2024 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan) | 979-8-3503-8684-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCE-Taiwan62264.2024.10674363

Unveiling the Black Box: An XAI-based Anti-Money Laundering Model

Pei-Yi Li*, Ting-Ting Chang†, Yu-Chiao Kuo‡, Chia-Yu Lin*, and Heng-Yu Chang§
* *Department of Computer Science and Information Engineering, National Central University, Taoyuan, Taiwan*
† *Department of Information Management, National Central University, Taoyuan, Taiwan*
‡ *Department of Computer Science and Engineering, Yuan Ze University, Taoyuan, Taiwan*
§ *Department of Digital Financial Technology, Chang Gung University, Taoyuan, Taiwan*
Corresponding Author: Chia-Yu Lin (sallylin0121@ncu.edu.tw)

***Abstract*—In the existing anti-money-laundering process, judging abnormal transactions still requires human resources, which is time-consuming and requires companies to pay many human costs. Many experts and scholars have used AI to identify abnormal trading behavior of accounts, but the problem of highly unbalanced data leads to poor model performances. In addition, the complex neural network of deep learning models is considered a black box, which is less likely to explain the model's results. Therefore, our research proposed an "XAI-based AI Anti-Money Laundering Model." We utilize the DNN model to detect laundering, with a recall of 0.94. By applying SHAP to the model, we evaluate the effectiveness of the dataset's ten features on the model. We find that "Payment Format" is the most crucial feature of the anti-money laundering model.**

***Index Terms*—Anti-money laundering, Explainable AI, SHAP, LIME, PDPs**

I. Introduction

As financial technology rapidly evolved, money laundering activities have sharply increased in recent years. According to the 2022 Police Monthly Report, about 32.3 billion NTD were involved in laundered funds. As a result, prevention of money laundering has become a top priority for the government and the financial industry. However, two issues need to be addressed. Firstly, financial institutions that have implemented automation processes still heavily rely on humans to make the final decision, which is time-consuming and expensive. Secondly, the current methodologies for building machine learning and deep learning models cannot determine the significance of each feature in predicting money laundering cases [1].

To solve the problems, this paper proposes an "XAI-based anti-money laundering model" to quickly and precisely detect abnormal transaction activities within financial accounts. As AI models become more complex, explainable artificial intelligence (XAI) is becoming increasingly important. To enhance the explainability and understanding of these so-called "black-box models" decision-making processes, we focus on using XAI tools such as SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and PDPs (Partial Dependence Plots). These tools help identify key features and factors influencing the model's decision-making process in complex financial environments. By improving transparency in model decisions, XAI technology leads to more effective and accurate anti-money laundering efforts.

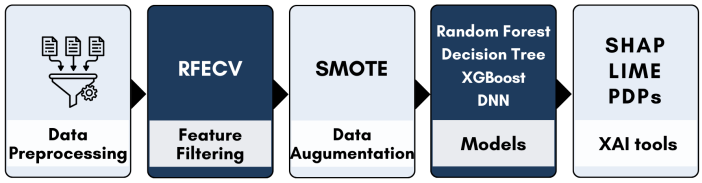

Fig. 1. Research architecture and process.

II. Methods

The diagram displayed in Fig. 1 outlines the research architecture. The data preprocessing module generates a more appropriate dataset. The feature selection module selects the most important features. The data synthesis module resolves the issue to address data imbalance concerns. Following this, we train the model and measure the efficacy of four models to select the optimal one. Our research emphasizes undertaking an in-depth exploration of three XAI tools. Ultimately, we aim to utilize SHAP to clarify the model's predictions and understand each feature's impact on the model.

A. Data preprocessing

As customer privacy presents a challenge in obtaining authentic data on money laundering, we employ the IBM Transactions for Anti Money Laundering dataset, which is publicly accessible on Kaggle [2]. This dataset comprises more than fifty thousand records with 11 feature columns. Our approach involves addressing missing values in the data and converting data types to integers or floats as required.

B. Feature selection and data synthesis

For datasets with many features, recursive Feature Elimination with Cross-Validation (RFECV) is designed to preserve impactful features by eliminating weaker features in each iteration. In addition, due to the extremely low proportion of abnormal transactions in the data, the model is prone to overfitting to these very few outlier points, potentially leading to inaccurate predictions on new data. To address this issue, we employ the Synthetic Minority Over-sampling Technique (SMOTE) to generate money-laundering data, a typical approach based on random oversampling.

TABLE I
MODEL EVALUATION

Model	Recall
Decision Tree	0.32
Random Forest	0.32
XGBoost	0.65
DNN	0.94

C. Build models

We implement decision tree, random forest, XGBoost, and DNN models to predict money laundering activities. Then, we selecte the model demonstrating the highest recall for the subsequent processes.

D. Select XAI tools for explaining model

1) SHAP: SHAP is based on game theory. [3] In machine learning, the model generates a prediction for each predicted sample, and each feature in that sample is assigned a Shapley value. This value quantifies the contribution of each model feature to the model's final output.

2) LIME: LIME aims to explain "why the model classifies a particular instance into a specific category." The analysis process of LIME begins by randomly perturbing the data to create new samples. Subsequently, the samples' weights are assigned based on similarity (e.g., feature distance). Finally, a simple linear regression model $g(z')$ is trained, and the results are interpreted based on the model coefficients [4].

3) PDPs: PDPs are a highly effective tool for Explainable AI (XAI) that can be used alongside other methods like SHAP and LIME. It is a globally applicable interpretive technique that works with any algorithm or model. PDP's computational aspect is straightforward and employs its regression partial dependence function to analyze the impact of a selected set of features on prediction results, presented in the form of line charts or contour plots [5].

III. EXPERIMENTS

A. Model Performance Evaluation

We compare the prediction performance of decision tree, random forest, XGBoost, and DNN models. We use "recall" metric to evaluate four models. This metric calculates the ratio of samples predicted as money laundering to the total number of actual money laundering transactions. From Table I, we can see that DNN model has the highest recall of 0.94.

B. Model Interpretation

We implement SHAP to interpret the DNN model and analyze the input features using a beeswarm plot in Fig. 2. The features are ranked based on their impact on the model's prediction. The graph reveals that "Payment Format" is the most crucial feature. Meanwhile, "Amount Paid" and "Amount Received" are less important. When a transaction's "Payment Format" aligns with median values of 3 or 4 (shown in purple), the SHAP value reaches a peak of 0.6. Transactions linked to "Payment Format" values 3 or 4, representing ACH (Automated Clearing House) and cash transactions, are notably prevalent in transactions flagged for potential involvement in money laundering. These specific transactions amount to 4,591, making up 88.7 percent of the total count, highlighting their substantial role in money laundering activities. This indicates a strong correlation with transactions potentially linked to money laundering activities.

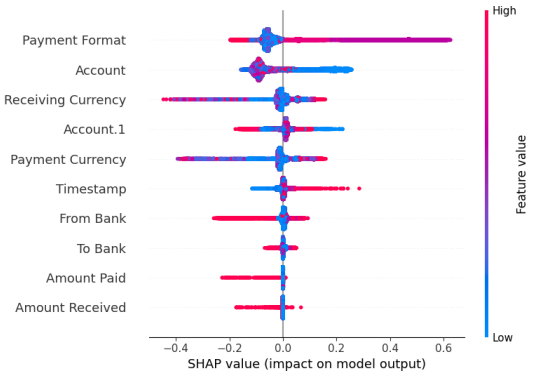

Fig. 2. SHAP's beeswarm diagram of DNN model.

IV. CONCLUSION

This paper analyzes XAI Tools and proposes an architecture of the "XAI-based AI Anti-Money Laundering Model" as a demonstrative case. The architecture consists of a data preprocessing module, a feature selection module, a data synthesis module, four prediction models, and SHAP. The XAI-based AI Anti-Money Laundering model can not only predict money-laundering transactions but also enhance its interpretability and help financial institutions better understand how the model works with XAI tools.

ACKNOWLEDGEMENTS

This work is sponsored by the National Science and Technology Council (NSTC) under the projects NSTC 110-2222-E-008-008-MY3 and NSTC 112-2622-8-A49-021.

REFERENCES

[1] Ebberth L Paula, Marcelo Ladeira, Rommel N Carvalho, and Thiago Marzagao, "Deep learning anomaly detection as support fraud investigation in brazilian exports and anti-money laundering," in *IEEE International Conference on Machine Learning and Applications*, 2016.

[2] "IBM Transactions for Anti Money Laundering (AML) — kaggle.com," https://www.kaggle.com/datasets/ealtman2019/ibm-transactions-for-anti-money-laundering-aml/data.

[3] Alvin E Roth, "Introduction to the shapley value," *The Shapley value*, pp. 1–27, 1988.

[4] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin, "Why should I trust you? explaining the predictions of any classifier," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[5] Christoph Molnar, *Interpretable machine learning*, Lulu. com, 2020.