Détection d'anomalies : une méthode appliquée aux transactions interbancaires

Hamza Chergui[*,**], Romain A. Alfred[**], Lylia Abrouk[*]
Ali Jabbari[**], Nadine Cullot[*]

[*] hchergui, ajabbari, ralfred@skaizengroup.fr
SKAIZen Group, Paris
[**] lylia.gouaich-abrouk,nadine.cullot@u-bourgogne.fr
hamza_chergui@etu.u-bourgogne.fr
Université de Bourgogne, Dijon

SWIFT [1] est une entreprise qui met à disposition un réseau interbancaire qui propose différents services tels que le transfert d'argent entre différents comptes bancaires. Ce réseau permet de réaliser des transactions financières entre plus de 11000 organismes bancaires à travers près de 200 pays. La détection d'anomalies dans la lutte contre le blanchiment d'argent est une tâche complexe pour les institutions financières. Leurs systèmes ne détectent qu'un faible pourcentage d'activités liées au blanchiment d'argent. Dans la littérature, plusieurs types d'approches d'apprentissage automatique ont été développées pour la détection d'anomalie liées au blanchiment d'argent. Ces techniques se divisent principalement en deux catégories : les techniques d'apprentissage supervisé et non supervisé. Dans ce travail, nous nous sommes focalisés sur l'utilisation de techniques d'apprentissage supervisé. Ces dernières appliquent un même principe, elles disposent d'un jeu de données labellisé contenant des transactions financières qui sont classées "normale" ou "anormale", la classe "normale" correspond aux transactions légitimes et la classe "anormale" correspond aux transactions liées à du blanchiment d'argent. Avec l'aide d'experts du domaine financier, des attributs sont sélectionnés et calculés pour mettre en valeur les transactions normales et anormales. Puis les jeux de données sont enrichis avec ces attributs pour entraîner un classificateur. Ensuite commence la phase de test et d'évaluation des résultats du classificateur à l'aide de mesures comme l'*accuracy* (précision), le F1-score ou le rappel qui permettent de vérifier l'efficacité de ses prédictions. Les approches proposées dans la littérature s'intéressent principalement à deux types d'anomalies qui portent sur les montants et les fréquences des transactions. Or, dans les transactions SWIFT, il existe également des types d'anomalies liés à d'autres attributs comme les pays ou les devises. Ces attributs correspondent à des variables dites catégoriques qui sont plus difficilement exploitables pour les algorithmes d'apprentissage supervisés. Du fait de la dimension interbancaire des transactions SWIFT, nous avons identifié un nouveau type d'anomalie propre aux transactions SWIFT lié à une combinaison inhabituelle entre le pays de l'émetteur, le pays du bénéficiaire et la devise. Dans ce travail, nous proposons une approche permettant de détecter à la fois les types d'anomalies présents dans l'état de l'art : sur les montants et les fréquences ainsi que ce nouveau type d'anomalie. Nous avons sélectionné un nouvel attribut en calculant les fréquences d'apparition des combinaisons uniques entre un pays émetteur, un pays bénéficiaire

1. https ://www.swift.com/

et une devise. Un réseau de neurones a été entraîné à partir d'un jeu de données de 500 000 transactions enrichi par notamment ce nouvel attribut ainsi que d'autres attributs figurant dans les approches étudiées. Nous avons également mené une étude comparative d'articles de l'état de l'art, traitant de la détection d'anomalies dans la lutte contre le blanchiment d'argent, en fonction des attributs pris en considération dans les anomalies détectées. Le tableau 1 montre que ces travaux ne s'intéressent qu'à deux types d'anomalies et ne manipulent pas de variables catégoriques. Nous avons obtenu des résultats satisfaisants de l'ordre de 83% d'accuracy sur trois types d'anomalies dont un qui est nouveau dans le domaine. Pour la suite de nos travaux, nous souhaitons étudier d'autres types d'anomalies et tester notre approche sur des données réelles au sein de systèmes d'institutions financières.

Référence (année)	Algorithme	Anomalie montant	Anomalie fréquence	Anomalie devise	Anomalie pays	Variable catégorique	*Accuracy*(%)
(Tang et Yin, 2005)	SVM	✓	✓				-
(Keyan et Tingting, 2011)	SVM	✓	✓				79.5
(Mubalaike et Adali, 2018)	Réseau de neurones	✓	✓				91.53
(Kumar et al., 2020)	Réseau Bayésien	✓	✓				81.00
Notre approche	Réseau de neurones	✓	✓	✓	✓	✓	83.27

TAB. 1 – *Tableau comparatif des papiers de la littérature selon des critères d'analyse.*

Références

Keyan, L. et Y. Tingting (2011). An improved support-vector network model for anti-money laundering. In *2011 Fifth International Conference on Management of e-Commerce and e-Government*, pp. 193–196. IEEE.

Kumar, A., S. Das, et V. Tyagi (2020). Anti money laundering detection using naïve bayes classifier. In *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, pp. 568–572. IEEE.

Mubalaike, A. M. et E. Adali (2018). Deep learning approach for intelligent financial fraud detection system. In *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, pp. 598–603. IEEE.

Tang, J. et J. Yin (2005). Developing an intelligent data discriminating system of anti-money laundering based on svm. In *2005 International conference on machine learning and cybernetics*, Volume 6, pp. 3453–3457. IEEE.

Summary

Money laundering activities are more frequent, financial institutions must reinforce their systems which are not efficient enough. Anomaly detection process in transactional data could help those systems. Through a study of the state-of-the-art, the existing techniques are limited for international interbank transactions. Therefore, we propose an anomaly detection approach to detect three types of anomalies in this kind of transaction using multiple variables including an index designed by ourselves. A neural network trained by a realistic data set gave us promising results in the detection of anomalies not addressed by the existing techniques. Finally, this approach aims to extend to other types of anomalies and data.