Réduction du risque du coût d'un modèle dans la détection de fraude financière.

Hamza Chergui*,**, Lylia Abrouk*, Nadine Cullot*, Nicolas Cabioch**

*Université de Bourgogne
hamza.chergui@etu.u-bourgogne.fr,
lylia.abrouk,nadine.cullot@u-bourgogne.fr
**SKAIZen Group
hchergui,ncabioch@skaizengroup.fr
https://skaizengroup.eu/

La lutte contre la fraude financière est une tâche complexe pour les institutions financières. Selon Knobel (2019), 98,9% des activités liées aux fraudes financières passent à travers les mailles du filet. Les institutions financières se doivent d'améliorer leurs systèmes sous peine de sanctions financières conséquentes des régulateurs du monde financier.

Notre travail s'inscrit dans les thématiques de recherche visant à améliorer la détection de fraude financière (DFF) avec des données provenant d'une société appelée SWIFT [1]. Cette dernière met à disposition un réseau interbancaire proposant différents services comme le transfert d'argent entre des institutions financières.

Ces dernières années, des travaux utilisant les techniques d'apprentissage automatique ont été étudiées pour la détection transactions frauduleuses. Elles permettent de pallier les limites des systèmes de détection de fraudes actuels basés sur des règles pré-définies, notamment avec des tâches de classification rapides et intelligentes à l'aide des modèles prédictifs.

Des nombreux travaux existent dans le domaine de la finance (Al-Hashedi et Magalingam, 2021) et plus particulièrement dans la détection de fraude par carte de crédit (Adewumi et Akinyelu, 2017).

Nous proposons d'organiser les techniques d'apprentissage automatique en 4 étapes : (1) **L'obtention des données** dans le milieu financier est difficile en raison des politiques de confidentialité des institutions financières. De ce fait, il existe une réelle disparité des jeux de données utilisés dans la littérature : des données publiques [2], synthétiques (Lopez-Rojas et al., 2016) et privées. (2) **L'extraction de caractéristiques** permet d'enrichir le jeu de données afin de distinguer les transactions frauduleuses des transactions légitimes. Dans les travaux liés à la détection de fraude financière, les travaux de Bhattacharyya et al. (2011) et Whitrow et al. (2009) renseignent les caractéristiques à calculer pour représenter le comportement des acteurs. (3) **L'entraînement d'un modèle prédictif** est basé sur un apprentissage supervisé, non supervisé ou semi-supervisé. Dans la DFF, l'apprentissage supervisé a pour but de classifier les transactions dans les classes *légitimes* ou *frauduleuses*. (4) **L'évaluation du modèle** s'effectue avec des mesures classiques de *précision*, *rappel* et *f1-score* (F1).

1. https ://www.swift.com/
2. https ://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

Ces différentes étapes nous permettent d'avoir une vue sur les techniques d'apprentissage automatique utilisées au sein de la DFF. Nous proposons, ainsi, une méthodologie pour entraîner un modèle sur des transactions SWIFT en plusieurs étapes : (1) la définition de nouvelles caractéristiques basées sur les spécificités des transactions SWIFT, notamment sur leur dimension internationales (nombreux pays et devises) et interbancaires (présence d'intermédiaires dans le circuit de la transaction), (2) le choix du meilleur algorithme pour notre jeu de données et (3) une évaluation basée sur une mesure de risque du coût afin de minimiser le coût de notre modèle.

Après avoir enrichi notre jeu de données, les résultats de nos expérimentations nous montrent que l'algorithme *XGBoost* est le plus adapté à nos données en obtenant le meilleur F1 (0.78). Nous avons également remarqué que le comportement des acteurs et leurs interactions sont importants pour la détection de fraude. Les caractéristiques sur les pays sont moins impactants, mais les meilleurs résultats sont obtenus quand elles sont combinées avec celles des acteurs. En effet, les algorithmes d'apprentissage ensembliste ont prouvé leur efficacité sur des jeux de données déséquilibrés.

Une partie de notre contribution est la définition d'une formule de risque de coût, calculé à partir des prédictions de notre modèle. Une prédiction est associée à un coût pour une institution financière, par exemple une transaction prédite comme frauduleuse à un coût qui représente le coût d'un expert pour analyser une transaction.

Pour minimiser le risque de coût de notre modèle, nous avons choisi le seuil de probabilité à partir duquel une transaction est considérée comme frauduleuse par le modèle. Par défaut, le seuil est de 0.5 avec un risque de coût de 29285 euros. Avec nos expérimentations, nous avons fixé le seuil à 0.45 associé à un risque de cout de 28690 euros tout en gardant le même F1. Pour nos travaux futurs, nous souhaitons étudier la manière dont nous pouvons identifier des types de fraudes dans un jeu de données de transactions frauduleuses.

Références

Adewumi, A. O. et A. A. Akinyelu (2017). A survey of machine-learning and nature-inspired based credit card fraud detection techniques. *International Journal of System Assurance Engineering and Management 8*(2), 937–953.

Al-Hashedi, K. G. et P. Magalingam (2021). Financial fraud detection applying data mining techniques : A comprehensive review from 2009 to 2019. *Computer Science Review 40*, 100402.

Bhattacharyya, S., S. Jha, K. Tharakunnel, et J. C. Westland (2011). Data mining for credit card fraud : A comparative study. *Decision support systems 50*(3), 602–613.

Knobel, A. (2019). Swift data can be a global vantage point for tackling global money laundering.

Lopez-Rojas, E., A. Elmir, et S. Axelsson (2016). Paysim : A financial mobile money simulator for fraud detection. In *28th European Modeling and Simulation Symposium, EMSS, Larnaca*, pp. 249–255. Dime University of Genoa.

Whitrow, C., D. J. Hand, P. Juszczak, D. Weston, et N. M. Adams (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data mining and knowledge discovery 18*(1), 30–55.