Fighting Money Laundering with Statistics and Machine Learning

Mrs. B. Uma Maheswari[1], K. Bharathi[2]

[1]Assistant Professor, Dept. of MCA, Audisankara College of Engineering and Technology (AUTONOMOUS), Gudur, AP, India.

[2]PG Scholar, Dept. of MCA, Audisankara College of Engineering and Technology (AUTONOMOUS), Gudur, AP, India.

ABSTRACT

Money laundering is a profound global problem. Nonetheless, there is little scientific literature on statistical and machine learning methods for anti-money laundering. In this project, we focus on anti-money laundering in banks and provide an introduction and review of the literature. We project a unifying terminology with two central elements: (i) client risk profiling and (ii) suspicious behavior flagging. We find that client risk profiling is characterized by diagnostics, i.e., efforts to find and explain risk factors. On the other hand, suspicious behavior flagging is characterized by non-disclosed features and hand-crafted risk indices. Finally, we discuss directions for future research. One major challenge is the need for more public data sets. This may potentially be addressed by synthetic data generation. Other possible research directions include semi-supervised and deep learning, interpretability, and fairness of the results.

1. INTRODUCTION

OFFICIALS from the United Nations Office on Drugs and Crime estimate that money laundering amounts to 2.1-4% of the world economy [1]. The illicit financial flows help criminals avoid prosecution and undermine public trust in financial institutions [2]–[4]. Multiple intergovernmental and private organizations assert that modern statistical and machine learning methods hold great promise to improve anti-money laundering (AML) operations [5]–[9]. The hope, among other things, is to identify new types of money laun- dering and allow a better prioritization of AML resources. The scientific literature on statistical and machine learning methods for AML, however, remains relatively small and fragmented [10]–[12].

The international framework for AML is based on recommendations by the Financial Action Task Force (FATF) [13]. For banks, any interaction with proceeds from crime practically corresponds to money laundering within the framework (regardless of intent or transaction complexity) [14]. Further- more, the framework requires that banks:

1) know the identity of, and money laundering risk associated with, clients, and

2) monitor and report suspicious behavior.

Note that we, to reflect FATF’s recommendations, are intentionally vague about what constitutes ”suspicious” behavior.

To comply with the first requirement, banks ask their clients about identity records and banking habits. This is known as know-your-costumer (KYC) information and is used to construct risk profiles. The profiles are, in turn, often used to determine intervals for ongoing due diligence, i.e., checks on KYC information.

To comply with the second requirement, banks use electronic AML systems to raise alarms for human inquiry. Bank officers then dismiss or report the alarms to national financial intelligence units (i.e., authorities). The process is illustrated in figure 1. Traditional AML systems rely on predefined and fixed rules [15], [16]. Although the rules are formulated by experts, they are essentially ‘if-this-then-that’ statements; easy to interpret but inefficient. Indeed, over 98% of all AML alarms can be false positives [17]. Banks are not allowed to disclose information about alarms and generally receive little feedback on filled reports. Furthermore, money launderers may change their behavior in response to AML efforts. For instance, banks in the United States must, by law, report all currency transactions over $10,000 (regardless of whether they constitute money laundering or not) [18]. In response, money launderers may employ smurfing (i.e., split up large transactions). Finally, as money laundering has no direct victims, it can potentially go undetected for longer than other types of financial crime (e.g., credit card or wire fraud).

In this paper, we focus on anti-money laundering in banks and aim to do three things. First, we propose a unified terminology for AML in banks. Second, we review selected exemplary methods. Third, we present recent machine learning concepts that may improve AML.

2.LITERATURE SURVEY

Canhoto [18] and Weber et al. [20], [21] stated that deep learning and machine learning beats the traditional methods of anti-money laundering. Particularly, Weber et al. [21] high- lighted the significance of ML regulations and provided the Elliptic dataset for detecting illegal Bitcoin transactions. Dif- ferent machine learning techniques were used to evaluate the Elliptic dataset, including logistic regression (LR), multilayer perceptrons (MLP), random forest (RF), and graph convolu- tional networks (GCNs). It was observed that RF technique achieved the high results with a precision, recall-store and F1-score of 0.95, 0.67, and 0.788, respectively. To classify and detect suspicious currency on the Bitcoin network, Lee et al.

[22] implemented the artificial neural network (ANN) and RF algorithms. The illegal and legal Bitcoin data were collected from various websites such as Blockchain Explorer and Silk Road. The F1-scores showed that the RF algorithm achieved a high rate of 0.98, while the ANN algorithm achieved a lower rate of 0.89. In the same regard, a novel method for predicting illegal currencies in the Bitcoin currency is proposed by Alarab et al. [23] using a graph convolutional neural network (GCN). The MLP and GCN were combined to enhance the model’s performance for which a 0.974 of accuracy was achieved under the project method. However, The same author [24] used RF, Extra Trees, Gradient Boosting, XGBoost, LR, and MLP, where RF outperformed with a rate of 0.82. Along similar lines, Ostapowicz and Zbikowski [25] implemented different algorithms on the Ethereum network to identify fraudulent accounts based on supervised learning approach.

The accounts were classified and analyzed as “not fraudulen” or “fraudulent” using SVM, XGBoost, and RF. It was observed that the RF algorithm achieved the best results with a detection precision of 85.71. In another study, eight different supervised machine learning techniques were presented and analyzed by Bhowmik et al. [26] to investigate illegal transactions on the blockchain network. These include Naive Bayes (NB), LR, MLP, SVM, RF, Ada Boost, etc. The results of the comparison study found that among the five algorithms, SVM, RF, and NB algorithms obtained the best results with an accuracy of 97%. In view of the same, Monamo et al. [27] also employed an unsupervised learning method based on trimmed k-means and a k-means in order to track down illegal behavior and detect fraudulent activity on the Bitcoin transactions. To classify these transactions, Monamo et al. [28] applied clustering algorithms and machine learning techniques in which several assumptions were imposed to categorize transactions into illegal and legal categories. In addition, different Bitcoin fraud activities were illustrated from both global and local perspectives by using kd- trees and trimmed k-means. To further investigate these two methods, three classification algorithms were used including the maximum likelihood-based, random forests, and boosted binary regression. Based on the obtained results, it was found that the random forest outperformed the other two classi- fication models. Related to the detection and classification of suspected Bitcoin network addresses, several studies have been reported in literature based on different approaches and techniques [13], [29]–[31]. In fact, the unsupervised models for detecting money laundering activities were found to be inadequate for the Bitcoin network as per Lorenz et al. [13]. Therefore, they have developed supervised learning models to identify illegal money laundering activities in the network. In their study, a rule-based technique was employed that showed low detection rates and high false-positive rates. By Lin et al. [29], suspected Bitcoin network addresses transactions were detected and classified by adding the distribution data of transactions, detailed transaction summaries, and time series as new statistics. The model performance was improved and the variance in data was increased. In this study, various machine learning techniques, including LR, SVM, AdaBoost, XGBoost, and LightGBM were implemented. However, Light- GBM achieves the best results as compared to the other techniques. A novel method based on a cascade of classifiers and entity characterization to assail bitcoin anonymity was proposed by Zola et al. [30]. In this study, three different algorithms, including the gradient boosting, random forest, and Adaboost, were used to identify illicit transactions on the Bitcoin blockchain network. The inter-entity transactions (organizations or people with multiple accounts) were also in- vestigated, and the classification performance was improved by utilizing 34 features. Bartoletti et al. [31] used data mining and machine learning-based approaches to detect Ponzi schemes related to the Bitcoin addresses. In their study, three machine learning algorithms were provided for evaluation including the Bayes network, random forest, and RIPPER. As a result, the random forest has been proven to detect 96% of addresses. However, it is worth mentioning that the project approach was tested against Ponzi schemes.

Kumar et al. [32] classified a 10000-transaction dataset to identify money laundering activities using Naive Bayes algorthoms. The obtained results showed that the proposed model achieved 81% accuracies. In another study, the light gradient boosting machine (LGBM) is proposed by Aziz et al. [33] to detect fraudulent transactions. The MLP, RF, and KNN were compared with the LGBM approach for the identification and classification of fraudulent Ethereum datasets. Relative to the other techniques, the LGBM algorithm has achieved the highest accuracy of 99.03.

Based on the above discussion related to existing literature, it is evident that machine learning algorithms play a vital role in the detection of suspicious transactions in money laundering activities. However, it is worth mentioning that there are still several problems and challenges associated with the detection process that require further improvements. In addition, it seems that there exist very few studies on using deep learning ap- proaches to detect money laundering activities. In view of the same, this paper mainly aims at using deep learning methods with machine learning to detect such suspicious activities in Cryptocurrency.

3. PROPOSED WORK

The money laundering transaction detection model includes five main stages i.e. data understanding, data preprocessing, data splitting, model training, model testing, and model eval- uation. Fig. 1 illustrates the methodological framework of the study. Several ML and DL algorithms are employed in this chapter for transaction classification e.g. NB, RF Classifier, KNN Classifier, and DNN.

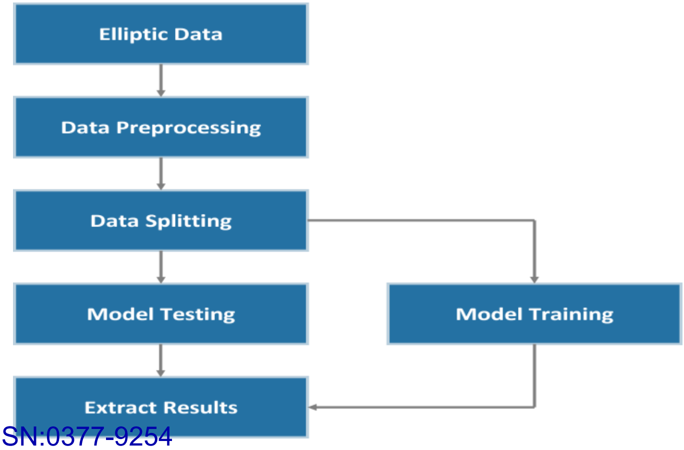

Fig. 1. System Model

A.Dataset

In this study, the Ellipse dataset1 created by Weber et al. [21] is employed to detect the cryptocurrency activities. Elliptic is a cryptocurrency monitoring company aimed to protect cryptocurrencies from illegal activity, and it has the largest publicly available dataset for transactions in cryptocurrencies.

B.Preprocessing of Data

As the model performance can be affected by irrelevant features, it is indeed necessary to detect and select the important features. Particularly, there are 166 features associated with each transaction in the elliptic dataset. Due to intellectual property rights, the elliptic company has not disclosed the details and nature of the features.

C.Training and Testing

In this subsection, the model training and testing for trans- action classification are discussed. Consequently, the trans- actions in Elliptic dataset will be classified into legal and illegal transactions. Particularly, the techniques employed in this study are based on supervised learning, which cannot be used when transactions have unknown labels. Therefore, such labels are omitted and not included in the training and testing phases as previously discussed. Essentially, the training set is utilized for model training and hyperparameter tuning. On the other hand, the testing set is utilized to evaluate the performance of the trained model. In the Elliptic dataset, there exist 46, 564 transactions which includes both legal and illegal transactions.

4. CONCLUSION

Money laundering represents a serious threat to govern- ments all over the world and it has been indeed challeng- ing. Various ML and DL techniques have been employed in literature to detect illegal transactions. However, there is still a serious need to further explore and develop suitable algorithms for detecting money-laundering activities, which was the main purpose of the study. Essentially, this research aims to determine the appropriate DL and ML algorithms for detecting money laundering using Elliptic BTC Dataset. To achieve this objective, the results of four algorithms are extensively analyzed and compared. These algorithms include three ML algorithms (RF, KNN, NB), and one DL (DNN). In addition, four key evaluation metrics were used to quantify the performance. These metrics include the precision, recall, F1-score, and ROC curve. the ML technique (RF) proved to be better at classifying fraudulent activities than DL. It was observed from the obtained results that the RF algorithm achieved the best results as compared to other algorithms.

5. REFERENCES

[1] U. W. Chohan, "The fatf in the global financial architecture: challenges and implications," 2019.

[2] W. Firmansyah and H. T. Atmadja, "Juridical analysis awareness of profession advocacy to financial transaction reports and analysis centre (ppatk) during prevent and eradicate money laundering crime," Journal of Multidisciplinary Academic, vol. 5, no. 4, pp. 308–314, 2021.

[3] R. Soltani, U. T. Nguyen, Y. Yang, M. Faghani, A. Yagoub, and

A. An, "A new algorithm for money laundering detection based on structural similarity," in 2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE, 2016, pp. 1–7.

[4] A. Salehi, M. Ghazanfari, and M. Fathian, "Data mining techniques for anti money laundering," International Journal of Applied Engineering Research, vol. 12, no. 20, pp. 10 084–10 094, 2017.

[5] C. Alexandre and J. Balsa, "A multiagent based approach to money laundering detection and prevention." in ICAART (1), 2015, pp. 230– 235.

[6] D. Savage, Q. Wang, X. Zhang, P. Chou, and X. Yu, "Detection of money laundering groups: Supervised learning on small networks," in Workshops at the Thirty-First AAAI Conference on artificial intelligence, 2017.

[7] G. Sobreira Leite, A. Bessa Albuquerque, and P. Rogerio Pinheiro, "Application of technological solutions in the fight against money laundering—a systematic literature review," Applied Sciences, vol. 9, no. 22, p. 4800, 2019.

[8] S. N. F. S. M. Nazri, S. Zolkaflil, and N. Omar, "Mitigating financial leakages through effective money laundering investigation," Managerial Auditing Journal, 2019.

[9] "Financial Crimes Enforcement Network. 2019. Application of FinCEN's Regulations to Certain Business Models Involving Convertible Virtual Currencies — FinCEN.gov, howpublished

= https://www.fincen.gov/resources/statutes-regulations/guidance/ application-fincens-regulations-certain-business-models ."

[10] S. A. Bieler, "Peeking into the house of cards: Money laundering, luxury real estate, and the necessity of data verification for the corporate transparency act's beneficial ownership registry," Fordham J. Corp. & Fin. L., vol. 27, p. 193, 2022.

[11] S. Butler, "Criminal use of cryptocurrencies: a great new threat or is cash still

king?” Journal of Cyber Policy, vol. 4, no. 3, pp. 326–345, 2019.

[12] M. Campbell-Verduyn, “Bitcoin, crypto-coins, and global anti-money laundering governance,” Crime, Law and Social Change, vol. 69, no. 2, pp. 283–305, 2018.

[13] J. Lorenz, M. I. Silva, D. Apar´ıcio, J. T. Ascensa˜o, and P. Bizarro, “Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity,” in Proceedings of the First ACM International Conference on AI in Finance, 2020, pp. 1–8.

[14] D. S. Demetis, “Fighting money laundering with technology: A case study of bank x in the uk,” Decision Support Systems, vol. 105, pp. 96–107, 2018.

[15] J. McPherson, K.-L. Ma, P. Krystosk, T. Bartoletti, and M. Christensen, “Portvis: a tool for port-based detection of security events,” in Proceed- ings of the 2004 ACM workshop on Visualization and data mining for computer security, 2004, pp. 73–81.

[16] Z. Chen, L. D. Van Khoa, E. N. Teoh, A. Nazir, E. K. Karuppiah, and

K. S. Lam, “Machine learning techniques for anti-money laundering (aml) solutions in suspicious transaction detection: a review,” Knowl- edge and Information Systems, vol. 57, no. 2, pp. 245–285, 2018.

[17] P. Tertychnyi, M. Godgildieva, M. Dumas, and M. Ollikainen, “Time- aware and interpretable predictive monitoring system for anti-money laundering,” Machine Learning with Applications, vol. 8, p. 100306, 2022.

[18] A. I. Canhoto, “Leveraging machine learning in the global fight against money laundering and terrorism financing: An affordances perspective,” Journal of business research, vol. 131, pp. 441–452, 2021.

[19] C. Yan, C. Zhang, Z. Lu, Z. Wang, Y. Liu, and B. Liu, “Blockchain abnormal behavior awareness methods: a survey,” Cybersecurity, vol. 5, no. 1, pp. 1–27, 2022.

[20] M. Weber, J. Chen, T. Suzumura, A. Pareja, T. Ma, H. Kanezashi,

T. Kaler, C. E. Leiserson, and T. B. Schardl, “Scalable graph learning for anti-money laundering: A first look,” arXiv preprint arXiv:1812.00076, 2018.

[21] M. Weber, G. Domeniconi, J. Chen, D. K. I. Weidele, C. Bellei,

T. Robinson, and C. E. Leiserson, “Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial foren- sics,” arXiv preprint arXiv:1908.02591, 2019.

[22] C. Lee, S. Maharjan, K. Ko, and J. W.-K. Hong, “Toward detect- ing illegal transactions on bitcoin using machine-learning methods,”in International Conference on Blockchain and Trustworthy Systems. Springer, 2019, pp. 520–533.

[23] I. Alarab, S. Prakoonwit, and M. I. Nacer, “Competence of graph con- volutional networks for anti-money laundering in bitcoin blockchain,” in Proceedings of the 2020 5th International Conference on Machine Learning Technologies, 2020, pp. 23–27.

[24] I. Alarab and S. Prakoonwit, “Effect of data resampling on feature importance in imbalanced blockchain data: Comparison studies of resampling techniques,” Data Science and Management, 2022.

[25] M. Ostapowicz and K. Z˙ bikowski, “Detecting fraudulent accounts on blockchain: a supervised approach,” in International Conference on Web Information Systems Engineering. Springer, 2020, pp. 18–31.

[26] M. Bhowmik, T. S. S. Chandana, and B. Rudra, “Comparative study of machine learning algorithms for fraud detection in blockchain,” in 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2021, pp. 539–541.

AUTHOR’S PROFILE

Mrs. B. UMA MAHESWARI currently working as Assistant Professor in Audisankara College of Engineering and Technology (AUTONOMOUS), NH-5, Bypass Road, Gudur, Tirupati (Dt.), Andhra Pradesh, India.

K. BHARATHI is pursuing MCA from Audisankara College of Engineering and Technology (AUTONOMOUS), NH-5, Bypass Road, Gudur, Tirupati (Dt.), Andhra Pradesh, India.