Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 11, November : 2024

A MACHINE LEARNING FRAMEWORK FOR MONEY LAUNDERING DETECTION BASED ON STATISTICAL ANALYSIS

[1] **V. Surendra Reddy,** Assistant Professor, [2] **N. Subramanyam,** Assistant Professor,
[3] **B.V.S Uma Prathusha,** Assistant Professor, [4] **SK. Karimuni**, Assistant Professor
[1234]Audisankara College of Engineering & Technology (AUTONOMOUS), Gudur (M), Tirupati (D), A.P, India-524101

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 paper, we focus on anti- money laundering in banks and provide an introduction and review of the literature. We propose 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.

Index Terms—Anti-money laundering, know-your-client, machine learning, literature review.

I. INTRODUCTION

Officials from the United Nations Office on Drugs and Crime estimate that money laundering amounts to 2.1-4% of the world economy. The illicit financial flows help criminals avoid prosecution and undermine public trust in financial institutions. Multiple intergovernmental and private organizations assert that modern statistical and machine learning methods hold great promise to improve anti-money laundering (AML) operations. The hope, among other things, is to identify new types of money laundering 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.

The international framework for AML is based on recommendations by the Financial Action Task Force (FATF). Within the framework, any interaction with criminal proceeds practically corresponds to money laundering from a bank

perspective (regardless of intent or transaction complexity). Furthermore, 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. 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. 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). 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 AML in banks and aim to provide a technical review that researchers and industry practitioners (statisticians and machine learning engineers) can use as a guide to the current literature on statistical and machine learning methods for AML in banks. Furthermore, we aim to provide a terminology that can facilitate policy discussions, and to provide guidance on open challenges within the literature. To achieve our aims, we (i) propose a unified terminology for AML in banks, (ii) review selected exemplary methods, and (iii) present recent machine

learning concepts that may improve AML.

II. LITERATURE SURVEY

A) Estimating illicit financial flows resulting from drug trafficking and other transnational organized crimes

This study's review of relevant reports concludes that the best estimate for the amount of criminally obtained money laundered by transnational organized crime is approximately 2.7 percent of the global gross domestic product (GDP) in 2009, which amounted to U.S. $1.6 trillion. The largest income for transnational organized crime apparently comes from the sale of illicit drugs, which accounts for 20 percent of all crime proceeds. This study's estimate of gross profits from global cocaine sales in 2009 is U.S. $84 billion, compared with approximately U.S. $1 billion paid to farmers in the Andean region. Most of the gross profits (retail and wholesale) were generated in North America (U.S. $35 billion) and in West and Central Europe (U.S. $26 billion). This report reminds readers that investments of illicit money into licit economies can cause problems that range from distortions of resources allocation to the "crowding out" of legitimate economic sectors. The largest outflows of illicit proceeds for laundering occur from countries in North America, South America, and Europe. These regions together account for
95 percent of all cocaine profit-related outflows worldwide. In terms of net outflows (outflows less inflows), the study model suggests that the main destination outside the regions where the profits were generated would be the Caribbean, with net inflows of approximately U.S. $6 billion.

B) Consequences of money laundering and financial crime

Money laundering is seen as critical to the effective operation of transnational and organized crime. However, money laundering effects a country's economy, government, and social well-being. This article briefly reviewed both the economic and social costs of money laundering. The economic effects of money laundering discussed included: (1) undermining the legitimate private sector; (2) undermining the integrity of financial markets; (3) loss of control of economic policy; (4) economic distortion and instability; (5) loss of revenue; (6) risks of privatization efforts; and (7) reputation risk.

The social costs of money laundering include allowing drug traffickers, smugglers, and other criminal to expand operations and the transfer of economic power from the market, government, and citizens to criminals. In extreme cases, money laundering can lead to a complete takeover of

legitimate government. Anti-money laundering efforts are both a critical and effective component of anti-crime programs. Money laundering presents a complex and dynamic challenge across the world. The shear global nature of money laundering requires global standards and increased international cooperation in order to reduce the ability of criminals to launder their proceeds and carry out criminal activities.

C) Application of technological solutions in the fight against money laundering A systematic literature review

With the growing interest in technological solutions aimed at combating money laundering, several studies involving the application of technology have been carried out. However, there were no records of studies aimed at identifying, selecting, rigorously analyzing and synthesizing the literature on solutions that adopt technology to combat money laundering. This paper presents a systematic review of the literature on the application of technological solutions in the fight against money laundering. Seventy-one papers were selected from the 795 studies initially retrieved for data extraction, analysis and synthesis based on predefined inclusion and exclusion criteria. The results obtained with the data analysis made it possible to identify a general categorization of the domains of application of the approaches, as well as a mapping and classification of the support mechanisms adopted. The findings of this review showed that, among the application domain categories identified, the detection of suspicious transactions attracted greater attention from researchers. Regarding the support mechanisms adopted, the application of data mining techniques was used more extensively to detect money laundering. Topics for further research and refinement were also identified, such as the need for a better description of data analysis to provide more convincing evidence to support the benefits presented.

III PROPOSED SYSTEM

The overview of our proposed system is shown in the below figure.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 11, November : 2024

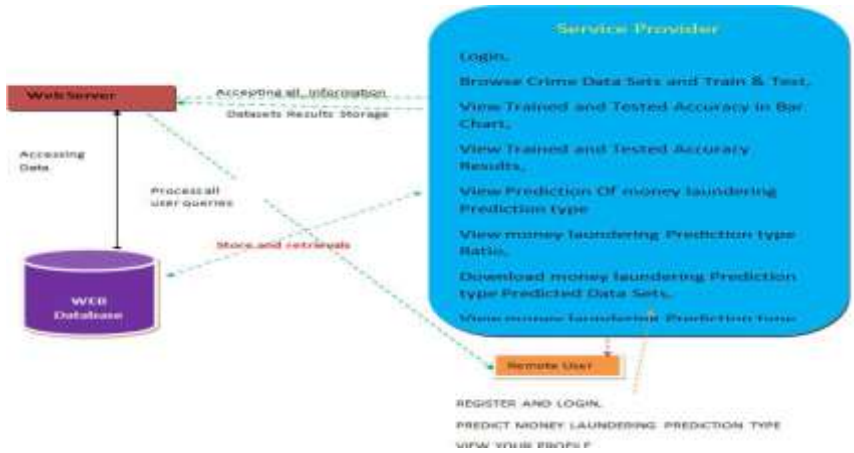

Fig. 1: System Overview

Implementation Modules

Service Provider

- In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse the dataset, view trained and tested results, view predicted money laundering prediction type , view predicted money laundering prediction type ratio and view all remote users

Train and Test Model

- In this module, the service provider split the Used dataset into train and test data of ratio 70 % and 30 % respectively. The 70% of the data is consider as train data which is used to train the model and 30% of the data is consider as test which is used to test the model

Remote User

- ✓ In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized username and password. Once Login is successful user will do some operations like predict money laundering prediction type, View Your Profile.

Classification

- In this module, user enter data and classify them using tested machine learning models

Implementation Algorithms

CNN

- In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of artificial neural network (ANN), most commonly applied to analyze visual imagery.

- CNNs are also known as Shift Invariant, based on the shared-weight architecture of the convolution kernels or filters that slide along input features and provide translation-equivariant responses known as feature maps.

Support Vector Machine

- In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis. An SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

Gradient Boosting Classifier

- Gradient Boosting is a powerful boosting algorithm that combines several weak learners into strong learners, in which each new model is trained to minimize the loss function such as mean squared error or cross- entropy of the previous model using gradient descent. In each iteration, the algorithm computes the gradient of the loss function with respect to the predictions of the current ensemble and then trains a new weak model to minimize this gradient. The predictions of the new model are then added to the ensemble, and the process is repeated until a stopping criterion is met.

IV RESULTS

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 11, November : 2024

Fig. 2: Login Page

Fig. 3: Models Accuracy Details

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 11, November : 2024

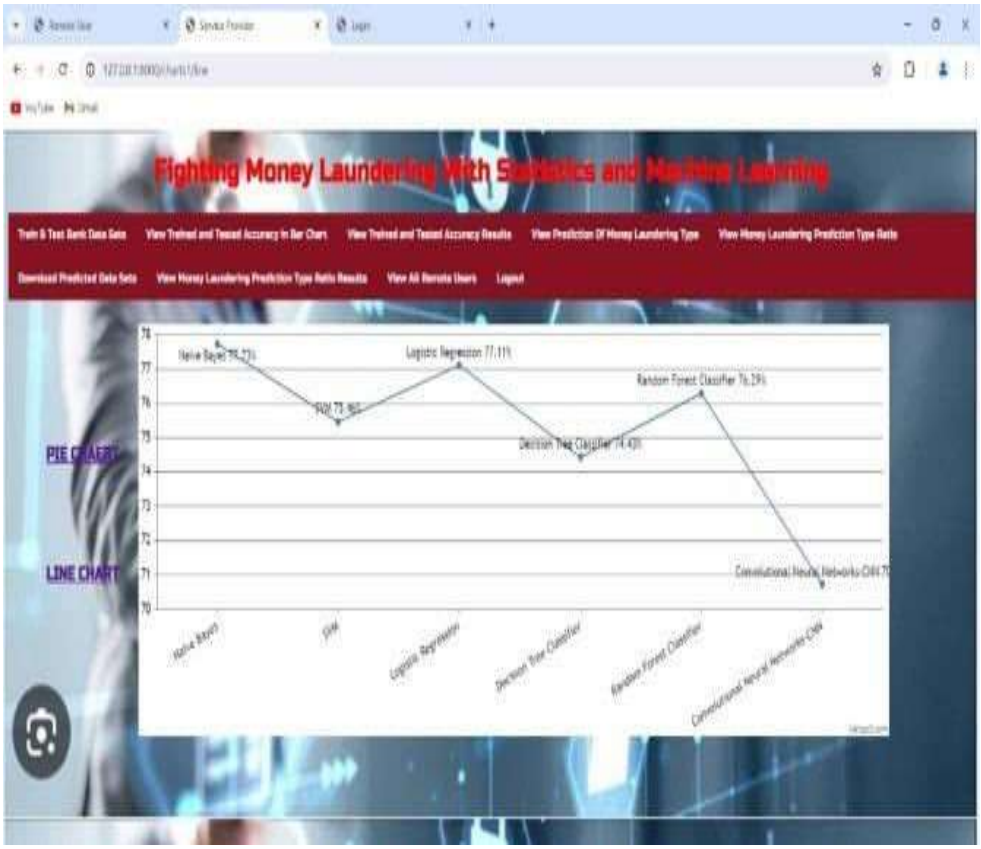

Fig. 4: Accuracy of Models in Line Chart

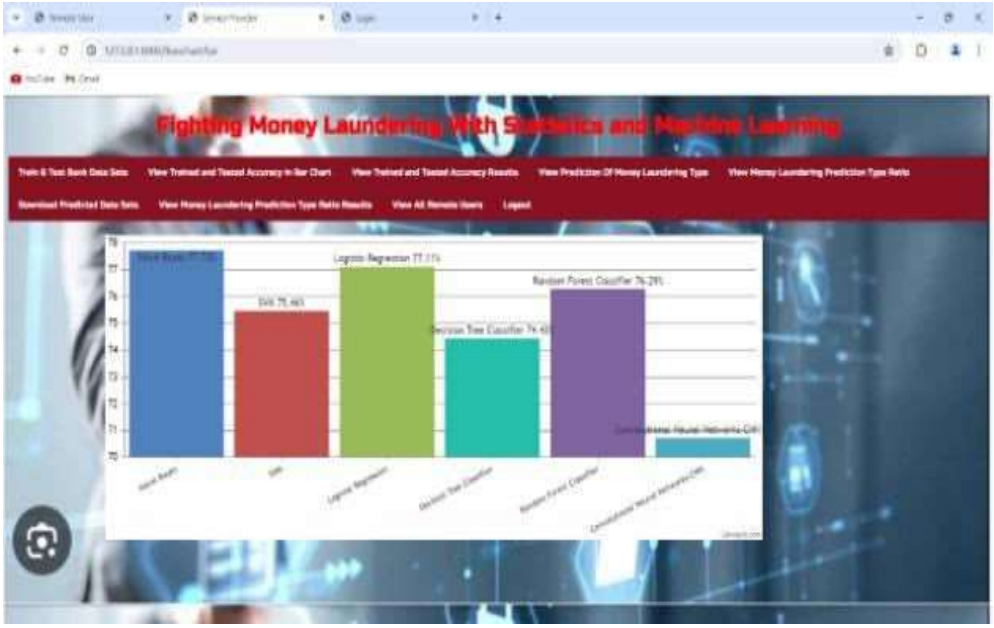

Fig. 5: Accuracy of Models in Bar Chart.

V CONCLUSION

Inspired by FATF's recommendations, we propose a terminology for AML in banks structured around two central tasks:(i) client risk profiling and (ii) suspicious behavior flagging. The former assigns general risk scores to clients (e.g., for use in KYC operations) while the latter raises alarms on clients, accounts, or transactions (e.g., for use in transaction monitoring). Our review reveals that the literature on client risk profiling is characterized by diagnostics, i.e., efforts to find and explain risk factors. The literature on suspicious behavior flagging, on the other hand, is characterized by non- disclosed features and hand-crafted risk indices. In general, we find that the literature on AML in banks is plagued by a number of problems. Two challenges are class imbalance and a lack of public data sets. To address class imbalance, a multitude of different data augmentation methods may be used. Motivated by the sensitivity of bank data, synthetic data generation may be a viable way to address the lack of public data sets.

REFERENCES

[1] T. Pietschmann, J. Walker, M. Shaw, D. Chryssikos, D. Schantz, P. Davis,C. Philip, A. Korenblik, R. Johansen, S. Kunnen, K. Kuttnig, T. L. Pichon, and S. Chawla, ''Estimating illicit financial flows resulting from drug trafficking and other transnational organized crimes,'' United Nations Office Drugs Crime, Vienna, Austria, Tech. Rep., 2011.

[2] J. McDowell and G. Novis, ''Consequences of money laundering andfinancial crime,'' Econ. Perspect., vol. 6, no. 2, pp. 6–8, May 2001.

[3] J. Ferwerda, ''The effects of money laundering,'' in Research Handbookon Money Laundering, B. Unger and D. Linde, Eds. Northampton, MA,USA: Edward Elgar, 2013, pp. 35–46.

[4] B. L. Bartlett, ''The negative effects of moneylaundering economic development,'' Platypus Mag., vol. 77, pp. 18–23, Dec. 2002.

[5] R. Grint, C. O'Driscoll, and S. Paton, ''New technologies and anti-money laundering compliance,'' Financial Conduct Authority, London,U.K., Tech. Rep., 2017. [Online].Available: https://www.fca.org.uk/publication/research/new-technologies-in-aml-final-report.pdf

[6] Opportunities and Challenges of New Technologies for AML/CFT, Financial Action Task Force, Paris, France, 2021. [Online]. Available: https://www.fatf-gafi.org/media/fatf/documents/reports/Op portunitiesChallenges-of-New-Technologies-for-AML-CFT.pdf

[7] The Wolfsberg Group. (2022). Wolfsberg Principles for Using Artificial Intelligence and Machine Learning in Financial Crime Compliance. [Online]. Available: https://www.wolfsbergprinciples.com/sites/default/files/wb/Wolfsberg%20Principles%20for%20Using%20Artificial%20Intelligence%20and%20Machine%20Learning%20in%20Financial%20Crime%20Compliance.pdf

[8] M. Biallas and F. O'Neill, ''Artificial intelligence innovation in financial services,'' Int. Finance Corp., World Bank Group, Washington, DC, USA,2020.

[9] S. Breslow, M. Hagstrom, D. Mikkelsen, and K. Robu. (2017). The new frontier in anti-money laundering. McKinsey and Company.

[10] G. S. Leite, A. B. Albuquerque, and P. R. Pinheiro, ''Application of technological solutions in the fight against money laundering—A systematic literature review,'' Appl. Sci., vol. 9, no. 22, p. 4800, Nov. 2019, doi:10.3390/app9224800.