International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 2, pp 143-147, February-2024

https://doi.org/10.47001/IRJIET/2024.802021

Machine Learning for Anti-Money Laundering and Fraud Detection

[1]Tasnim Shamsuddin Pathiyaparambath, [2]Prof. P. Gopika

[1]PG Student, Dept. of Computer Science and Engineering, EASA College of Engineering and Technology, Tamilnadu, India
[2]Professor, Dept. of Computer Science and Engineering, EASA College of Engineering and Technology, Tamilnadu, India

***Abstract* - Financial institutions must meet international regulations to ensure not to provide services to criminals and terrorists. They also need to continuously monitor financial transactions to detect suspicious activities. Businesses have many operations that monitor and validate their customer's information against sources that either confirm their identities or disprove. Failing to detect unclean transaction will result in harmful consequences on the financial institution responsible for that such as warnings or fines depending on the transaction severity level. The financial institutions use Anti-money laundering (AML) software sanctions screening and Watch-list filtering to monitor every transaction within the financial network to verify that none of the transactions can be used to do business with forbidden people. Lately, the financial industry and academia have agreed that machine learning (ML) may have a significant impact on monitoring money transaction tools to fight money laundering. Several research work and implementations have been done on Know Your Customer (KYC) systems. To overcome this problem we propose an efficient Anti-Money Laundering System which can able to identify the traversal path of the Laundered money using the Hash-based Association approach and success in identifying agents and integrators in the layering stage of Money Laundering by Graph-Theoretic Approach. Also, detect credit card fraud. Also, it will minimize the compliance officers' effort, and provide faster processing time.**

Keywords: Data mining, Anti-Money Laundering, Fraud Detection, Suspicious Transaction, AML, Supervised Learning Methods.

I. INTRODUCTION

Money Laundering is a criminal activity to disguise black money as white money. It is a process by which illegal funds and assets are converted into legitimate funds and assets. Money laundering occurs in three stages: Placement, Layering, and Integration. It leads to various criminal activities like Political corruption, Smuggling, Financial Frauds, etc. The Reserve Bank of India (RBI), has issued guidelines to identify suspicious transactions and send it to the Financial Intelligence Unit (FIU). FIU verifies if the transactions are suspicious or not. The process is time consuming and not suitable to identify the illegal transactions that occur in the system. To overcome this problem we propose an efficient Anti-Money Laundering System which can able to identify the traversal path of the Laundered money using the Hash-based Association approach and success in identifying agents and integrators in the layering stage of Money Laundering by Graph-Theoretic Approach. Also, detect credit card fraud.

Money laundering is a process of converting unaccountable money into accountable money. Day to day the technology is getting updated and in this fast-changing technology, many merits, as well as demerits, is associated. With the advent of E-Commerce, the world has been so globalized and further the technology has made everything so user-friendly that with a single click of a button, many transactions can be performed. Fraud Detection is mandatory since it affects not only the financial institution but also the entire nation. This criminal activity is appearing more and more sophisticated and perhaps this might be the major reason for the difficulty in fraud detection. This criminal activity leads to various adverse effects ranging from drug trafficking to financial terrorism.

Dealing with illegal companies and people will lead to direct fines, and suspend business or harm the institution reputation. Considering a substantial number of transactions, and the significant amount of illegal entities, it is a must to implement an automated system to assure that financial institutions meet the compliance regulations. Currently, financial institutions perform vast volumes of transactions per day, and there are high chances of missing a suspicious transaction or even transactions. Moreover, while large and medium-sized institutions can hire armies of experienced compliance officers, small-sized firms cannot afford to do the same.

Payments related fraud is a key aspect of cyber-crime agencies and recent research has shown that machine learning techniques can be applied successfully to detect fraudulent transactions in large amounts of payments data. Such techniques have the ability to detect fraudulent transactions that human auditors may not be able to catch and also do this on a real time basis. In this project, we apply multiple supervised machine learning techniques to the problem of fraud detection using a publicly available simulated payment transactions data. We aim to demonstrate how supervised ML techniques can be used to classify data with high class imbalance with high accuracy. We demonstrate that exploratory analysis can be used to separate fraudulent and non fraudulent transactions. We also demonstrate that for a well separated dataset, tree based algorithms like Random Forest work much better than Logistic Regression. Digital payments of various forms are rapidly increasing across the world. Payments companies are experiencing rapid growth in their transactions volume. For example, PayPal processed ~$578 billion in total payments in 2018. Along with this transformation, there is also a rapid increase in financial fraud that happens in these payment systems. Preventing online financial fraud is a vital part of the work done by cybersecurity and cyber-crime teams. Most banks and financial institutions have dedicated teams of dozens of analysts building automated systems to analyze transactions taking place through their products and flag potentially fraudulent ones. Therefore, it is essential to explore the approach to solving the problem of detecting fraudulent entries/transactions in large amounts of data in order to be better prepared to solve cyber-crime cases.

II. PROBLEM FORMULATION

It is common that money launderers divide the pillage into multiple parts and make sequences of financial transfers or commercial transactions, into E-wallet thus manually tracking activities of money laundering is incredibly challenging. Smurfing is another general practice for money laundering, in which the individual divides the big amount of cash into small multiple deposits and spread into multiple accounts to hide from detection. Another commonly used practice for money laundering is wire transfers, currency exchanges, cash smugglers or mules are the individuals who involved in cross borders money laundering practice. Other money laundering techniques involve making an investment in commodities inclusive of gemstone and gold that may be effortlessly shifted to other jurisdictions, cautiously spending in and selling precious belongings such as actual property, playing counterfeiting and creating shell agencies. While conventional money-laundering methods are still used, the internet has positioned a brand new spin on an antique crime. Using the net allows money launderers to effortlessly keep away from detection. The upward thrust of the establishment of online banking, anonymous online price services, peer-to-peer transfers the usage of cellular telephones and using digital currencies like Bitcoin, have made finding the illegal transactions of money even greater difficulty.

The exploration process in the underwater is a complicated method.

One famous example of the AML penalties is the HSBC bank penalty that occurred in 2012. The bank paid $1.9billion because of the weak and insufficient money laundering controls that made the bank to be used to launder a river of drug money flowing out of Mexico. Another example was in 2009, Switzerland Credit Suisse Group was fined $536 million because helping Iran and other countries to move billions of dollars through the US banking system. Chart is Research asked a question on a survey. The International Monetary Fund has defined money laundering as the process of assets being spawned by criminal activities to hide or make obscure any connections established between the funds and their illegal origins has defined this as a process of cleaning ‘dirty money’ or changing the money from illegal to legal one, which is normally derived from criminal activities and hard to trace. In the meantime, has also regarded money laundering as something simple to explain: It is the conversion of assets produced from criminal activities into assets that cannot be traced back to crimes committed previously, and the assets will appear to have been originated from legitimate sources. Money laundering can bring about upsetting economic, social, and political consequences for countries, especially for developing countries and those countries with vulnerable financial systems. According to a study done by Young (2014), money laundering imposes a negative effect on the global economic and financial growth. It can, unfortunately, steer resources to unproductive activities and even give ways to corruption and crimes. As a result, money laundering could impede legitimate businesses and cripple banks and financial institutions. There are some obstacles or challenges that have been faced by financial institutions in a way to identify money-laundering activities there are many methods of money laundering, including surfing, which involves depositing cash at various branches of a financial institution, buying secured bank instruments and postal orders, and buying shares and currency exchanges in money services businesses.

For the convenience of comparison, we unify the entries of the two sentiment dictionaries with the sentiment dictionary in the field of e-commerce.

III. BASIC IDEA OF OUR SCHEME

The AML solutions have essential elements that work together to fight money laundering and ensure that the financial institutions are not making business with risky clients. The first element is Know Your Customer (KYC) which is the process of verifying the identity of the clients and assessing the potential risks of making a business relationship with them. The second element is Record-Keeping; this can be achieved by archive historical records for a specific period depends on the regulation of each region. The third is the Suspicious Activity Monitoring element which focuses on monitor client account. The fourth element is Customer Behaviour that requires studying the client transactions pattern. The fifth is the Customer Due Diligence (CDD) element that validates the identity of the client before opening an account by checking documents such as passports, photo card driving licenses, and at least two recent utility bills. The sixth and last element is the watch list filtering system that we are focusing on in this paper. The Watch-list Filtering system helps financial institutions to apply the required regulations and stay updated with any change or update on the regulations and blacklists' content. This system can filter all transactions and customers against all types of blacklisted entities.

This methodology served as the deliverables of the project. It describes the results of each phase that was tried out and do a comparison between them to identify which is the best technique to address the fraud detection problem.

Each phase of the project has an output that describes the findings in that phase. These deliverables were used in this final project are explained below.

Methodology Phases Project Deliverables Report on the summary of the data set and each variable Understanding the data set it contains along with necessary visualizations Exploratory Data Analysis Report on analysis conducted and critical findings with a full description of data slices considered Hypothesis about the separation between fraud and nonfraud transactions Visualizations and charts that show the differences between fraud and non-fraud transactions Python code of the analysis performed Modeling Report on the results of the different techniques tried out, iterations that were experimented with, data transformations and the detailed modeling approach Python code used to build machine learning models Final Project Report Final report summarizing the work done over the course of the project, highlighting the key findings, comparing different models and identifying best model for financial fraud detection.

IV. RELATED WORK

1) Applying data mining in money laundering detection for the Vietnamese banking industry, D. Cao and P. Do's. In 2006, The World bank warned that Vietnam is becoming an easy target for money laundering activities because of the weakness in the inspection, supervision, auditing and customer relationship management systems. The level of using cash and several kinds of payment method make transactions become out of control. Meanwhile banking plays an important role in cleaning dirty money in the money laundering process. Thus, the requirement for a mechanism to detect money laundering is growing with great importance for all over the world.

2) The application of social network analysis algorithms in a system supporting money laundering detection, R. Dreáewski, J. Sepielak, and W. Filipkowski's. An example of link analysis is the system that supports money laundering detection. Dreaewskiet al. proposed a system consist of three components that support money laundering detection. The first component is clustering; it refers to constructing graphs that represent the flow of money and captures only suspicious money transfers between groups of accounts. The second component is mining for frequent patterns in clusters. The third component is data visualization, and concerns about displaying the result and the transactions flow.

3) Data mining techniques and its applications in banking sector, K. Chitra and B. Subashini's. The idea was to use the EM cluster in grouping the data into a similar cluster by building a model using the historical transaction behavior for each bank. Data mining process can be broken down to the following iterative sequence of following steps. Data Integration In a production environment, there could be multiple databases storing same information. These heterogeneous data sources are combined in a common source. Data Transformation and Data Reduction Data are transformed or consolidated by performing summary or aggregation operations so that they are simpler to handle for the mining operations.

4) Suspicious activity reporting using dynamic Bayesiannetworks, S. Raza and S. Haider's. For the risk scoring aspect, Sudhakar and Reddy in used the DT algorithm and proposed a two-step loan credibility prediction system that made it easy for the financial institutions to make the right decision to approve or reject the loan request of the customers with the application of DT algorithm. The authors have clarified that credit risk management is critical for a successful bank loan process. Building this model will need five main phases, including problem understanding, data understanding, data filtering, system modeling, and system evaluation.

5) Study on anti-money laundering service system of online payment based on union-bank mode, Q. Yang, B. Feng, and P. Song's. Finally, for the geographic capability aspect, Yang proposed an AML service system for a union bank to detect money laundering on online payment using the neural network algorithm. The logical framework for this proposed method contains five sequential layers: database layer, basic data resource base layer, data analysis layer, application service layer, and the interface layer. The database layer gathers transaction information. Then, the basic data resource

ISSN (online): 2581-3048

Volume 8, Issue 2, pp 143-147, February-2024

https://doi.org/10.47001/IRJIET/2024.802021

layer contains a knowledge base, case base, and other useful information that enabled the discovery of money laundering cases.

V. CONCLUSION

In conclusion, we successfully developed a framework for detecting fraudulent transactions in financial data. This framework will help understand the nuances of fraud detection such as the creation of derived variables that may help separate the classes, addressing class imbalance and choosing the right machine learning algorithm. We experimented with two machine learning algorithms – Logistic Regression and Random Forest. The Random Forest algorithm gave far better results than Logistic Regression indicating tree-based algorithms work well for transactions data with well differentiated classes. This also emphasizes the usefulness of conducting rigorous exploratory analysis to understand the data in detail before developing machine learning models. Through this exploratory analysis, we derived a few features that differentiated the classes better than the raw data.

VI. FUTURE WORK

The frequent accounts mustn't be the sole criteria for locating out the suspicious dealing as there is also a case once the dealing doesn't occur often but even then they are illegal. To trace out such cases further parameters ought to be thought of. Through this project, we demonstrated that it is possible to identify fraudulent transactions in financial transactions data with very high accuracy despite the high-class imbalance. We provide the following recommendations from this exercise - Fraud detection in transactions data where transaction amount and balances of the recipient and originator are available can be best performed using tree-based algorithms like Random Forest Using dispersion and scatter plots to visualize the separation between fraud and non-fraud transactions is essential to choose the right features To address the high-class imbalance typical in fraud detection problems, sampling techniques like under sampling, oversampling, SMOTE can be used. However, there are limitations in terms of computation requirements with these approaches, especially when dealing with big data sets. To measure the performance of fraud detection systems, we need to be careful about choosing the right measure. The recall parameter is a good measure as it captures whether a good number of fraudulent transactions are correctly classified or not. We should not rely only on accuracy as it can be misleading.

REFERENCES

[1] (2012). Study Guide for The CAMS (Certified Anti-Money Laundering Specialists). [Online]. Available: https://www.acams.org/

[2] R. J. Lowe, ``Anti-money laundering The need for intelligence,'' J. Financial Crime, vol. 24, no. 3, pp. 472_479, Jul. 2017.

[3] M. Hinkle and D. Stewart. (2014). Modernizing Anti-Money Laundering Practices.

[4] C. Gatti and J. Eligon. (Dec. 2009). Iranian Dealings Lead to a Fine for Credit Suisse. [Online]. Available: https://www.nytimes.com/2009/12/16/business/16bank.html/nytimes.co

[5] Chartis. (2017). RiskTech100 RiskTech100 2018. [Online]. Available: https://www.risktech100.com/2018-report

[6] R. Ramachandran. (2014). OFAC Name Matching and False Positive Reduction Techniques. [Online]. Available: https://www.cognizant.com/InsightsWhitepapers/OFAC-Name-Matching-and-False-Positive-Reduction-Techniques-codex1016.pdf

[7] Standards MT. (Nov. 2019). Standards Release Guide. [Online]. Available: https://www.swift.com/standards

[8] A.Lait and B. Randell, ``An assessment of name matching algorithms,'' Series-Univ., Tech. Rep., 2006. [Online]. Available: https://www.semanticscholar.org/paper/An-Assessment-of-Name-Matching-Algorithms-Lait-Randell/bd88761329102ea617c1c3173cf11efac4ae7876

[9] M. Mohri, A. Rostamizadeh, and A. Talwalkar, Foundations of Machine Learning. Cambridge, MA, USA: MIT Press, 2012.

[10] Chartis. (2019). Artificial Intelligence in Financial Services, 2019: Demand-side Analysis. [Online]. Available: https://www.chartisresearch.com/technology/artficial-intelligence-ai/artificial-intelligence-financial-services-2019-demand-side-analysis-10716.

[11] Dr. Anna Saro Vijendran, D. Nethra Pingala Suthishni "Reduction of RoutingOverhead Using Cluster-Fuzzy Algorithm In MANET" International Journal of Scientific & Technology Research (IJSTR), VOLUME 9, ISSUE 2.

[12] KS Senthilkumar, An Implementation Of Eavarp And Eavar-Nca In Underwater Sensor Network "Turkish Journal of Computer and Mathematics Education (TURCOMAT) ", VOLUME 12, ISSUE-10.

[13] S. Gnanapriya "SENTIMENT ANALYSIS BASED ON FINE GRAINED FEATURE REPRESENTATIONOF DOMAIN SENTIMENT DICTIONARY "ISSN: 2229-6956 (ONLINE) ICTACT JOURNAL ON SOFT COMPUTING, APRIL 2023, VOLUME: 13, ISSUE: 03.

Citation of this Article:

Tasnim Shamsuddin Pathiyaparambath, Prof. P. Gopika, “Machine Learning for Anti-Money Laundering and Fraud Detection” Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 8, Issue 2, pp 143-147, February 2024. Article DOI https://doi.org/10.47001/IRJIET/2024.802021

****