ST-FRAUD : un outil de détection de transactions frauduleuses

Hamza Chergui [1,2], **Lylia Abrouk** [1], **Nadine Cullot** [1], **Nicolas Cabioch** [2]

1. *Laboratoire d'Informatique de Bourgogne - EA 7534*
   Université de Bourgogne
   hamza.chergui,lylia.abrouk,nadine.cullot@u-bourgogne.fr
2. *SKAIZen Group*
   hchergui, ncabioch@skaizengroup.fr

RÉSUMÉ. *ST-Fraud est un outil de détection de transactions frauduleuses utilisant les techniques d'apprentissage automatique. Il est composé de trois moduleS : l'entraînement d'un modèle de classification capable de distinguer les transactions frauduleuses des légitimes, l'identification de types de fraudes pour améliorer le modèle, et le contrôle des transactions détectées frauduleuses avec des informations utiles pour les experts chargés de vérifier ces transactions.*

ABSTRACT. *ST-Fraud is a fraud transaction detection tool that uses machine learning techniques. It consists of three modules: training a classification model capable of distinguishing fraudulent transactions from legitimate ones, identifying types of fraud to improve the model, and monitoring detected fraudulent transactions with useful information for experts responsible for verifying these transactions.*

MOTS-CLÉS : *apprentissage automatique, outil, détection de fraude, finance*

KEYWORDS: *machine learning, tool, fraud detection, finance*

1. Introduction

Ces dernières années, plusieurs travaux s'intéressent à la lutte contre la fraude financière. Les systèmes actuels ne sont pas assez efficaces, nos travaux portent sur l'amélioration de ces systèmes avec des techniques basées sur l'apprentissage automatique (Hamza *et al.*, 2023 ; Chergui *et al.*, 2023) pour les institutions financières faisant parties du réseau SWIFT.
SWIFT est un réseau international connectant plus de 11 000 organisations bancaires réparties dans près de 200 pays à l'aide de transactions. Les systèmes de détection de fraude actuels sont basés sur des règles prédéfinies, ils détectent un faible nombre de transactions frauduleuses et génèrent un nombre élevé de faux positifs (transactions

légitimes détectées frauduleuses). Les fraudeurs adaptent constamment leur manière de frauder afin de contourner ces règles. Pour renforcer ces systèmes, les techniques basées sur l'apprentissage automatique sont un réel atout : *"elles offrent la puissance numérique et la flexibilité fonctionnelle nécessaires pour identifier des modèles complexes"* (Dixon *et al.*, 2020).
Dans cet article, nous présentons notre outil ST-Fraud conçu pour détecter les transactions frauduleuses et faciliter leurs vérifications par les experts chargés de contrôler les transactions bloquées par les systèmes de lutte contre la fraude. Ce travail s'inscrit dans le cadre d'une collaboration avec l'entreprise SKAIZen Group sur des transactions du réseau SWIFT.

2. L'outil ST-Fraud

ST-Fraud est une application composée de trois modules :

1. L'import de données, le calcul de caractéristiques et l'entraînement de modèles
2. L'identification des types de fraudes et l'amélioration du modèle
3. L'analyse des transactions frauduleuses

2.1. Entraînement du modèle de classification :

Le premier module (figure 2.1) permet d'entraîner un modèle de classification binaire. Pour cela, un jeu de données labellisé importé est séparé en un jeu d'entraînement et un jeu de test. Les caractéristiques nécessaires pour distinguer les transactions légitimes des frauduleuses sont calculées sur des périodes prédéfinies. Ensuite, un algorithme de classification est sélectionné pour entraîner le modèle. Sa performance est évaluée avec le f1-score : une mesure de moyenne pondérée entre le rappel et la précision d'un modèle.

2.2. Analyse des clusters de transactions frauduleuses

Le deuxième module (figure 2.2) est dédié à l'identification des types de fraudes du jeu de données. Nous formons un sous-jeu de données composées des transactions prédites frauduleuses par le modèle entraîné dans la section 2.1. Ensuite, nous réalisons un clustering de ces transactions avec l'algorithme k-means. Puis, nous analysons les clusters à l'aide du modèle de classification et le framework SHAP capable d'interpréter ses prédictions. Nous avons choisi deux types de visualisations intégrés dans le framework SHAP: les heatmaps pour une visualisation locale des transactions au sein des clusters, et les beeswarm pour une visualisation globale des transactions entre les clusters. Ces visualisations vont permettre aux experts d'associer un type de fraude aux clusters. Le modèle est ensuite amélioré en un modèle de classification multi-classes capable d'attribuer aux transactions les classes légitimes et frauduleuses. Cela permet aux experts de faciliter leurs enquêtes pour bloquer ou libérer les transactions détectées frauduleuses par le système.

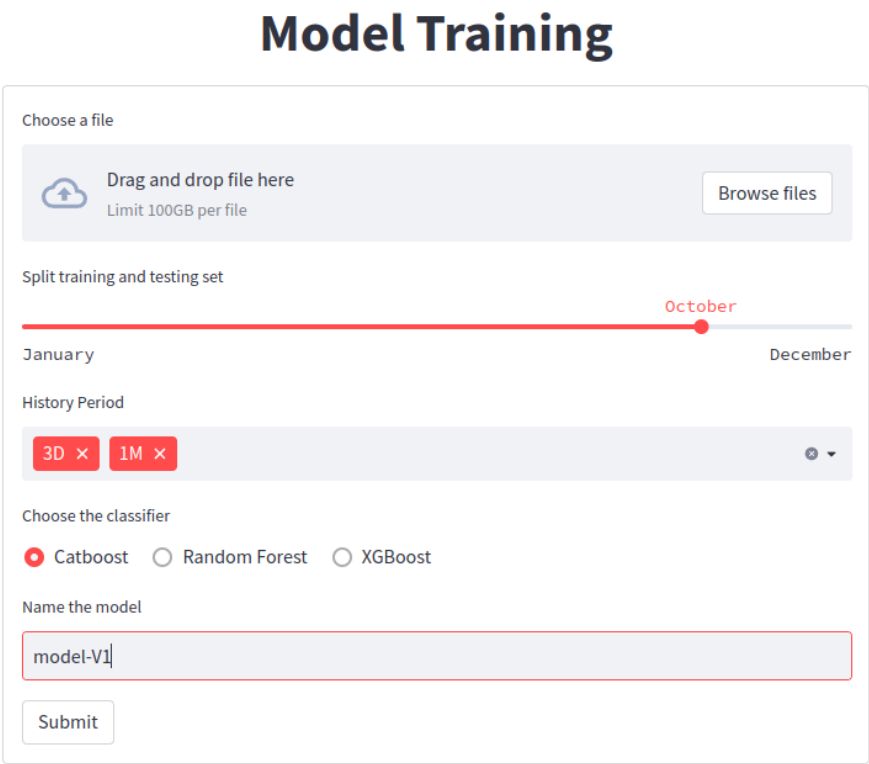

FIGURE 1. *Entraînement d'un modèle*

2.3. Examen des transactions frauduleuses :

Le dernier module de l'outil est réservé aux experts financiers chargés de contrôler les transactions frauduleuses. Nous présentons sur ce module (figure 2.3) les informations relatives à chaque transaction. Nous utilisons le modèle de classification multiclasses pour afficher le pourcentage d'appartenance aux classes légitimes et frauduleuses. Nous avons ajouté des informations non réelles sur les acteurs par manque de données. Un projet de recherche est mené pour la collecte d'informations sur les acteurs à partir de sources externes (DBpedia, articles) (Auger *et al.*, 2022).

Bibliographie

Auger B., Chergui H., Chehade Y., Kadri J. E., Abrouk L., Cabioch N. (2022). Construction d'une ontologie dans le domaine financier pour la détection de fraudes. In *Inforsid'23*, p. 157–162.

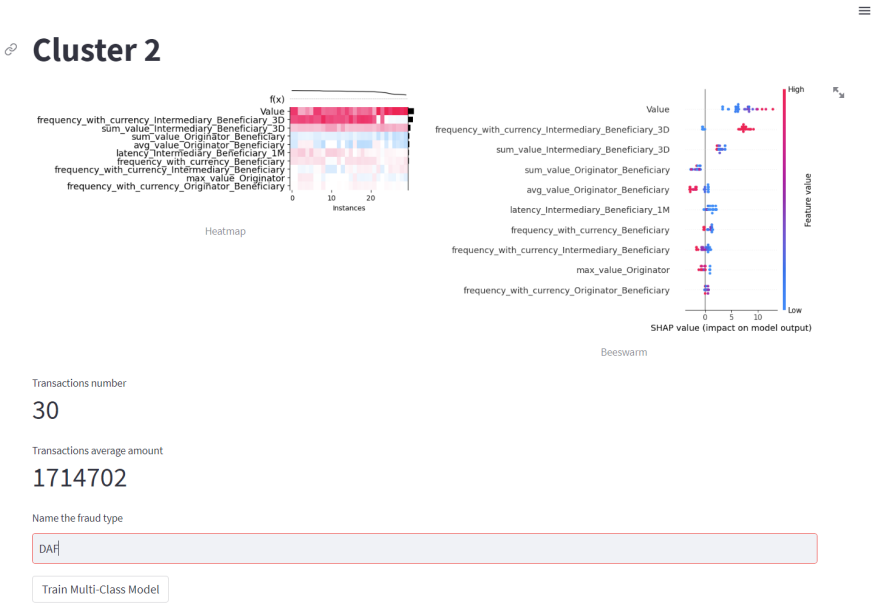

FIGURE 2. *Identification d'un type de fraude*

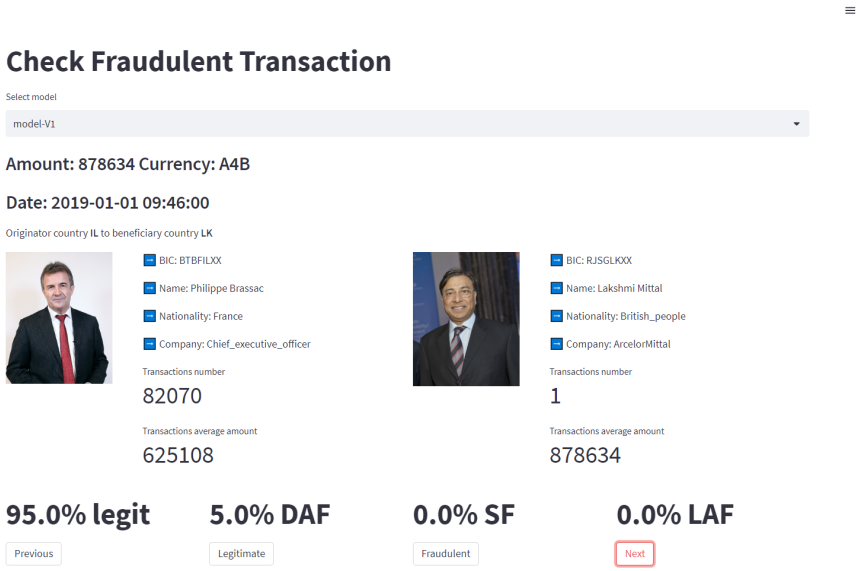

FIGURE 3. *Examen d'une transaction frauduleuse*

Chergui H., Abrouk L., Cullot N., Cabioch N. (2023). Réduction du risque du coût d'un modèle dans la détection de fraude financière. In *Extraction et gestion des connaissances, EGC 2023*, vol. E-39, p. 641–642.

Dixon M., Halperin I., Bilokon P. (2020). *Machine learning in finance: From theory to practice*. Springer International Publishing.

Hamza C., Lylia A., Nadine C., Nicolas C. (2023). Semi-supervised method to detect fraudulent transactions and identify fraud types while minimizing mounting costs. *International Journal of Advanced Computer Science and Applications*, vol. 14, nº 2.