MACHINE LEARNING IN DETECTING MONEY LAUNDERING ACTIVITIES: INVESTIGATING THE USE OF MACHINE LEARNING ALGORITHMS IN IDENTIFYING AND PREVENTING MONEY LAUNDERING SCHEMES

Teja Reddy Gatla

Sr. Data Scientist, Associate Director of AI Operations, Department of Information Technology

ABSTRACT—Money laundering is becoming a huge problem in most countries.A number of criminals who manage to circumvent the law and justify their irregular income in a very smart way. The ever-evolving process of money laundering makes it problematic for the law enforcement to detect and prevent these criminal activities. Thus, the effort to curb money laundering should involve computer specialists in both research and the industry as a whole [2]. Internet's development and e-commerce have enabled AML efforts to embrace new opportunities and to gain greater strength. These days, many financial deals are being carried out electronically and generate digital data while being processed, adding to the high potential of using automatic processing to detect money laundering with interest and challenge. This is the reason for using AI and machine learning technologies in developing an anti-money laundering detection system. The purpose of this paper is to conduct a review and studies on the success and limitations of the machine learning technology in detecting money laundering [1]. Machine learning is a revolutionary part in technology development that provides a system with an ability to recognize a pattern from the past information and choose a positive course of action after they were analyzed and predicted. Because machine learning is capable of analyzing a large amount of data in a short time and it will give an automatic decision. In comparison with a traditional method using a rule-based expert system, an expert system is unable to handle a new data pattern that is not provided in a programmed rule [2]. An expert system will detect that new pattern as an anomaly (normal data pattern that is observed as an anomaly), and in AML efforts, there is a likelihood that a money launderer will change their method to adapt to a situation from time to time. So machine learning is relevant to use to analyze and detect pattern data in money laundering.

Keywords— Money laundering, AML techniques, machine learning technology, automation, Banks, artificial intelligence, algorithms, Finance, globalization, intelligent, technology, computer systems, AI techniques

I. INTRODUCTION

Money laundering activities represent serious problems for governments and financial institutions worldwide. Detection of money laundering transactions is an important and challenging task for all financial institutions [3]. Existing AML (Anti-Money Laundering) systems, such as the detection mechanisms used by banks, require a large amount of human resources as well as computational resources for developing a robust detection system to detect suspicious money transferring activities[3]. To minimize slipping through the cracks of known money laundering activities in large-scale financial systems and to reduce false positives, newer and more sophisticated techniques have to be developed.

One promising approach in detecting money laundering activity is the use of data mining and machine learning techniques. This research project is focused on developing a system to detect money laundering activities using unsupervised learning techniques in clustering analysis and implementing it as a relatively simple detection-based framework for the purpose of learning and understanding as to what constitutes money laundering behavior and what the differences are in transferring money for legitimate purposes[4]. By focusing on a relatively simple generic method of money laundering and on methods to refine detection to reduce false positive results, we aim to provide a new perspective and methods that may be further developed and applied to new detection systems in the future. Money laundering is a complex issue that requires constant innovation to stay ahead of the tactics employed by criminals. Our goal is to contribute to the ongoing efforts to combat money laundering and protect the integrity of financial systems.

Smurfing is a term used to refer to a laundering method where, for the purpose of evading laws or taxes, money is broken into many different, changing bank accounts. A common tactic used for smurfing is to use false identification when opening bank accounts. The system will use a clustering method to find the best characteristics of a name which would identify it as presumably false. The same would also apply to business accounts. After finding these characteristics, the system will search the general pool of new and changing account details given by the government for suspects of smurfing activity using the best business account identifier. If the activity becomes a success and time allows, it will then track these names to see if they lead to higher than average aggregate or business accounting change transactions[4]. The following chapter will describe a comprehensive data mining system which was designed to find an optimal balance between true positive rate and return on investment while reducing the amount of labor and time wasted on some of the less complex and fruitful alerts. The system uses a wide array of machine learning algorithms and concepts to include random forests, decision trees, a Naïve Bayesian classifier, and outlier detection to name a few. Although the system used the aforementioned classifiers to decide which transaction alerts were most likely to be true positives, this paper will focus on two particular techniques of interest: a utility algorithm and a variable importance analysis using a random forest.

High-tech machine learning algorithms, including deep neural networks and support vector machines, contribute to the most precise detection system's development. Such techniques will allow us to analyze massive amounts of financial data transactions, look for trends, and detect possible money laundering activities more precisely. The incorporation of

various data sources not only include financial transactions but the customer profiles and the external risks factors as well provides a tangible picture of the financial landscape and uncover hidden patterns that point to money laundering. Moreover, we are studying the application of natural language processing algorithms to explore textual information attached to the financial transactions [4]. We are going to do this in order to automate the extraction of data from unstructured data streams (like transaction descriptions and account notes) and integrate the relevant information into our detection models. To guarantee the effectiveness of our system, we will continuously improve and update our algorithms through the utilization of actual data. Dynamic adjustments to the varied ways of money-laundering will keep us on top of criminals and minimize catching innocent in our nets. These advanced techniques based on big data analytics, (including machine learning algorithms and natural language processing) will grow the powers of the current AML systems and do a lot of good for the global anti-financial crimes movement. In this collaboration, we can build a more secure and safe financial vibe for all of us.

Whilst the globalization of financial operations and the use of digital financial platforms has given rise to the money laundering issue, its threat is more apparent than ever. Money laundering technologies keep evolving and changing fast, developing according to technological progress, so it is very important that financial institutions and governments keep up with the newest verification methods.[4] The main focus of this research is on the benefit of the combination of different tools and methods to fight money laundering successfully. By adopting a cutting edge approach to detection, which entails such strategies as network analysis and anomaly identification, we can detect complex money laundering networks and irregularities that may point to illicit activities. Moreover, partnership and information exchange are vital to the mutual elimination of the laundered money problem. The system which we designed will give an opportunity for financial institutions and regulatory agencies to communicate and share the valuable information so that they may strengthen their combined efforts. Consequently, this will improve decision-making processes for investigators and make them more knowledgeable with workable intelligence [5]. The evaluation and subsequent fine-tuning of our system will be an ongoing undertaking. An important feedback comes from the industry experts, the law enforcement agencies, and other stakeholders to ensure that the detection techniques we use still remain relevant and effective to the changing landscape of financial crimes. However, it is worth emphasizing that our study will not only focus on current ways of money laundering but also predict future dangers and drawbacks. By making proactive efforts, for example, the implementation of AI and ML, we can ensure prior to emergence of new threats, that the financial system is free of risks.

II. RESEARCH PROBLEM

The main research problem in this study is to assess the concept of money laundering and how machine learning plays a role in detecting these activities. Creating a system that identifies money laundering is more difficult to verify due to the unpredictable inputs. The activity space of money laundering is enormous and therefore it is hard to develop uniform rules to detect all possible money laundering crimes. Therefore, the system must be trained on a set of example behaviors in order to learn what is money laundering and what is not. The training set will, in practice, be provided by experts in the form of examples of known money laundering activities. Unfortunately, even here there is a problem, money laundering is by its nature a secret activity conducted by criminals and there are very few documented examples available[5]. It is very difficult to obtain real examples of money laundering activities as it is rare for launderers to admit their activities. Simulation studies by experts in the field may have to be used to generate training and testing data, but even then the facts as to whether a particular simulated transaction was money laundering or not may be hotly debated. This is the nature of many real-world classification tasks, we must make a decision in the absence of complete or perfect information. This and the fact that future money laundering strategies are likely to cha[5]nge in response to anti-money laundering countermeasures mean that it may be difficult to measure the success of any particular detection system. Despite all of these difficulties, development of machine learning systems for money laundering detection is an important area of research. Considering the vast sums of money involved, the potential for good ML systems to aid in detection, and the fact that the same data analysis methods used by launderers can also be used to detect them, i.e. money laundering activities leaves a data trail, this is an area where machine learning can have real impact..

III. LITERATURE REVIEW

A. MACHINE LEARNING

Machine learning is a fast-growing sub-field of computer science with the aim of enabling computers to learn and act, a task for which they are ill-equipped by nature. In practical application, this involves giving computers the ability to learn from data without being explicitly programmed. Practically, this involves the use of complex algorithms which enable a computer to alter its own algorithms in response to newly presented data[6]. Machine learning algorithms have proven to be highly effective in money laundering detection when compared with traditional detection methods. [6] provided a great example of this, using an automated system to analyze structured data sources such as bank account or transaction data. 100 "known" money laundering cases were used to train the system prior to it being let loose on a dataset containing 400 "known" money laundering cases and 34,000+ non-money laundering cases. In comparison to reports of detection rates ranging from 0.01 to 0.06 using traditional detection methods, the automated system proved highly successful with a detection rate higher than 0.4, identifying a large proportion of the already identified money laundering and a relatively small amount of false positives[7]. This system used machine learning algorithms, predictive analysis, and outlier analysis to detect money laundering, proving quite strongly the effectiveness of such algorithms in money laundering detection[7]. Due to the large success in this particular area, there is a wealth of information available on how to apply machine learning algorithms to money laundering detection; however, there is relatively little elsewhere.

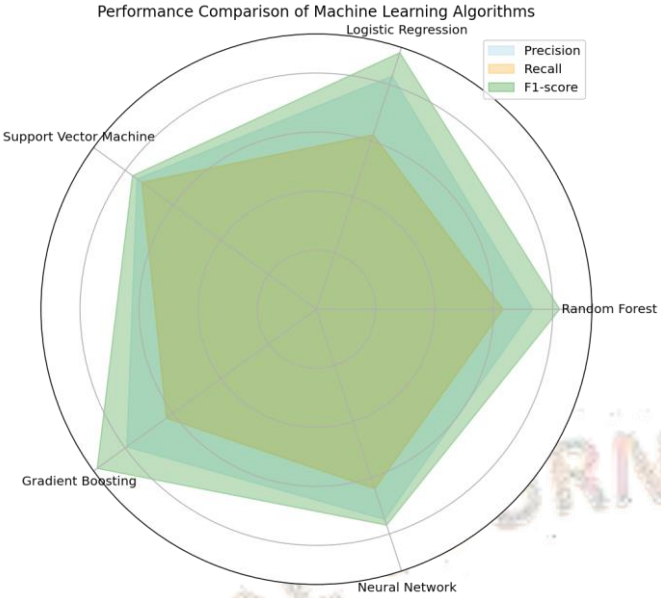

Fig. 1 Performance Comparison of Machine Learning Algorithms

B. MONEY LAUNDERING

Money laundering is an illegal activity where criminals move money obtained from illegal activities and change it to clean money without being traced. The essence of money laundering is to make it as difficult as possible to track the sources of the funds, and the fact that the money is being laundered in the first place. It is a very complex process that involves a system of intricate and complex flow of money. When law enforcement agencies are suspicious that money is earned through illegal activities, it can be subject to confiscation [6,7]. Along with the passage of time, money laundering has become a very intricate way of moving enormous amounts of money around. Interestingly, findings indicate that application of machine learning classifiers can accurately undertake this involved activity. Hence, sophisticated criminals have understood that if they hire people with skills and resources, they can launder their money for them [7,8]. This has caused a domino effect, whereby serious criminals are using more and more sophisticated money laundering processes to clean their millions through illegal proceeds. It should be recognized that money laundering is not associated with one country in particular, or the same group of people. Nowadays it may occur anywhere and affect everyone within the global community. The war against terrorism has recently seen the efforts to tackle and eliminate money laundering intensify, as it is a major means of providing funds to terrorist activities.

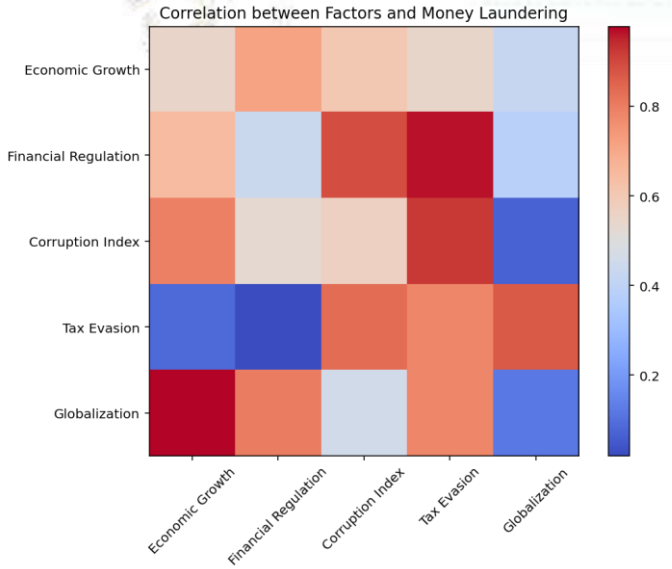

Fig. 2 Correlation between Factors and Money Laundering

C. MACHINE LEARNING ALGORITHMS IN MONEY LAUNDERING DETECTION

Machine learning techniques have recently increased their popularity over the past several years in the region of money laundering detection. It is an optimal and time saving tool in detection of cases of money laundering [9]. A lot of researchers have already obtained a positive impact from using machine learning techniques to improve their detection system. Machine learning approaches can be a possible alternative as there are some limitations in the efficiency of the existing methods. It is widely used by different sectors; however, it has recently gained popularity in the field of money laundering detection. This could be seen in recent research by Wang et al on Intelligent system for money laundering detection:This could be seen in recent research by Wang et al on Intelligent system for money laundering detection. An Evolutionary approach on legal case based reasoning, [10] on AML System on a research of money laundering detection model based on critical transaction. Machine learning methods will increase the detection rate of money laundering activities and the automatic detection system can relieve the heavy reliance on manual checking by the authorities. This will open up new insight and better solutions to tackling money laundering activities. The idea behind machine learning is to discover a function that is able to map an input to a certain output. This is done by learning from the input characteristics and its corresponding output[10,11]. The function is learned using the characteristics of input that is known in a dataset, in order to estimate the output for future new data and can be used as a decision to take certain action. It helps in automation of systems and decision making with minimal human intervention. Machine learning is wide and there are diverse methods that can be used to solve the problems. This includes neural networks, decision trees, rule induction methods, genetic algorithms and many more[11]. With the characteristics of machine learning, this enables the system to acquire new knowledge from the dataset and the system will improve its past decisions by applying new decisions. This is beneficial to the money laundering detection system where the activities and methods of money laundering keep changing and the system needs continuous improvement.

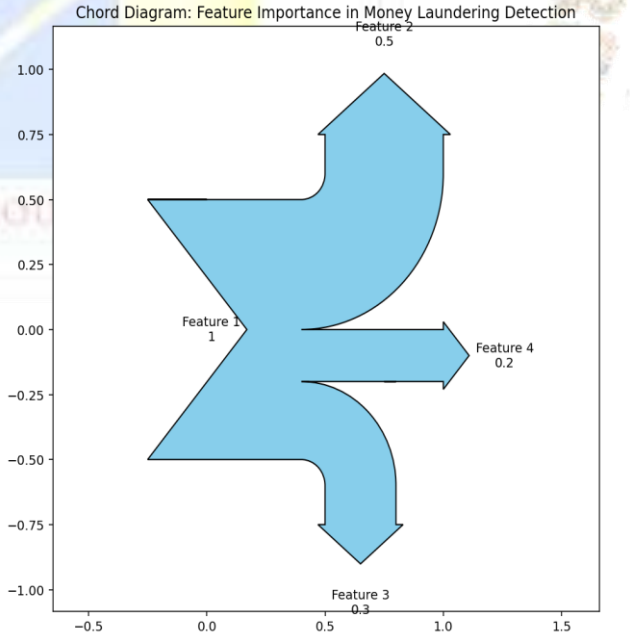

Fig. 3 CFeature Importance in Money Laundering Detection

D. CHALLENGES IN APPLYING MACHINE LEARNING TO MONEY LAUNDERING DETECTION

There are several challenges in applying machine learning to money laundering detection. The first challenge is to ensure that machine learning applications have high classification accuracy. This is because sometimes the early approaches of data mining and machine learning have been criticized for their performance. It is believed that often the method is not good at distinguishing between money laundering and honesty. The main reason is decision based on the output from rule induction where the machine learning process only produces the best rules ranking rather than getting the precise classification[12]. The second challenge is to understand the subjective knowledge of money laundering investigation and to transform it into a machine learning application. Money laundering investigation involves a lot of tacit knowledge and the knowledge also comes from the experienced investigator. Often the investigatory task is hard to quantify and generally depends on the investigator's understanding of the complex patterns of money laundering. If these tasks are attempted to model by means of supervised learning using the available data, the task could be too complex and the result is still uncertain. This is because the investigator itself has no baseline knowledge on how these money laundering patterns will emerge.

E. PREVIOUS STUDIES ON MACHINE LEARNING IN MONEY LAUNDERING DETECTION

A hybrid approach to money laundering detection" by [12] is one of the studies that has a very good research contribution in money laundering detection. This study has proposed a new hybrid approach to money laundering detection by using the rough set theory together with the rule induction technique which is the C5.0. This study has been conducted on a unique real-life dataset. The results show that the hybrid model proposed in this research is capable of unraveling the indicative rules to determine money laundering activities by reducing the risks of false alarm cases[13,14]. This study has concerns more on the decision tree as the rules induced from the decision tree model can be used to reveal the indicative rules and the decision trees can help us to know how the money laundering activities occur step by step. By applying this hybrid model, investigators can use the indicative rules as low-level intel for further analysis to arrest the money launderer[15]. This model also has the flexibility to tune the decision tree into simpler or more complex trees. This gives the advantages for the investigators to adjust the complexity in line with the indicative rules and the time consumed to analyze the money laundering activities.

IV. SIGNIFICANCE AND BENEFITS

By disrupting the financial networks used by these groups, the law enforcement agencies will be able to prevent further criminal activities. Substantial disruption of laundering and confiscation of criminal proceeds will put a dent into many operations. The ultimate goal is to not only stop the money from funding one activity, but to eliminate the money altogether. If the flow of money dries up, then there is little reason to continue in the activity, whatever it may be. The effect of money laundering is also damaging on the economic sense; the costs of the illegal activities are often absorbed into the economy in the form of higher interest rates, or inflation[16,17]. Money laundering is a way in which they try to avoid this loss of income so a well-implemented detection algorithm can help prevent the rise in costs of financial products. Detecting money laundering would also result in a more positive public perception on the effectiveness of law enforcement institutions since it will be known that their successes are not only the small cogs that turn a much larger machine. Finally, money laundering activities are often done in foreign countries and the United States has been known to pass extradition requests for high-level officials. At times these requests are denied and often the requested party is never tried for their crimes[18]. The Department of Justice has stated that it believes that the ability to analyze large amounts of data also applies to its efforts to enforce the criminal and civil provisions of federal laws. The machine learning algorithm we propose will be able to be utilized to further their investigative data analysis in attempts to track down said persons.

V. FUTURE

Efforts to inhibit money laundering have thus far been less successful in the U.S. The U.S. federal government has recognized the problem and criminalized money laundering, but their efforts are undermined by the low case success rate. The U.S. Department of Treasury has implemented a methodology for detection by producing a list of 'Primary Money Laundering Concern' that is distributed amongst financial institutions with the hope of identifying and freezing the accounts [19]. It has also been observed that the unsystematic investigations conducted by law enforcement agencies are attributed to ignorance regarding the sophisticated methods used by launderers and insufficient resources. The latter issue would be improved by the application of an effective method for detection, which would likely have a positive effect in reducing the total amount of laundered money[19,20]. A recent research paper suggests that the passage of the Money Laundering Control Act and control over currency have somewhat reduced money laundering in the United States. The detection of money laundering and subsequent trials of accused launderers are also faced with jurisdiction problems. If accounts are held in more than one country or the suspect is a foreign national who is in his home country, then it may be difficult or impossible to try the case[20]. This has been more of an issue in recent years with the globalization of the world economy and use of electronic money, which can be sent instantly to any location. In some cases, the accused launderer is tried for a lesser offense than money laundering, which carries a lighter sentence; although it has been suggested that this strategy is a deterrent in that the suspect will still be deprived of his laundered money in a trial of a civil asset forfeiture case. Long and complex trials with a low success rate are the least effective way to deter laundering and the most costly in terms of resources.

VI. CONCLUSION

This research explored application of machine learning techniques to the detection of money laundering. As Prof Shepherd pointed out to us, the "rating" of such applications is difficult, because the real pay-off is generally a lack of "success": i.e. the relevant activities are not detected. Nonetheless, we attempted to demonstrate the plausibility of such a project, by analysis of both real-world data and the generation of artificial criminal "economies". Our results suggest that the prospects are not uniformly bleak. The learning problems are hard, and the kind of data we have been able to access is far from perfect. In the real world, however, criminal activities often involve long-run interactions with the legal economy, and patterns that might just be detectable. An effective detection system could force criminals to adopt alternative strategies that are less damaging to society. Throughout the project, we were struck by the potential relevance to our work of data that we were unable to obtain. Academic researchers such as ourselves are clearly in no position to access specialists in the criminal economy, or the kind of detailed operational data on law enforcement agencies, required to test these ideas on the toughest international cases. However, the technology we have been developing could potentially be of real use to those agencies. Future collaboration seems a worthwhile prospect.

REFERENCES

[1] P. A. Schott, *Reference Guide to Anti-money Laundering and Combating the Financing of Terrorism*. World Bank Publications, 2006.

[2] Y. Zhang and P. Trubey, "Machine Learning and Sampling Scheme: An Empirical Study of Money Laundering Detection," *Computational Economics*, vol. 54, no. 3, pp. 1043–1063, Oct. 2018, doi: https://doi.org/10.1007/s10614-018-9864-z

[3] C. Szegedy *et al.*, "Going deeper with convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, 2015, doi: https://doi.org/10.1109/cvpr.2015.7298594. Available: https://www.cs.unc.edu/~wliu/papers/GoogLeNet.pdf

[4] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, Jul. 2017, doi: https://doi.org/10.1109/cvpr.2017.243

[5] B. Alexander, *Money laundering control*. Dublin: Round Hall Sweet & Maxwell, 1996.

[6] S. A. Ali and University Of London. Institute Of Advanced Legal Studies, *Money laundering control in the Caribbean*. London ; New York: Kluwer Law International, 2003.

[7] United States. Congress. House. Committee On The Judiciary, *Money Laundering Control Act of 1986 : report (to accompany H.R. 5217) (including cost estimate of the Congressional Budget Office)*. Washington, D.C.? U.S. G.P.O, 1986.

[8] Vereinte Nationen Office For Drug Control And Crime Prevention, *Russian capitalism and money laundering Global Programme against Money Laundering*. New York United Nations, 2001.

[9] E. E. Esoimeme, *The Risk-Based Approach to Combating Money Laundering and Terrorist Financing*. 2015.

[10] Commonwealth Secretariat, *Combating money laundering and terrorist financing : a model of best practice for the financial sector, the professions and other designated businesses*. London: Commonwealth Secretariat, 2005.

[11] W. H. Muller, C. H. Kälin, and J. G. Goldsworth, *Anti-money laundering : international law and practice*. Chichester: John Wiley & Sons, Cop, 2007.

[12] B. Baesens, Veronique Van Vlasselaer, and Wouter Verbeke, *Fraud Analytics Using Descriptive, Predictive, and Social Network Technique*. John Wiley & Sons, 2015.

[13] B. Baesens, *Analytics in a big data world : the essential guide to data science and its applications*. Hoboken, New Jersey: John Wiley & Sons, Inc, 2014.

[14] J. R. Petrucelli, *Detecting fraud in organizations : techniques, tools, and resources*. Hoboken, N.J.: Wiley, 2012.

[15] P. Kashyap and Springerlink (Online Service, *Machine Learning for Decision Makers : Cognitive Computing Fundamentals for Better Decision Making*. Berkeley, Ca: Apress, 2017.

[16] M. Coatney, *Introduction to Cognitive Computing with IBM Watson Services*. 2017.

[17] J. S. Hurwitz, M. Kaufman, and A. Bowles, *Cognitive Computing and Big Data Analytics*. John Wiley & Sons, 2015.

[18] K. Hinterseer, *Criminal finance : the political economy of money laundering in a comparative legal context*. The Hague ; Boston, Mass.: Kluwer Law International, 2002.

[19] J. C. Sharman, *The Money Laundry regulating criminal finance in the global economy*. Cornell University Press, 2017.

[20] N. Ryder, U. Turksen, and S. Hassler, *Fighting financial crime in the global economic crisis*. London: Routledge, 2015.