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.