audit, standard auditing procedures may prove insufficient. These limitations suggest that there is a need for additional analytical procedures for the effective detection of management fraud. It has also been noted that the increased emphasis on system assessment is at odds with the profession's position regarding fraud detection since most material frauds originate at the top levels of the organization, where controls and systems are least prevalent and efficient (Kirkos et al, 2007).

Applying data mining to fraud detection as part of a routine financial audit can be challenging and, as we will explain, data mining should be used when the potential payoff is high. In general, when it comes to fraud detection for a given audit client, the audit team would make three major decisions:

(1) What specific types of fraud (e.g., revenue recognition, understated liabilities, etc.) should be included in the audit plan for a particular client?

(2) What sources of data (e.g., journal entries, emails, etc.) would be provided evidence of each type of fraud?

(3) Which data mining technique(s) (e.g., directed or undirected techniques) would be the most effective for finding potential evidence of fraud in the selected data? Developing answers for

each of these questions are significant individually, but, in combination, answering these questions is challenging.

These challenges may encourage the audit team to continue to use traditional – but less diagnostic – analytical and substantive procedures. However, as we will discuss in this paper, each of the populations for each of these three questions can be intelligently reduced so that the application of data mining to fraud detection becomes more manageable and will have a higher potential for a successful payoff. We also recognize that data mining techniques and associated software can have a steep learning curve. Further, if used improperly, data mining can produce many false positives and spurious patterns that will require auditors to expend time to subsequently investigate. (Gray and Debreceny, 2014)

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 for the criminal organization to use it effectively. Dealing with 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 are 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. (Manjunath, 2015)

Review the past work

Identifying various methods that can help money laundering detection is essential. Our goal here is to introduce different methods of money launderingdetection. Given the importance of the issues in the past, in this part, we introduce ways that are presented from 2005 and 2017 to detect money laundering. In recent years, the importance of detecting this fraudulent Activity has been more understood due to adverse and harmful effects to the country's economy, and during this time, various methods are provided for money laundering detection, that the most important of these researches have been identified in **Figure 1** with the time of publication.

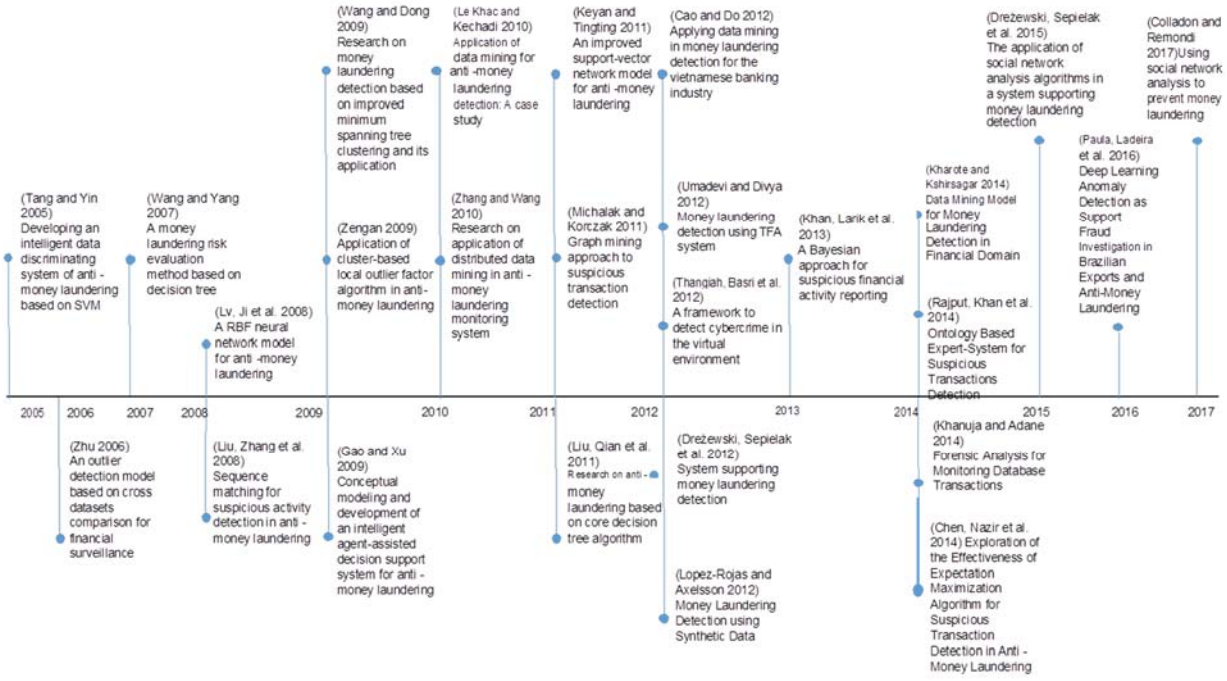

Figure 1. Published Articles in the Field of Money Laundering with the Year of Their Publishing

In this part, we will try to extend the research presented in the area of money laundering detection to the main areas of data mining and machine learning. We have presented a brief definition of the concept for each part, and then the research title, along with the type of methods, the main purpose and the algorithm, method or technology used in this study are expressed in the table provided for each part.

Clustering: Clustering is a process that classifies the data into different groups, and the members of each group have the most similarities to each other and the members of each group have the least similarities to another group. The best performance of a clustering algorithm will be apparent when the clusters are away from each other as far as possible (Han, Kamber, and Pei 2011). In anti money laundering, clustering is typically used for grouping transactions with bank accounts in different clusters that have the most similarities with each other. These techniques help us to detect patterns for suspicious transactionsequenceor present models to identify the accounts or the riskier customers. One of the most important challenges facing the clustering of financial transactions is the size and the amount of data, for example, we are facing thousands or millions of transactions per unit time in this method (Rohit and Patel 2015). Table 1 shows the clustering methods for the money laundering detection.

Rule-based methods: We can observe two approaches in data mining, classification - prediction and clustering approach (Han, Kamber, and Pei 2011). Rule-based methods are considered classification and prediction methods. In rule-based methods, we are facing a set of rules that are expressed in the language of logic, and actually, we use a series of logical rules to classify the factors (Pang-Ning, Steinbach, and Kumar 2005). In Table 2, we review the summary of the rule-based methods.

Table 1. Classification of Money Laundering Detection Methods based on Clustering

Source	Methods	Title	The main objective	Technology /algorithms/ methods
Zhu 2006	Clustering	An outlier detection model based on cross dataset comparison for financial surveillance	The money laundering detection using comparison every customer transaction with the same customer's transaction history	Creating a profile for customers
Zengan 2009	Clustering	Application of cluster-based local outlier factor algorithm in anti-money laundering	The money laundering detection using clustering techniques combination and Outliers	Cluster-based local outlier factor algorithm
Wang and Dong 2009	Clustering	Research on money laundering detection based on improved minimum spanning tree clustering and its application	Clustering method based on improved minimum spanning tree is created based on the criterion of dissimilarity, which the minimum spanning tree is built based on these criteria, and this tree is clustered into k clusters.	Minimum spanning tree
Le Khac and Kechadi 2010	Clustering	Application of data mining for anti-money laundering detection: A case study	Providing a solution based on the knowledge that detects the patterns of money laundering by combining data mining techniques and natural computing.	Neural networks Genetic algorithm Heuristics
Cao and Do 2013	Clustering	Research on Money Laundering Detection based on Improved MinimumSpanning Tree Clustering and Its Application	Clustering financial data to detect money laundering, according to the algorithm CLOPE	Algorithm CLOPE

Table 2. Rule-Based Methods in Money Laundering Detection

Source	Methods	Title	The main objective	Technology /algorithms/ methods
Khan et al 2013	Prediction	A Bayesian approach for suspicious financial activity reporting	It creates a model of the user's past activities, and this model will be a gauge of future customer activities. If the transaction or customer financial activities have significant deviations to the pattern, the new transaction of the user will be the suspicious money laundering activities	Bayesian networks
Rajput et al. 2014	Classification and prediction	Ontology Based Expert-System for Suspicious Transactions Detection	In this study, the provided approach to detect suspicious activities is founded by monitoring independent transactions.	Ontology / Semantic Web
Khanuja and Adane 2014	Classification	Forensic Analysis for Monitoring Database Transactions	Providing a methodology for continuous monitoring of the monetary and financial systems by preset instructions	Dempster – Shafer Theory
Panigrahi, Sural, and Majumdar 2009	Classification and prediction	Detection of intrusive activity in databases by combining multiple pieces of evidence and belief update	Providing an intrusion to the database detection method using four components, including rule-based component, the combine believes component, historical database sensitive to security events and Bayesian learning component.	Intrusion Detection

Neural Networks: the Neural network is a method that uses a set of connected nodes and mimics the human brain function. This method is based on computer models of biological neurons. A multilayer neural network contains a large number of units (neurons) linked together in a pattern of communication. First, the network is taught using a set of paired data to draw inputs and outputs. Then, the connection weight between established neurons and network will be used to determine the classification of experimental data sets (Han, Kamber, and Pei 2011).

A neural network consists of three layers of input, hidden and output. As shown in Figure 2, the input variables are represented by a series of a node. Each of input nodes (neurons) is connected to the neurons in the hidden layer by a communication link that each communication link has a weight, which represents the weight of communication link between these two neurons and ultimately, the output nodes indicate the classifications. (Bhattacharyya et al. 2011; West and Bhattacharya 2016).

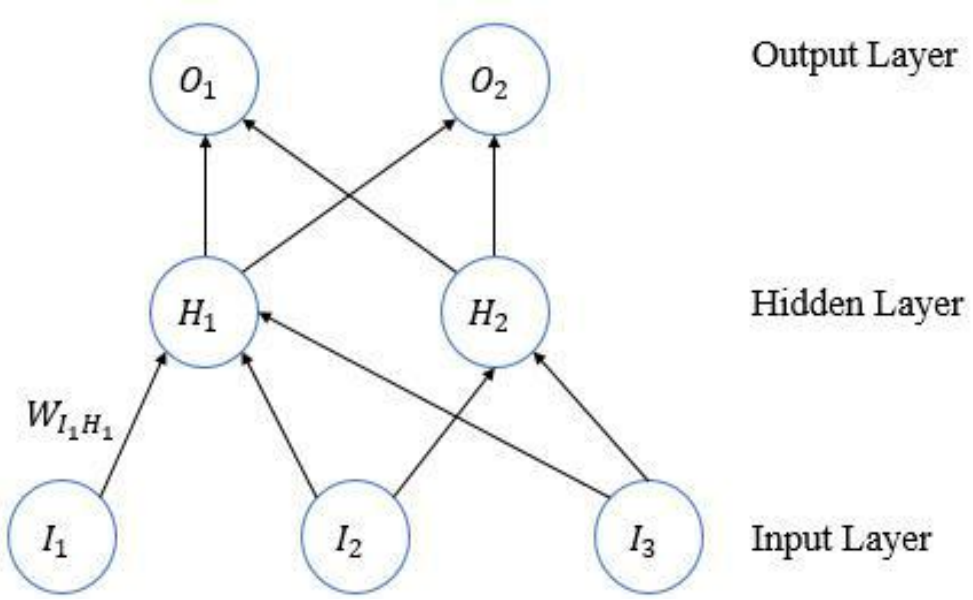

Figure 2: (West and Bhattacharya 2016)

The number of input nodes depends on the number and type of data sets characteristics, and the number of output nodes depends on the type of classification operation. The use of neural networks has many advantages, including the ability to deal with confusing data and usability at a time when there is very little knowledge about the issue (Han, Kamber, and Pei2011). Table 3 explains the raised research in the discussion of money laundering detection.

Table 3. The Use of RBF Neural Network for Money Laundering Detection

Source	Methods	Title	The main objective	Technology /algorithms/ methods
Lv, Ji, and Zhang 2008	Prediction	Classification	A way to detect money laundering is provided using neural networks which are composed of three layers.	Neural Network RBF[1]

1 Radial Basis Function

Support vector machines: SVM is a supervised learning method, which is used for classification. Support vector machines similar to neural networks can obtainapproximations with the desired degree of accuracy for each multivariable function. So, it is very useful to model the nonlinear and complex systems and processes, including detecting activities related to financial fraud. SVM goal is to find a separator super-vector of data points belonging to two classes, with a maximal margin. From a the geometric perspective, it is the gap between the super-vector and the nearest training samples. From another perspective, the margin is defined as the amount of space or separation between the two classes, which is defined by the super -vector (Bhattacharyya et al. 2011) and (West and Bhattacharya 2016). Figure 3 shows the display of optimized super-vector in SVM.

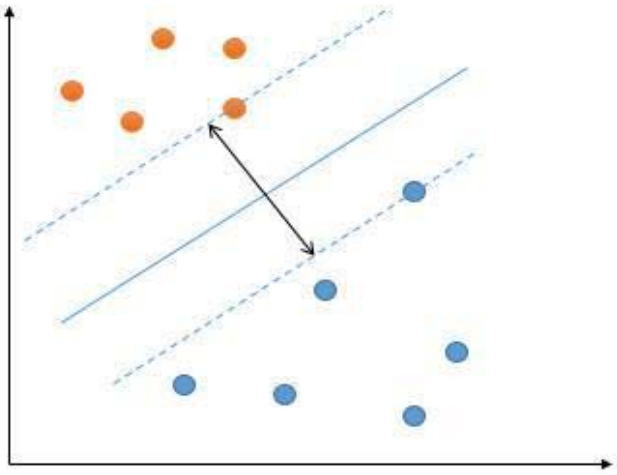

Figure 3. A Sample of SVM (West and Bhattacharya 2016)

Table 4. Classification of Methods Based on Support Vector Machines for Money Laundering Detection

Source	Methods	Title	The main objective	Technology /algorithms/ methods
Tang and Yin 2005	Classification	Developing an intelligent data discriminating system of anti-money laundering based on SVM	Using statistical learning theory to improve the anti money laundering. In this paper, a method is presented for detecting unusual activities using support vector machines.	Statistical learning theory Support Vector Machine
Keyan and Tingting 2011	Classification	An improved support-vector network model for anti-money laundering	Selecting a classification parameter suitable to detect suspicious activities using support vector machines is vital. To solve this problem, in this study, a cross-validation method is presented.	Support Vector Machine Cross Validation

Decision tree: The structure of a decision tree is a tree topology similar to a flowchart. The highest node in the tree is the root node, and the leaf nodes represent categories and distribution of categories. The decision tree is a classification or prediction technique that each non-leaf test node test specifies a feature and every branch out from this node shows the result of this test (Zhang and Zhou 2004). Figure 4 shows a sample of a decision tree. Unlike neural networks, decision trees deal with the production rules. The prediction obtained from a tree is explained in the form of a series rule in a decision tree, a whilethe result of prediction is only expressed in neural networks and how to achieve them is hidden in the network itself.Also, unlike neural networks, there is no requirement for the data to be necessarily numerical in the decision tree (Han, Kamber, and Pei 2011).

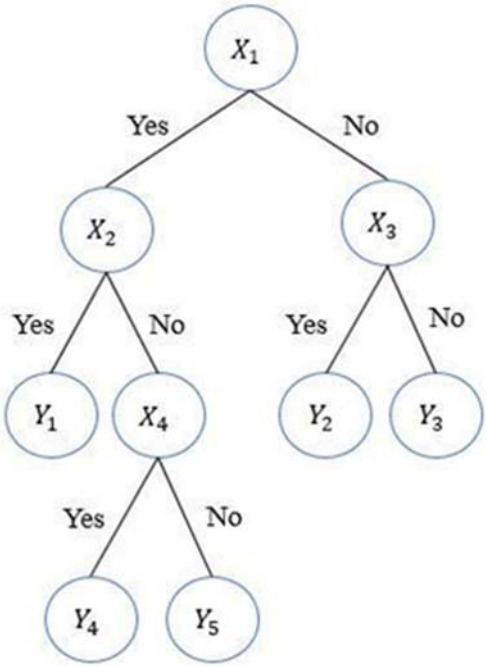

Figure 4. A Sample of a Binary Decision Tree With Input Values Xi and Yi Represents Label of The Classes To Classify The Input Values (West And Bhattacharya 2016)

Table 5. RepresentingThe Decision Tree-Based Detection of Money Laundering

Source	Methods	Title	The main objective	Technology /algorithms/ methods
Wang and Yang 2007	Prediction	A MONEY LAUNDERING RISK EVALUATION METHOD BASED ON DECISION TREE	Using a decision tree to determine the rules of money laundering risk through the accounts of customers	Decision tree
Liu et al. 2011	Prediction, Clustering	Research on Anti-Money Laundering Based on Core Decision Tree Algorithm	The combination of decision trees and clustering algorithms to detect money laundering	K-means algorithm, and BIRCH algorithm

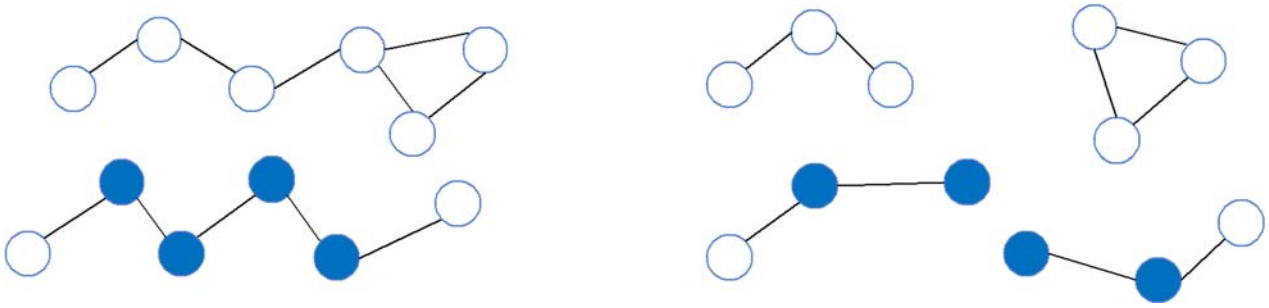

Figure 5. Networks with Different Relationships (Han, Kamber, AndPei 2011)

Decision forest or random forest is a collection of several decision trees that avoids instability and excessive risk education (bias) that may occur in a single tree. Another way that causes less complexity in trees is pruning trees. Pruning a decision tree leads to removing parts of the tree (below the tree) that do not participate in the accuracy of test samples (Bhattacharyya et al. 2011) and (West and Bhattacharya 2016). In Table 5, two examples of using a decision tree for anti money launderin ar shown.

Social networks: In recent years, the Social Network theory has attracted increasing attention. Social network analysis regarding data mining is called link analysis or link mining. For the modeling of social networks, the relationship between entities will be displayed in the form of links in a graph (Han, Kamber, and Pei 2011). A sample of social relations with different relations has been shown in Figure 5.

A social network represents relationships between social entities such as friends, professionals or writers. In the last decade, due to the increasing growth of communication technologies and Web-based services, the growth and penetration of these networks have been widespread, and many people had the experience and interacted with social networks. (Bindu, Thilagam, and Ahuja 2017) Through online social networks, a huge amount of facilities will be provided for interaction and cooperation between independent individuals, regardless of the geographical distance between them. Each network is a massive database of millions of users and their activities. Unfortunately, due to the liberalization of the use of most social networks, the information contained on these networks is a good place to delinquent users (Fire, Goldschmidt, and Elovici 2014). Therefore, network analysis will be instrumental to explore relationships or suspicious transactions for money laundering detection.

Social networks are dynamic. New links show a new inter action between objects.In the prediction of links, a snapshot of the social network at the time "t" will be placed at our disposal, and we are asked to predict the edges that will be added to this network, in the period t to t + 1. In this case, we are looking forward to use the real attributes of the model and to expose the development that can model the evolution of a social network (Han, Kamber, and Pei 2011).

Other Methods: In this part, methods will be introduced which specifically do not belong to a particular group of data mining methods and they are a combination of several methods, presenting a framework or introducing a new model to detect money laundering by means of statistics, methods, and procedures of data mining, machine learning, intelligent agents and other techniques, so it has been tried to collect these studies separately in a table. Table 7 shows these methods.

Table 6. The Methods of Money Laundering Detection Using Social Network Analysis

Source	Methods	Title	The main objective	Technology /algorithms/ methods
Dreżewski, and Filipkowski 2015	Social networks	The application of social network analysis algorithms in a system supporting money laundering detection	Presenting a new application for money laundering detection in a great deal of banking data. In this study, researchers have added a social network analysis component to their old system (Dreżewski, Sepielak, and Filipkowski 2012) to add a detailed analysis of human relationship to the money laundering detection system.	Money Laundering Detection System (MLDS)
Colladon and Remondi 2017	Social networks	Using social network analysis to prevent money laundering	Introducing a new approach to sorting and mapping relationships between data and presenting a prediction model based on social networks criteria to assess the risk of profiles that are involved in the business. The social network analysis is used to predict the involvement of accounts for customers who are involved in processes of money laundering.	Decision support systems

Table 7. Other Introduced Methods in Money Laundering Detection

Source	Methods	Title	The main objective	Technology /algorithms/ methods
Liu, Zhang, and Zeng 2008	Suspicious activity detection Suspicious activity reports	Sequence matching for suspicious activity detection in anti-money laundering	Presenting a sequence matching algorithm to detect money laundering detection, using sequential detection of suspicious transactions. This method takes advantage of the two references to identify suspicious transactions: history of each person's account and transaction information with other accounts	Time Series Euclidean distance
Gao and Xu 2009	Clustering Frequent patterns	Conceptual modeling and development of an intelligent agent-assisted decision support system for anti-money laundering	Presenting an analyzer system of transactions with the ability to detect money laundering. In this method, the analysis algorithms such as Transaction mining algorithm and Frequent mining algorithms were used.	Money transfer analysis

Zhang and Wang 2010	Distributed Data Anti-Money System Business Intelligence	Research on application of distributed data mining in anti-money laundering monitoring system	Presenting an architecture for real- time anti money laundering system based on an analysis of the transactions	Real Time Monitoring System
Michalak and Korczak 2011	Graph mining	Graph mining approach to suspicious transaction detection	Modeling the three main stages of money laundering (placement, layering, and integration) in the form of a graph and detecting the suspicious subgraphs.	Machine Learning /Fuzzy logic
Umadevi and Divya 2012	Clustering Frequent pattern mining visualization	Money laundering detection using TFA system	The plan presented in this study consists of 4 parts: entering banking transactions, then running frequent pattern mining algorithms and mining transactions to detect money laundering. Clustering transactions and suspicious activities to money laundering and finally display them on a chart.	K-means Sequence Miner algorithm
Thangiah, Basri, and Sulaiman 2012	Framework	A framework to detect cyber crime in the virtual environment	In this paper, approaches, tools, and techniques used for detecting money laundering are presented in virtual environments, and also a general framework for combating cyber crime is presented.	
Dreżewski, Sepielak, and Filipkowski 2012	Clustering Frequent pattern mining visualization	A system supporting money laundering detection	Monitoring banking transactions like a policeman, transactions are logged into the system, and analysis algorithms are run on it. Then, the results are visualized to be easier for analysts understand them.	FP-Growth FPClose FPMax Sequence Miner algorithm
MCA, PHIL, and PRABAKARAN 2014	Classification	Money laundering analysis based on time variant behavioral transaction patterns using data mining	Presenting an approach of money laundering detection using behavioral patterns. Behavioral patterns specify the methods of transfer between accounts, the amount range of moving money, the number of destination accounts and so n	Behavioral Pattern Time Variant
Paula et al. 2016	Deep Learning	Deep Learning Anomaly Detection as Support Fraud Investigation in Brazilian Exports and Anti-Money Laundering	In this study, a methodology known as a Cross Industry Standard Process for Data Mining (CRISP- DM) has been introduced that is composed of 6 Sections. The core of this method is Data Understanding that uses a pattern of deep learning for modeling and learning. Data in this methodology enters CRISP-DM cycle, and anomalies are detected based on model learning.	CRISP-DM DeepLearning Linear Principal Component Analysis(PCA)

Results

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, the 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.

According to what was mentioned in the previous parts, detecting activities related to money laundering is necessary and inevitable for the economy, industries, banks and financial institutions. The aim of this study was to review research conducted in the field of fraud detection with an emphasis on detecting money laundering and examine deficiencies based on data mining techniques. Due to the high volume of daily transactions in banks and financial institutions, the possibility of automated systems that can interact with the massive data, is essential. Methods of detecting money laundering in these systems are usually based on provisions, which include a set of predefined rules and threshold values . In addition to this approach, data mining techniques are very convenient to detect money laundering patterns and detect unusual behavior. Given that the instances of money laundering are changing and money launderers use newer methods, therefore, unsupervised data mining techniques will be more effective to detect new patterns of money laundering and can be crucial to enhance learning models based on classification methods. Of course, the development of new methods will be very useful to increase the accuracy of performance. Also,the systems that can draw the final reports in summarized and with graphics, increase the influencing factors and the simplicity of understanding the situation for analysts of financial fraud. It should also be noted that the research field of financial fraud detection, including the detection of money laundering, is very talented regarding theoretical research and application development.

References

1. Bhattacharyya, Siddhartha, Sanjeev Jha, Kurian Tharakunnel, and J Christopher Westland. 2011. 'Data mining for credit card fraud: A comparative study,' Decision Support Systems, 50: 602-13.

2. Bindu, PV, P Santhi Thilagam, and Deepesh Ahuja. 2017. 'Discovering suspicious behavior in multilayer social networks,' Computers in Human Behavior.

3. Cao, Dang Khoa, and Phuc Do. 2012. 'Applying data mining in money laundering detection for the Vietnamese banking industry.' in, Intelligent Information and Database Systems (Springer).

4. Colladon, Andrea Fronzetti, and Elisa Remondi. 2017. 'Using social network analysis to prevent money laundering,' Expert Systems with Applications, 67: 49-58.

5. Dreżewski, Rafał, Jan Sepielak, and Wojciech Filipkowski. 2012. 'System supporting money laundering detection,' Digital Investigation, 9: 8-21.

6. 2015. 'The application of social network analysis algorithms in a system supporting money laundering detection,' Information Sciences, 295: 18-32.

7. Efstathios Kirkos, Charalambos Spathis, Yannis Manolopoulos. 2007, "Data Mining techniques for the detection of fraudulent financial statements." Journal: Expert Systems with Applications, Volume 32, Issue 4, pp. 995–1003.

8. Fire, Michael, Roy Goldschmidt, and Yuval Elovici. 2014. 'Online social networks: threats and solutions,' IEEE Communications Surveys & Tutorials, 16: 2019-36.

9. Gao, Shijia, and Dongming Xu. 2009. 'Conceptual modeling and development of an intelligent agent-assisted decision support system for anti-money laundering,' Expert Systems with Applications, 36: 1493-504.

10. Glen L. Gray ., Roger S. Debreceny. (2014). A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. Journal: International Journal of Accounting Information Systems, Volume 15, Issue 4, Pp 357– 380.

11. Han, Jiawei, Micheline Kamber, and Jian Pei. 2011. Data mining: concepts and techniques: concepts and techniques (Elsevier).

12. “International trade and trading operations” (“Меѓународна трговија и трговско работење”). Prof. D-r Ljupcho Stojcheski, Assoc. D-r Ivana Stojchevska. Publisher: MCSP, Skopje 2019.

13. Keyan, Liu, and Yu Tingting. 2011. "An improved support-vector network model for anti-money laundering." In Management of e-Commerce and e-Government (ICMeCG), 2011 Fifth International Conference on, 193-96. IEEE.

14. Khan, Nida S, Asma S Larik, Quratulain Rajput, and Sajjad Haider. 2013. 'A Bayesian approach for suspicious financial activity reporting,' International Journal of Computers and Applications, 35.

15. Khanuja, Harmeet Kaur, and Dattatraya S Adane. 2014. 'Forensic Analysis for Monitoring Database Transactions.' in, Security in Computing and Communications (Springer).

16. Le Khac, Thien An, and M Kechadi. 2010. "Application of data mining for anti-money laundering detection: A case study." In Data Mining Workshops (ICDMW), 2010 IEEE International Conference on, 577-84. IEEE.

17. Michalak, Krzysztof, and Jerzy Korczak. 2011. "Graph mining approach to suspicious transaction detection." In Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on, 69-75. IEEE.

18. Umadevi, P, and E Divya. 2012. "Money laundering detection using TFA system." In Software Engineering and Mobile Application Modelling and Development (ICSEMA 2012), International Conference on, 1-8. IET.