Global Scientific JOURNALS

GSJ: Volume 9, Issue 2, February 2021, Online: ISSN 2320-9186
www.globalscientificjournal.com

MOBILE MONEY LAUNDERING USING DATA MINING METHODES: A REVIEW

Humaira Fatima, Dr.Hamid Ghous, Malik Mubashir Hussain

Humaira Fatima E-mail humaira6082@gmail.com doing m.phil in computer science
Humaira fatima masters degree program in computer science in University institution of southern Punjab mulatan, Pakistan ph# 03376082377 E-mail: humaira6082@gmail. Malik mubashir is masters degree in computer science in University, BZU, pakistan ph#03018630005 E-mail: mubashee@isp.edu.pk.

KeyWords

Mobile money laundering, data mining supervised unsupervised methods.

ABSTRACT

Mobile money laundering has one of the main challenges for the world. Mobile money laundering is a process of sending money through illegal ways using mobile applications. In past, many researchers have worked on detection of the transactions involved in mobile money laundering using data mining techniques. . In this review paper, we have analyzed the previous literature related to mobile money laundering using data mining methods. We have categorized the literature into supervised and unsupervised learning methods. After reviewing the literature, the limitations and future work is also suggested.

INTRODUCTION:

The process of mobile money laundering money is acquired through that even the investigated companies cannot find the important origin of prosperity. Mobile money laundering is method that used people for converting money that's acquired from illegaled ways. Mobile money laundering is illegald method that creating huge amounts of money bring about through criminal method. Mobile money laundering increases day by day, peoples used many process to acquire the mobile money laundering in different ways. Many researchers worked on the mobile money laundering to overcome this transaction and applied different methods to detect the mobile money laundering. Mobile money laundering is process that much utilize in Pakistan. This act exists treating with terrorism and terrorism financing through freezing mobile money laundering is the method to change illicit money into licit money. The financial action task force is an orginazation that helps the authorities to go after

money criminals and stop the funds. . Money launderers almost consist of three steps:

1) **Placement**: The first step is placement linked to installments the money in external banks.

2) **Layering**: The second step is layering associates to implement money through, money take away. And not show the important secret of money.

3) **Integration**: The third step is integration relevant to help the money for funding and other organization.

BACKGROUND:

As describe above about the mobile money laundering, this part also defines the mobile money laundering reviews. FATF is global companies that managing the mobile money laundering. Mobile money laundering is the largest problem in the world. Many researchers' worked on the money laundering and used different methods such as hybrid methods and different datasets also used but some gaps are also left to overcome the mobile money laundering. Mobile money laundering is converting through illegal way. Money laundering started from frontage as he sets up launder mates across the city in order to disguise the front of money earn from alcohol sale. Some researchers worked on the cash withdrawals.

DATA MINING:

Data mining is the process which contains the deep learning and machine learning methods. Different techniques are used to data mining to convert the raw data into actionable data. Data mining methods are divided in two categorize supervised and unsupervised.

SUPERVISED MACHINE LEARNING METHOD:

Supervised is the machine learning method that survey data from input to output way.

1. Regression
2. Classification
3. Naïve bayes model
4. Random forest model
5. Neural network
6. Linear regression
7. Decision tree
8. Support vector machine

UNSUPERVISED MACHINE LEARNING METHOD:

Unsupervised learning is a type of algorithm that read patterns from unlabeled data.

K-means for clustering problems

1. Clustering algorithm
2. KNN
3. Neural network
4. K-means
5. Fizzy algorithm
6. Hierarchal
7. C-means
8. Gaussian mixture

LITERATURE REVIEW:

Literature review has been divided in two categories supervised and unsupervised.

SUPERVISED METHOD:

Francis et.al (2020) proposed prediction for mobile money fraud. They utilized gradient boosted tree, support vector machine and naïve byes. These methods were applied on imbalanced dataset from KAGGLE that contains 9 attributes and consists of 99% negative class and 0.14% positive class .They used pre-processing techniques random, under sampling and over sampling. The Methods support vector machine and naïve byes performed better than gradient boosted tree with the accuracy of 64%, precision 93% and f-1 score 40%. In future this exploration can be stretched out by focus on real time data for detecting transaction in mobile money detection.

Nit in et.al (2020) utilized imbalanced dataset from KAGGLE. They worked with methods of SVM, decision tree, logistic regression, K-nearest and naïve byes .They used some preprocessing procedures hybrid, under sampling and hybrid sampling. The end result showed that RF, DT and logistic regression classifiers perform better than Naïve Bays and k- nearest neighbor. The random forest, decision tree and logistic regression better performance than Naïve Bays and k-nearest with the accuracy of 100%. In future work effectiveness of this innovative machine learning approach by considering scenario in other similar problem domain.

Martin et.al (2020) worked with methods of supervised machine learning naive bays, NN, SVM and neural network. They used some preprocessing techniques data sources data refinement training and test data. They used large dataset from largest Norway bank. The result showed that training and test data performed better than naïve bays random forest neural network and support vector machine with the accuracy of 95%. In future this research can be extended by how cash flow by the financial network around all system and party?

ALHANOUF et.al (2020) applied the imbalance dataset. They applied some preprocessing techniques is random sampling, over under sampling and cross validation. They utilized some methods support vector machine, decision tree, NN, multilayer preceptor, gradient boosted KNN. The possible outcome indicated that neural network method performed best with imbalance dataset in term of its accepter operating characteristics 76% and kappa statics 0.341%. In future work verify the reliability of the statistical approach.

HUYEN (2020) worked with some methods logistic regression, RF and SVM. They exercised synthetic financial dataset and elliptic dataset. They utilized some preprocessing technique data preprocessing, model building and model evaluation. The end result showed that random forest gave high accuracy 97.53% than logistic regression and support vector machine. In future work extend by look at the effectiveness of detecting money laundering cases through unsupervised machine learning.

Ana ISABLE CANHOTO (2020) applied the training dataset. They utilized some methods supervised machine learning, naïve Bayesian model, neural network, random forest and support vector machine. They made use of some preprocessing techniques Artifact features and system affordance. The end result showed that due to the lack of high quality large training dataset having money laundering methods have limited scope for using supervised machine learning method. In future work shows that system can add into the automated analysis system.

ABDALBASIT et.al (2020) utilized the large amount of diverse dataset. They utilized some preprocessing techniques link analysis and fraud detection. They utilized some methods support vector machine, decision tree, naïve bays classification, KNN neural network. The possible outcome showed that machine learning method will showed that this process dealing with a huge amount of data with a high level of accuracy and good quality.

FETHI et.al (2020) applied the large dataset. They utilized some preprocessing technique trading decision making and order execution management. They utilized some methods ANN and SVM. The possible outcome indicated that NN and SVM method ensuring that systematic coverage of all trading related actives from a practical perspective. In future work identifies the gap and opportunities for new expanding field.

JORGE IVAN et.al (2020) exercised withdrawal dataset, various dataset have 52512 records and 9 variables and the frequency of withdrawal dataset is 0.4286. They worked with method support vector machine, decision tree and prototype k. They made use of some preprocessing techniques link analysis, anomaly detection and risk scoring. Result showed that support vector machine trained the dataset to identify the outliers and frequency of support vector machine is 0.091852.

MARK ESHWER (2019) proposed data mining for statistical analysis of money laundering transactions. The statistical methods: multiple regressions, logistic regression, clustering multiple, clustering hierarchal, and clustering particle were applied on ….dataset. They made use of some pre- processing techniques data selection, data preparation, data discovery, data evolution and data reporting. The consequences indicated that clustering method is more efficient and useful to detect the suspicious transaction than logistic regression. In future work differentiate between legitimate and suspicious transaction.

Kang (2019) used the synthetic dataset from KAGGLE. The machine learning methods gradient boosted and random forest was used for classification of fraud and non-fraud. The outcome indicated that both models accomplish high accuracy but boosted tree performed better than random forest with the accuracy of 0.99%. In future work focus on the real life transaction to test the tree models and adopt their methodology like k-means and logistic regression..

Ratha pecth (2019) proposed fraud detection in mobile money by machine learning methods support vector machine, multilayer preceptor and naïve byes. They applied PAYSIM data set based on public PAYSIM dataset contains no missing values and contained seven column step, type, name, and amount. They made use of pre-processing Techniques such as selecting data, sampling data, collecting data, features selection, testing and validating. The outcome demonstrated that support vector machine performed better than multilayer preceptor and naïve byes with the accuracy of 95%. In future work needs to developed the more methods to overcome the detection problem.

JOSE et.al (2017) proposed the imbalanced dataset from international transaction. They utilized methods logistic regression, DT, NN and RF. They made use of some pre-processing techniques smote algorithm, oversampling, correlation, own elaboration, and variables distribution. The end result showed that random forest get better result 96% negative true and 98% positive true than decision tree neural network and logistic regression. In future work centre around focus on limited police investigation resources for companies.

Stefan et.al (2016) worked with the method of decision tree and clustering. They made use of some preprocessing techniques data preprocessing and multi agent based simulation. They proposed synthetic dataset. The outcome demonstrated that decision tree and cluster are more understandable and useful than machine learning algorithm. In future work build a model for the reproduction of mobile money transaction that improves the result of realistic synthetic dataset.

MARINA SOLIN (2010) exercised dataset from World Bank. They utilized some preprocessing techniques innovation, occurring and regulation. They worked with methods support vector machine, decision tree and gradient boosted. The possible outcome showed that mobile money laundering silence or registered with competent authorities. In future work order to keep regulation to effective for future.

ZENGANG MAO YE (2007) utilized transaction dataset. They utilized some methods regression, decision tree, case based reasoning and support vector machine. They utilized some preprocessing techniques legitimacy legal, illegitimacy, data collection and database reconstruction. The possible outcome indicated that the proposed key of the frame work lies in money laundering network analysis include link analysis for community generation and network destabilization. In future work focus on improving money laundering network structure analysis in addition to visualization by more by used of more unsupervised techniques.

Ahmad salehi et.al (2017) worked with the methods NN, SVM, decision tree and social network. They made use of some preprocessing techniques placement, layering, integration and clustering. They exercised historical and operational dataset. The outcome demonstrated that neural network support vector machine decision tree will be very useful to increase the accuracy of best performance than social network. In future work increase the influencing factor and understand the situation of financial fraud.

XINGI et.al (2009) utilized financial dataset from real word. They made use of some preprocessing techniques covering, clustering and proportion of anomaly points. They utilized some methods neural network, SVM and DT. In future work extended by what method to set the parameters k effectively for best improvement of the algorithm and integrate with other classical data.

SUNAN et.al (2007) used sample and uniform data. They utilized drug traffickling and summgling techniques. They used decision tree methodes. The end result showed that 12% out of thousand customers are considered as AML.

NHIEN et.al (2011) utilized the some preprocessing technique data quality, data volume, and heterogeneity data. They applied large dataset. They worked with methods clustering and support vector machine. The end result showed that clustering is important method that can efficiently applied for anti money laundering than support vector machine. The future work showed that need an efficient framework for integrating data mining techniques that can deal with different level for anti money laundering from transaction to multi organization.

HOSSIEN et.al (2018) applied the large dataset. They utilized some preprocessing techniques variable future selection, complexity and difficulty of data quality assurance. They utilized some preprocessing methods k-means clustering, SVM, Decision tree and neural network. The possible outcome indicated that k-means clustering performed 60% classification than neural network, decision tree and SPM. In future work new technologies trend in the era of big data can also continuously alter the research direction.

JUN TANG et.al (2005) made use with the preprocessing technique data reporting and data filtering. They utilized heterogeneous dataset from Wuhan branch of agriculture bank. They utilized some methods support vector machine and pattern recognition. The possible outcome showed that support vector machine is efficient for anti money laundering data reporting system and reconstruction than pattern recognition.

MARIA et.al (2014) practiced the synthetic dataset. They employed some preprocessing technique fraud chain and fraud detection. They utilized some methods logistic regression, decision tree, neural network, SVM. The possible outcome showed that transaction log containing approximately 460000 transactions so 10000 for end user and compare with classical fraud detection 99.81% precision and 90.81% recall. In future work to detect other type of fraud detection in mobile money laundering services such as agent fraud.

BORIS et.al (2002) applied the temporal dataset. They made use of some preprocessing technique data selection, forecast horizon and hypothesis evaluation. They employed some methods decision tree and neural network. The outcome indicated that neural network explaining money laundering techniques using DT and relational data mining methodology. Future work showed that developing practice software tool that make easy to use in data mining application.

XINWEI et.al (2009) proposed real world banking data. They utilized some preprocessing techniques optimal design pool learning stochastic approximation threshold hyper plane. They worked with some methods logistic distribution and support vector machine. They consequence indicated that support vector machine preformed the best efficiency and accuracy than the logistic distribution. Future work showed that generalizing the active learning criterion for the non linear threshold surface.

ALXENDER et.al (2016) exercised the heterogeneous dataset. They employed some preprocessing techniques clustering evaluation, numeric attributes, rules evaluation and generation cluster assessment report. They utilized some methods SVM, NN and clustering. The possible consequences indicated that we obtained confusion matrix with accuracy 99% and classification rate is 0.0683%.

YAN et.al (2018) practiced the actual transaction dataset by U.S financial institution. They implemented sampling schemes preprocessing technique. They utilized some methods bays decision tree, LR, RF, SVM and artificial neural network. The possible outcome indicated that optimal value of bays logistic regression is 1 decision tree value is 0.85, random forest is 5000, support vector machine is 0.5 and neural network is 32. In the upcoming work ANN can be conducted allow to deep learning of the data.

LIU et.al (2011) applied heterogeneous data set. They utilized preprocessing technique cross validation. They utilized method support vector machine. The possible outcome indicated that the accuracy of support vector machine is 79.42%.

GAO ZENGAN (2009) utilized synthetic dataset. They utilized some preprocessing technique data collection, data analyze, suspicious modeling and system modeling. They utilized clustering method. The possible outcome indicated that clustering threshold value Alpha 75% beta 4%and gamma 0.15%. In future work related subjective character of the accounts admin always opens to our future research.

Title	Author/year	Dataset	preprocessing	Methods	Results	Future work
"Fraud Detection in Mobile Money Transactions Using Machine learning"	"Kang, Haimeng Fall 2019"	Synthetic dataset from Kaggle.com dimensions 6362620 records	correlation	Random forest and gradient boosting	F1 score Precision 0.99487	Focus on to find the more life real transaction record and other methodologies..
"fraud de-	"Ratha pech"	Paysim	Collecting data sampling	Support vector	Accuracy preci-	Need to More

tection in mobile money transfer as binary classification problem"	13 June 2019"	Dataset based on public paysim dataset and contains no missing values and contains seven column step ,type name ,amount, old balance ,new balance etc.	data selecting features training model testing and validating	machine Multilayer preceptor on Naive bays	sion should be 95% recall and f-score.	advanced methods are to be developed to overcome the detection problem.
"Mobile Money Fraud Prediction A Cross-Case Analysis on the Efficiency of Support Vector Machines, Gradient Boosted Decision Trees, and Naïve Bayes Algorithms"	"Francis Effirim Botchey Zhen Qin Kwesi Hughes-Lartey 31 july 2020"	Imbalanced dataset from KAGLE generated from transactional log and contain 9 attributes. dataset consist of negative classes 99% and positive class 0.129%	Random under sampling over sampling	Gradient boosted tree Support vector machine naïve byes algorithm	Accuracy 64% Precision 93% to 100% Recall F-1 score 40% to 100%	Focused on real time data driven to detect the money laundering.
"Detecting money laundering transactions with machine learning"	"23 January 2020 Martin Jullum, Anders Løland and Ragnar Bang Huseby"	Large dataset from Norway largest bank	Data sources data refinement training and test data	Supervised learning method Naive Bayesian random forest Neural network.	Test and train data performed better than supervised machine learning with the transaction of 95%.	Future work showed that how the work of cash flow around the financial account. .
"Anti money laundering and counter terrorists financing threats posed by mobile money"	"james wishker Mark Eshwer lokana 23 January 2019"		Rapidity elusiveness Anonymity Lack of oversight Peer to peer transaction Mobile to mobile transaction	Support vector machine decision tree gradient boosted	Result shows that AML most effective and optimum efficiency mitigate the factors.	Find the probability of initial threats.
"using machine learning for financial fraud detection in the accounts of companies investigated for money laundering"	"jose A.Alvarez-Jareno Elena badal Valero 7 2017"	Unbalanced dataset from international transactions.	Smote algorithm oversampling sample description own elaboration variables distribution correlation	DT ,LR Random forest and neural network.	Random forest get better result 96% true negative and 98% true positive than logistic regression neural network and decision tree	Future work focus on investigate limited resources for companies.
"anti money laundering system a systematic literature review"	"ALHANOUF ABDURAHMAN SALEH ALSUWAILEM ABDUL KHADER JILANI SAUDGAR 25 MAY 2020"	Imbalanced dataset		Super vector machine decision tree neural network multilayer preceptor Gradient boosted KNN	Result showed that neural network performed best accuracy with imbalance dataset 0.76% .	Find the real ability of statistical approach.

"money laundering detection using synthetic data"	"Stefan Axelsson edgar alonoso 4 october 2016"	Synthetic dataset	data pre-processing mobile money simulation	decision tree clustering	Result showed that decision tree and cluster that are more understandable than machine learning algorithm.	focuses a model that improve the result of realistic synthetic dataset.
"machine learning in money laundering detection"	"HUYEN VU 2020"	Synthetic financial dataset elliptic dataset	Data preprocessing model building and model evaluation	Random forest logistic regression support vector machine	Result shows that the random forest gave high accuracy 97.53% than logistic regression and support vector machine.	Future work look at the effectiveness of detecting money laundering cases through unsupervised machine learning methods.
"leveraging machine learning in the global fight against money laundering and terrorism financing	"Ana ISABLE CANHOTO 2020"	Training dataset available for querying.	Artifact features system affordance	Supervised method Naïve Bayesian model random forest model neural network SPM	Result shows that deed to the Lack of best quality large training dataset not use the supervised machine learning method..	Future work shows that not all illicit system feed into the automated system.
"methodology for assessing money laundering and terrorist fencing risks"	" Marina Solin junary2010"	Dataset from world bank	Innovation occurring regulation techniques are applied.	Support vector machine Decision tree Gradient boosted	Result shows that mobile money laundering provides silence or registered with competent authorities.	In future order to keep regulation to effective for future.
"A frame work for data mining based anti money laundering based"	"ZENGAN MAO YE (2007)"	Transaction dataset	Legitimacy illegitimacy data collection database reconstruction data preprocessing	Decision tree regression case base reasoning Support vector machine	. Result shows that the describing money laundering criminals is build on money laundering network generation which include link analysis community generation and data destabilization.	
"predicting fraud in mobile money transfer using case based reasoning "	"21 November 2017" Adeyinka adedoyon KAPETANAKIS SAMAKOVITIES	Sample dataset	Case based reasoning Case similarity	SVM, DT gradient boosted KNN	Case based reasoning model Accuracy 97% Recall and precision 93% positive class 98% than kNN algorithm.	Future work reducing computation cast to improved the scale ability of CASE BASED reasoning

"data mining techniques for anti money laundering"	"Ahmad salehi mehdi ghanzanfari and Mohammad Fatian" "2017"	Historical dataset	Placement layering integration clustering	Neural network SPM decision tree social network	Result showed that neural network support vector machine and decision tree perform best accuracy than social network	increase the inflecting factor and simply understand the process of financial fraud.

"Developing an intelligent data discriminating system of anti money laundering based on SVM "	"JUN TANG JIYAN YIN 21 AUGUST 2005"	Heterogeneous dataset from agriculture bank.	Data reporting data filtering	Support vector machine pattern recognition	Result shows that support vector machine is efficient for Anti money laundering data reporting system and reconstruction than pattern recognition.	
"Revealing fraud chains in mobile money transfer"	"Maria zhdanova jurgen repp Roland Rieke 2014"	Synthetic dataset	Fraud chain fraud detection	Decision tree logistic regression neural network SPM	Result shows that transaction log containing approximately 460000 transactions for 10000 end users and classical fraud detection is 99.81% precision 90.18% recall.	To detect remaining types of fraud detection in mobile money laundering services such as servant fraud.
"A hybrid approach for detecting suspicious accounts in money laundering using data mining techniques"	"May 2016 Ch.suresh Dr.k.thammi reddy"	Synthetic Transactional data set from multiple bank	Placement integration layering	Hybrid approach method is used	Result show that accuracy of the anti money laundering describing the suspicious account is placed in the layering stage of data mining..	Developed a system which can identify the best relation between suspicious accounts those using concept like ontology.
"research on money laundering detection based on improved minimum spanning tree clustering and its application"	"XINGQI WANG GUANG DONG 2009"	FINANCIAL DATASET from real word	Covering cluster and proportion of anomaly points.	SPM neural network decision tree	Result shows that money laundering transaction detect experiments those apply on financial data from real world showed the effective performance..	In future work focus we will improve our algorithm into real word financial data warehouse system to assist financial constitute for

						money laundering.
”A money laundering risk evaluation method based on decision tree”	“SU NAN WANG JIAN GANG YANG 2007”	Sample data uniform ruled data	Drug trafficking smuggling bribing	Decision tree	12% out of 160 thousand customers are considered as AML.	
“A RBF neural network model for anti money laundering”	“LIN TAO LV NAJI JIU LONG ZHANG AUG 2008”	Financial transaction record dataset acquire from commercial bank.	Lest square algorithm recursive clustering	Neural network Support vector machine	Result shows that neural network proposed method have been the highest detection rate and the lowest false positive rate than SPM.	
“an investigation into data mining approaches for anti money laundering”	“NHIEN AN LE KHAC SAMMER MARKOS M.O NELLI 2011”	Large dataset	Data quality data volume and heterogeneity data	Clustering support vector machine	Result shows that clustering are important method that can efficiently applied for AML than support vector machine.	Future work shows that need an effective frame work for data mining techniques that can use with different levels of AML from transaction to multi organization.
”Data mining for financial applications”	“BORIS KOVALERCHUK EVGENII VITYAE 2002”	Temporal dataset	Data selection forecast horizon hypothesis evolution	Neural network decision tree	Result shows that neural network discovering money laundering techniques using decision rules and relational data mining methods..	Future work shows that developing practical software tool that make easy to operate in data mining application.

"Active learning through sequential design with applications to detection of money laundering"	"XINWEI DENG V.ROSHAN JOSEPH AGUS SUDJIANTO 2009"	Real word banking data		Logistic distribution Support vector machine	Result shows that the support vector machine performance shows the efficiency and accuracy than the logistic distribution is better.	Future work shows that generalizing the active learning criterion for the non linear threshold surface.
"client profiling for an anti money laundering system"	"ALXENDER Claudio Balsa joao 11 January 2016"	Heterogeneous dataset	Clustering evaluation numeric attributes rules evaluation and generation cluster assessment report	Support vector machine neural network clustering	Result shows that we obtained a confusion matrix with accuracy 99% and classification rate is 0.0683% .	
" MACHINE LEARNING AND SAMPLING SCHEME AN EMPIRICAL STUDY OF MONEY LAUNDERING DETECTION"	"YAN ZHANG PETER TURBY 12 OCTOBER 2018"	Actual Transaction data by U.S financial institution.	Sampling schemes	Bays logistic regression DT random forest SVM ANN.	Result showed that the optimal value of bayes logistic is 1 and decision tree value is 0.85 and random forest is 5000 support vector machine is 0.5 and neural network is 32.	Future work multilayer ANN can be conducted allow to deep learning of the data.
"sequence matching for suspicious activity detection in anti money laundering"	"XUAN LIU PENGZHU DAJUN ZENG 2008"	Real financial dataset from Chinese financial institution.	Normal transaction suspicious transaction	SVM ,DT	Result showed that by selecting different detecting feature we achieve highest sensitivity 0.988 and specify 0.442.	Influence of different features on suspicious activity detection.
"An improved support network model for anti money laundering"	"Liu Keyan Yu Tingting 2011"	Heterogeneous data set	Cross validation	Support vector machine	Result showed that the accuracy of support vector machine is 79.42%.	

”Machine learning applications in trading and financial Markets”	”FETHI A. RABHI NIKOLAY MEHANDJIEV ALI BAGHDADI 2020”	Large dataset	Trading decision making order execution management	SVM ANN	Result showed that support vector and neural network method ensuring systematic coverage of al trading related activities from a practical perspective.	Identifies the gap and opportunities for new expanding field.
”machine learning methodologies against money laundering in non banking correspondents”	”JORGE IVAN OLMER GARCIA BEDOYA OCTOBER 2020”	Withdrawal dataset various dataset have 52512 records and 9 variables and withdrawal frequency is 0.4286.	Link analysis anomaly detection risk scoring	SUPPORT VECTOR MACHINE decision tree k prototype	Result showed that support vector machine trains a dataset to identify the outliers and support vector machine frequency is 0.91825	

UNSUPERVISED METHOD:

Joana et.al (2020) practiced the bit coin transactional dataset from belli. Dataset consists of 203 and 769 transactions. Transactions 21% as licit and 2% illicit. They utilized supervised method logistic regression and unsupervised method anomaly detection. They made use of some pre-processing techniques random sampling and isolation forest. The consequence indicated that supervised method logistic regression performed better by using a just 5% label than unsupervised method. In future this examination can be expanded focus on the unsupervised illicit activity.

DR.G KARISHANPARIYA (2020) utilized transactional dataset. They employed some preprocessing techniques correlation, data cleaning, transactional log, generation relational, link weight computation and decision cluster. They worked with method regression and decision tree. The end result showed that regression performed 90% accuracy than decision tree. In future work can be extended that classifier the resultant data under cleaning and data mining process.

XUAN et.al (2020) proposed real financial data from commercial bank. They made use of some preprocessing techniques normal transaction and suspicious transaction. They employed methods transaction recognition, pattern recognition, time series recognition. They outcome indicated that sensitivity of scan statics is 0.516% and the specify of 0.949. Future work showed that increase the sensitivity of our scars algorithm.

JOSE DE et.al (2020) utilized some preprocessing techniques own elaboration. They utilized some methods unsupervised clustering self organization, strict competitive and neural gas. The outcome indicated that c- means algorithm based on the compactness and separation used for the clustering process than the strict competitive, organization map and neural gas. Future work shows that applying the model in European countries to analyze detecting real cases of terrorism financing.

FLORIS et.al (2020) applied the training and test dataset which is 80% training dataset and 20% test dataset. They utilized some preprocessing techniques own elaboration. They used some methods decision tree, CIT, RF and neural network. The outcome indicated that random forest performing better with the precision of 0.984% than DT and NN. In future work the model will be trained and tested on real data to

compare the performance on real life process.

.

IBRAR HUSSAIN et.al (2020) proposed the synthetic dataset by the PAYSIM simulator. The utilized some preprocessing techniques clustering is used to detect the financial fraud and statistically technique are used to detect fraud in different field. They utilized some methods decision tree, decision table, naïve byes k-nearest and neural network. The possible consequence that decision tree and naïve byes gives the lowest accuracy precision, recall, F-score than the neural network and k-nearest. In future work deep learning can be applied to identifying the evolving fraudulent patterns in mobile system.

Leno UUSKULA (2019) proposed imbalanced dataset from MONESE LTD and dataset split into training dataset and test dataset random 70% to train set and 30% test set . They made use of some preprocessing technique oversampling and under sampling. They utilized methods logistic regression, random forest and k means clustering. The outcome indicated that accuracy of logistic regression is 0.0431 and random forest accuracy is 0.1128.

Adeyinka et.al (2017) proposed predicting fraud in mobile money transfer using case based reasoning on dataset. They made use of some preprocessing techniques case based reasoning and case based similarity. They utilized KNN algorithms. The end result showed that case based reasoning model achieved accuracy 97%, recall and precision 93%. In future work will be spotlight on reducing computation cast to improve the scale ability of case based reasoning.

Ch.suresh et.al (2016) exercised the synthetic transactional dataset from multiple banks. They made use of preprocessing techniques placement, integration, and layering. They worked with hybrid approach method. The end result showed that efficiency of the money laundering describing that the suspicious account is placed in the layering stage of data mining. In future work focus on develop a system which identify the relation between suspicious accounts using concept like ontology.

RUI et.al (2011) made use of some methods clustering, k-means algorithm core and decision tree. They applied large dataset. They utilized some preprocessing method leaves connection mode. The result shows that core decision tree performed better than k-means clustering. In future work measure to wider the range of application.

BY JHOON et.al (2004) utilized the merchandise dataset from U.S. They exercised some preprocessing technique overvaluing and undervaluing. They possible outcome indicated that obvious information be sent to the U.S custom agency in 24 hours in advance the shipment from a foreign part. Future work showed that increasing the quality of intelligence information.

.

SU NAN et.al (2007) utilized method decision tree. They utilized sample dataset and uniform ruled data. The utilized some preprocessing techniques drug trafficking, smuggling, bribing. The end result showed that 12% cutomers out of 160 thousands considered for AML.

.

NHIEN et.al (2010) utilized transaction data from multinational bank. They made use of some methods neural network and k-means clustering. They utilized some preprocessing techniques customer identification and transaction analysis. The end result showed that k-means clustering performed 65% detection than neural network. In future work extended by improving the learning process to Handel the problem of very large dataset.

R.CORY et.al (2010) utilized the large dataset. They utilized some preprocessing techniques placement, integration, layering, detection and avoidance. They utilized some methods logistic regression, linear regression, cluster analysis and NN. The end result showed that use of new methodologies that could increase the peace enforcement ability to detect reduces and prevent money laundering activities.

NIHA et.al (2015) utilized the historical dataset. They made use of some preprocessing technique crime detection link, analysis financial, crime reporting system discovering and discrimination. They worked with some methods case based reasoning, decision tree, neighbor retrieval, artificial intelligence and neural network. The outcome demonstrated that data mining could potentially be used to lessen and even prevent crime for the forth coming year. In future work shows that increasing the effectiveness and efficiency of criminal and intelligence analysis.

LIN TAO et.al (2008) practiced the financial transaction dataset acquire from commercial bank. They worked with methods neural network and support vector machine. They made use of some preprocessing technique least square algorithm technology and clustering algorithm technique. The possible outcomes indicated that neural network proposed method have the high detection rate and the low false positive rate than support vector machine.

MAHESH et.al (2014) utilized the some preprocessing techniques data importer, data visualization, profile generation, suspected sequence sets. They utilized some preprocessing methods k-means clustering. They used dataset consist of no of objects used in the algorithm. The possible outcome showed that k-means clustering extract pattern from record and learn user decision pattern frequent pattern from association.

SERGIO et.al (2021) proposed the training and validation dataset gives similar mean square. They brought into play some preprocessing techniques own elaboration and non uniform distribution of data. They utilized methods k-means clustering, neural network and artificial intelligence. The possible outcome indicated that (Mean absolute error) in training dataset is 0.3317 and validation set 0.4059 and (Mean square error) in training data is 0.5577 and validation set 1.4602. In future work showed that the appropriate level of detail for the definition of preventive and detection measure in the different communes where the city is organized.

ZHONGEFIE et.al (2003) proposed data mining investigating money laundering crimes utilized the Bi-part dataset here we do not have explicit and direct access to the communicate information between data items. They utilized some preprocessing technique correlation analysis, community generation, local correlation, global correlation, identification and link hypothesis. They utilized some methods k-means algorithm, clustering algorithm and pattern recognition. The possible outcome indicated that collection of 332 documents in 20 minutes completed the model generation on a PIII-800 with 512 MB memory running windows 2000.

Victoria et.al (2014) applied the wine dataset. They utilized proximity based technique for the data distribution model. They utilized some methods support vector machine, neural network, KNN, linear regression and clustering. The possible outcome indicated that support vector machine covers minimal dataset which effectively covers the data distribution through a small subset and neural showed the classification accuracy and regression combines the multiple attribute into single attribute.

Title	Author/year	Dataset	preprocessing	Methods	Results	Future work
"machine learning methods to detect money laundering in the bit coin block chain in the in the presence of label scarcity"	"joana lorezn maria ines silva 29 May 2020"	Real bit coin transactional dataset taken from Bellei database consist of 203,769 transaction 21% are labeled and as licit and as 2%	Random sampling isolation forest	Supervised method Logistic regression Anomaly detection unsupervised	Result showed that supervised method logistic regression performed better by using a just few 5%labeld than unsupervised .	Focus the Unsupervised illicit activity

		illicit.				
"detecting money laundering with benford 's law and machine learning"	"Lenno UUSKULA 2019"	Imbalanced dataset from MONESE Ltd and dataset split into train and test dataset random 70% to train set and 30% test set.	Oversampling under sampling	Logistic regression random forest k-means clustering	Result shows that logistic regression accuracy is 0.0431 and random forest accuracy is 0.1128.	
"identifying suspicious money laundering transaction based on collaborative relational data screening model using decision classifier in transactional database"	"DR.G KARISHNAPARIYA 27-02-2020"	Transactional dataset	Correlation Data cleaning transactional log generation relational link weight computation decision from cluster	Regression decision tree	Result shows that regression perform 97% accuracy than decision classifier	Future work shows that classifier the resultant data under cleaning and data mining process.
"research on anti money laundering based on core decision tree algorithm"	"RUI LIU XIA LONG QIAN SHU MAO SHUAI ZHENG ZHU 2011"	Large dataset	Leaves connection modes.	Clustering k-means algorithm core decision tree	Core decision tree perform better than k-means clustering.	Future work measure to wider the range of application.
"detecting money laundering and terrorists financing via data mining"	"By JHON S.ZDANOWICZ 2004"	Merchandise trade database from U.S	Overvaluing undervaluation	KNN	Result showed that important information sent to the U.S agency in 24 hours advance to the shipment from other country port.	Increasing the quality of intelligence information.
"Towards a new data mining based approach for anti money laundering in an international investment bank"	" NHIEN AN LE KHAC SAMMER MARKOS MOHAND TAHAR KECHADI 2010"	Transaction dataset from multinational	Customer identification Transaction analysis	k-means clustering neural network	Result showed that k-means clustering 65% detection than neural network	Future work focus on improve the learning process to handle the problem of very huge dataset.
"Data mining for statistical analysis of money laundering transaction"	"Mark ESHWER LOKANAN 3 March 2019"	Training dataset	Understanding data Data selection data preparation data discovery data evolution data reporting	Multiple regression Logistic regression Clustering multiple clustering hierarchical clustering parti-	Result showed that statistical methods clustering very efficient and useful for detecting suspicious transac-	Future work differentiate between legitimate and suspicious transaction

				tion	tion than logistic regression.	
”Exploring data mining technologies as tool to investigate money laundering”	“R.CORY WATKINS K.MICHAEL REYNOLDS RON DEMARA 27 OCT 2010”	Large dataset	Placement integration layering detection avoidance	Logistic regression cluster logistic regression , ANN	Result showed that use new data mining methodologies that could increase the peace enforcement’s ability to detect reduce and prevent money laundering activities.	
“DIGITILAZTION AND BIG DATA MINING IN BANKING”	“HOSSEIN HASSANI XU HUANG EMMANUEL SILVA 20 JULY 2018”	Large dataset	Data pre-processing variable future selection complexity and difficulty of data quality assurance.	K-means clustering neural network.	Result shows that k-means clustering perform 60% classification than DT, NN and SVM.	Future work shows that new technologies in the area of big data can also change the direction of research.
“detection of anomalous transaction in mobile payment system”	”IBRAR HUSSAIN MUHAAMAD ASIF DECEMBER 2020”	Synthetic dataset generated by the paysim simulator.	Clustering technique is used to detect the financial fraud. Statistically based techniques are used to detect fraud in different field.	Decision tree decision table naïve Bays K-NEAREST neural network	Result shows that decision tree and naïve bays gives the low accuracy precision and recall F score than the neural network and k-nearest.	Future work shows that deep learning can be applied to identify the evolving fraudulent patterns in mobile system.
”money laundering and terrorism financing detection using neural networks and an abnormality indicato	“jose de jesus rocha Salazar maria jesus Segovia –vargas Maria del mar Camacho minano 5 december 2020”		Own elaboration	Unsupervised clustering strict competitive learning self organizing maps c-means neural gas.	Result shows that c-means algorithm based on the compactness and separation used for the clustering process than the harsh aggressive learning self organization and neural gas.	Future work shows that apply the model in Europe eastern countries to detect real cases of terrorism financing as it does in proxy cases.
“Data mining necessity for crime detection”	“NIHA MISHRA POOJA SHELKE FEBRARURY 2015”	Historical dataset	Crime detection Link analysis financial crime reporting system discovering discrimination	Case based reasoning decision tree nearest neighbor retriev-	Result shows that data mining could be possibly used	Future work shows that increasing the accuracy and

				al artificial intelligence neural network	to minimize the crime for the forth coming year.	efficiency of criminal and intelligence analysis.
"the role of machine learning in digital forensics"	"Abdalbasit mohammad qadir Asaf varol june 19 2020"	Large amount of diver's dataset.	Link analysis fraud detection	decision tree naïve Bays classification KNN neural network	Result shows that machine learning use this process with large amount of data with a highest level of accuracy and better quality result.	Future work to predict criminal in behavior.
"A scan statics based suspicious transaction detection model for anti money laundering in financial institute"	"XUAN Liu Penghzhu Zhang 2010 "	Real financial data from commercial bank	Normal transaction suspicious transaction	Transaction recognition pattern recognition time series recognition	Result shows that the sensitivity of scan statics is 0.516 and the specificity 0.949.	Future work shows that increase the sensitivity of SARSs algorithm.
"applied machine learning in social sciences neural network and crime prediction"	"SERGIO Luis Nanez Alonso Javier Jorge Vazquez 2021"	Training and validation dataset gives similar mean square error	Non uniform distribution of data own elaboration	k-means clustering neural network artificial intelligence	Result showed that mean absolute error 0.3317 in training data and 0.4095in validation data. Mean square error 0.5577 in training and 1.4602 in validation.	The appropriate level of detail for the definition of preventive and detection measure in the different communes where the city is organized.
"Applying data mining in investigating money laundering crimes"	"Zhongfei MARK ZHANG JHON J SALENRO PHILIP S.YU 2003"	Bi party dataset	Community generation correlation analysis link local correlation global correlation identification link hypothesis	k-means algorithm clustering algorithm pattern recognition	Result showed that collection of 332 documents for the complete the model generation on a PIII-800 with 512 MB running windows 2000.	
"application of cluster based local outlier local algorithm in anti money laundering."	"Gao zengan 2009"	Synthetic dataset	Data collection data analyze suspicious modeling system modeling	Clustering	Result showed that clustering threshold alpha 75% beta 4% and gamma 0.15%.	Future work showed that relative subjective character of the account admin remains open to our

						future research.
”A survey of outlier detection methodologies”	”VICTORIA J. HODE JIM AUSTIN 17 MAY 2014”	Wine data set	Proximity based technique for the data distribution model.	k-NN neural network linear regression clustering	Result showed that neural and support vector machine which covers the minimal dataset which most effectively covers the data distribution through small subset. And neural network showed the classification accuracy. Regression combines the multiple attributes into single attribute.	

LIMITATIONS:

The use of supervised and unsupervised methods will help to detect the money laundering from the huge datasets. Some limitations were also found in above literature review such as applying on real time datasets, lack of hybrid models and use of feature selection methods.

In future recommendation, we suggest of developing hybrid models based on feature selection methods to handle this huge transactional datasets. We recommend of using more real-time with deep learning methods. :

Conclusion

Mobile money laundering increasing with the passage of time, we need an automated system to detect the money laundering at the real-time. Money related transactional data is a huge dataset. Bundle of interest from previous researcher have been effective to deal with this huge amount of data using data mining methods. This review paper is structured based on supervised and unsupervised learning methods of data mining. The limitations show that there is a very limited work on real-time money laundering detection and lack of hybrid models. There is a need of hybrid models to predict money laundering at real-time.

References

1. Cao, D. and Do, P. (2012),"Applying data mining in money laundering detection for the Vietnamese banking industry", in Pan, J. S., Chen, S. M., Nguyen, N.T. (Eds), Intelligent Information and Database Systems. ACIIDS 2012. Lecture Notes in Computer Science, Vol. 7197, Springer, Berlin**,** Heidelberg.
2. Salehi, A., Ghazanfari, M. and Fathead, M. (2017),"Data mining techniques for anti money laundering", International Journal of Applied Engineering Research, Vol. 12No. 20, pp. 10084-10094.
3. Zdanowicz John S. Detecting Money Laundering and Terrorist Financing Via Data Mining [J] . Communications of the ACM, 2004, pp.53-55.
4. Alexander, C. and Balsa, J. (2015),"Client profiling for an anti-money laundering system", arXivpreprint arXiv:1510.00878
5. Demotes, D.S. (2018),"Fighting money laundering with technology: a case study of bank x in the UK",
6. Decision Support Systems, Vol. 105, pp. 96-107.Gao, Z. and Ye, M. (2007),"A framework for data mining-based anti-money laundering research",
7. Journal of Money Laundering Control, Vol. 10No. 2, pp. 170-179
8. Lopez-Rojas, E.A. and Axels son, S. (2012), Money Laundering Detection Using Synthetic Data, the 27th
9. Annual Workshop of the Swedish Artificial Intelligence Society (SAIS), 14-15 May 2012,
10. Linkoping University Electronic Press, Orebro, pp. 33-40 Savage, D. Wang, Q. Chou, P. Zhang, X. and Yu, X. (2016),
11. "Detection of money laundering groups using supervised learning in networks", arXiv preprint arXiv:1608.00708.
12. Schott, P.A. (2006),Reference Guide to anti-Money Laundering and Combating the Financing of Terrorism, TheWorldBank,Washington, DC.
13. The Norwegian Money Laundering Act, Chapter 3 (2009),"The Norwegian money laundering act,chapter 3", In Norwegian, available at: https://lovdata.no/dokument/NL/lov/2009-03-06-11#KAPITTEL_3(accessed 15 January 2018)
14. Gao, Z., & Ye, M. (2007). A framework for data mining- based anti-money laundering research. Journal of Money Laundering Control, 10(2), 170–179.
15. Goo , S., & Xu, D. (2010). Real-Time Exception Management Decision Model (RTEMDM): Applications in Intelligent Agent-Assisted Decision Support in Logistics and Anti-Money Laundering Domains.

16. Proceedings of the 43rd Hawaii International Conference on System Sciences, Honolulu, HI, USA, 2010, 1–10 Drezewski, R., Sepielak, J., & Filipkowski, W. (2015).
17. The application of social network analysis algorithms in a system supporting money laundering detection. Information Scienece, 295, 18–32.
18. Ehsan, U., Harrison, B., Chan, L. & Riled, M.O. (2018). Rationalization: a neural machine translation
19. approach to generating natural language explanations. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES'18), New Orleans, LA, USA, 2–3 February 2018, 81–87.
20. hang, Cheng-wei, and Yu-bo Wang. 2010. "Research on application of distributed data mining in anti-money laundering monitoring system." In 2010 2nd International Conference on Advanced Computer Controls.
21. Ravisankar, P., Ravi, V., Raghava Rao, G., Bose, I.: Detection of financial statement fraud and feature selection using data mining techniques. Decius. Support Syst.50. 491–500 (2011). Science Direct.
22. Bekirev, A.S., Klimov, V.V., Kuzin, M.V., Shchukin, B.A.: Payment card fraud detection using neural network committee and clustering. Optical Memory and Neural Networks (Information Optics) (2015).
23. Gao, Z. and Ye, M. (2007),"A framework for data mining-based anti-money laundering research", Journal of Money Laundering Control, Vol. 10No. 2, pp. 170-179.
24. Lopez-Rojas, E.A. and Axels son, S. (2012),Money Laundering Detection Using Synthetic Data, the 27th Annual Workshop of the Swedish Artificial Intelligence Society (SAIS), 14-15 May 2012, Link oping University Electronic Press. Orebro, pp. 33-40.
25. arandela, R., Valdovinos, R., Sanchez, J., & Ferri, F. (2004). The imbalanced training sample problem:Under or over sampling? InJoint IAPR international workshops on statistical techniques in pattern recognition (SPR) and structural and syntactic pattern recognition (SSPR)(pp. 806–814).
26. Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review.Statistical Science, 17,235–249.
27. Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines.ACM Transactions on Intelligent Systems and Technology (TIST), 2,27.
28. Cortes, C., & Vapnik, V. (1995). Support-vector networks.Machine Learning, 20,273–297.
29. Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function.Mathematics of Control.
30. Larose, D. T., & Larose, C. D. (2015).Data mining and predictive analytics. New York: Wiley.Le-Khac, N. A.,Markos, S., &Kechadi,M. T. (2010). Adatamining-based solution for detecting suspicious money laundering cases in an investment bank. InSecond international conference on advances in databases, knowledge ,and data application.
31. Choi, Dane and Lee, Kyungho. Machine Learning based Approach to Financial Fraud Detection Process in Mobile Payment System. IT Convergence Practices (INPRA), vol. (5) 4, 12-24, (2017).

32. Watkins, R.C., et al.: Exploring Data Mining technologies as Tool to Investigate Money Laundering. Journal of Policing Practice and Research: An International Journal 4(2), 163–178 (2003).
33. L. Lv, N. Ji, J. Zhang, "A RBF Neural Network Model for Anti-money Laundering," Proc. 2008 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR 2008), 2008, pp. 209-215.
34. Demetis, D.S. (2018),"Fighting money laundering with technology: a case study of bank x in the UK",Decision Support Systems, Vol. 105, pp. 96-107.
35. Lopez-Rojas, E.A. and Axelsson, S. (2012),Money Laundering Detection Using Synthetic Data, the 27th Annual Workshop of the Swedish Artificial Intelligence Society (SAIS), 14-15 May 2012,Link«opingUniversity Electronic Press Orebro, pp. 33-40.
36. Savage, D. Wang, Q. Chou, P. Zhang, X. and Yu, X. (2016),"Detection of money laundering groups using supervised learning in networks", arXiv preprint arXiv:1608.00708.
37. iyuan Chen, Amril Nazir, Ee Na Teoh, Ettikan Kandasamy Karupiah, et al.2014. Exploration of the effectiveness of expectation maximization algorithm for suspicious transaction detection in anti-money laundering. In 2014 IEEE.
38. European Union. 2018. Directive (EU) 2018/843 of the European Parliament and of the Council of 30 May 2018 amending Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, and amending Directives 2009/138/EC and 2013/36/EU.
39. The People's Bank of China, China anti-money laundering report 2005[R], China Financial Publishing House, Beijing, 2006.6. (In chinese) [9] Zhiyuan Chen, Le Dinh Van Khoa, Ee Na Teoh, Amril.