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/

Résumé. La lutte contre la fraude financière est un enjeu majeur pour les institutions financières. Ces dernières années, plusieurs approches basées sur l'analyse des transactions bancaires ont été proposées pour la détéction de fraude. Dans ce travail, nous proposons une approche basée sur les techniques d'apprentissage automatique. Le but est la détection de fraudes financières sur des transactions internationales et interbancaires du réseau SWIFT. Nous entraînons un modèle avec des caractéristiques calculées à partir des spécificités des transactions SWIFT. Nous définissons une mesure de risque du coût sur les prédictions d'un modèle que nous souhaitons réduire avec notre méthodologie. Les expérimentations ont été menées sur un jeu de données réel et validées par les experts du domaine.

1 Introduction

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.
Les flux financiers sont composés de transactions et doivent être analysés par les institutions financières en étudiant les comportements des acteurs impliqués. Une transaction analysée frauduleuse est soit bloquée par le système, soit laissée dans le flux avec une alerte. Ces deux situations nécessitent une analyse manuelle d'un expert pour la gestion de ces transactions. Les systèmes implémentés dans les institutions financières sont basés sur des règles pré-définies, leurs mécanismes présentent une faiblesse exploitée par les fraudeurs : ils identifient ces règles et adaptent leur manière de frauder pour les contourner. Notre travail s'inscrit dans les travaux de recherche en collaboration avec l'entreprise SKAIZen Group qui vise à améliorer la détection de fraude avec des données provenant d'une société, appelé SWIFT [1]. Il s'agit d'une

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

société mettant à disposition un réseau interbancaire proposant différents services comme 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. Ces dernières années, des travaux utilisant des algorithmes d'apprentissage automatique ont été utilisés pour la détection transactions frauduleuses. Elles permettent de pallier les limites des systèmes de détection de fraudes basés sur des règles pré-définies, notamment avec des tâches de classification réalisées avec des modèles prédictifs. À travers notre travail, nous souhaitons présenter les différentes techniques d'apprentissage automatique et observer leurs utilités dans le domaine de la détection de fraude financière (DFF). Nous proposons ensuite une méthodologie pour entraîner un modèle qui sera évalué avec une mesure de risque du coût. Cette dernière est associée aux coûts financiers des prédictions du modèle et nous permettra de choisir un seuil de probabilité à partir duquel nous considérons une transaction comme frauduleuse. La suite de l'article est organisée de la manière suivante : dans la section 2, nous dressons un état de l'art des techniques d'apprentissage automatique dans le domaine des fraudes financières. Dans la section 3, nous présentons notre approche que nous validons avec des expérimentations sur un jeu de données réel dans la section 4. Enfin, nous concluons et abordons nos perspectives de recherche dans la section 5.

2 Travaux liés

Les méthodes basées sur les techniques d'apprentissage automatique pour la détection de fraudes peuvent être un réel atout pour les institutions financières grâce à leurs prédictions rapides et intelligentes. De 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 de présenter les techniques d'apprentissage automatique en plusieurs étapes (Chergui et al., 2022) :

L'obtention des données dans le milieu financier est difficile en raison de la politique de confidentialité des institutions financières. Il existe une réelle disparité des jeux de données utilisés dans la littérature. Nous trouvons des jeux de données publics [2], synthétiques (Lopez-Rojas et al., 2016) et privés. Une telle disparité rend difficile la comparaison des approches qui utilisent des schémas de données ainsi que des types de fraude différents. De plus, il est souvent difficile de reproduire les approches, car les algorithmes ne sont pas partagés.

L'extraction de caractéristiques permet d'enrichir le jeu de données afin de distinguer les transactions frauduleuses des transactions légitimes. Dans la littérature de la DFF, les travaux de Whitrow et al. (2009) et Bhattacharyya et al. (2011) conventionnent les caractéristiques à calculer, notamment en s'intéressant aux volumes et aux moyennes des montants des transactions réalisés par les acteurs sur différentes temporalités. En outre, les jeux de données comportant des fraudes sont déséquilibrés, il existe un nombre plus élevé de transactions légitimes que de transactions frauduleuses. Le pourcentage de transactions frauduleuses représente moins d'1% sur ces jeux de données. Ce déséquilibre peut impacter négativement l'apprentis-

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

sage. Ainsi, des techniques *d'over/under sampling* ont été développées pour soit augmenter le pourcentage de la classe minoritaire, soit de diminuer le pourcentage de la classe majoritaire.

L'entraînement d'un modèle prédictif est basé soit 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*. Des travaux existent sur la comparaison de différents algorithmes (ex : Random Forest, SVM, Naive Bayes, ...) comme ceux de Varmedja et al. (2019) et Khatri et al. (2020). Les algorithmes d'apprentissage non supervisé regroupent les données dans des clusters, la détection de fraude peut s'opérer de deux manières : soit avec des clusters jugés frauduleux, où toutes les transactions appartenant à ces clusters seront considérées comme frauduleuses (Le Khac et Kechadi, 2010) ; soit avec des algorithmes tels que *Isolation Forest* (Liu et al., 2008) ou *Local Outlier Factor* (Breunig et al., 2000) permettant d'identifier des anomalies au sein des clusters. Ces algorithmes ont montré leur efficacité auprès des travaux de John et Naaz (2019) ou Mishra et Chawla (2019). En outre, les techniques basées sur les réseaux de neurones ne sont pas performantes sur des données tabulaires (Borisov et al., 2021), ces techniques ne sont pas très présentes dans la littérature.

L'évaluation du modèle s'opère avec des mesures classiques de *précision*, *rappel* et *f1-score* (F1). Il y a également l'*AUC-PR* (aire sous la courbe de précision et de rappel), une mesure utilisée avec des jeux de données déséquilibrés qu'on retrouve dans des travaux de la DFF (Bahnsen et al., 2013). Des mesures de risque du coût ont vu le jour dans le domaine de la DFF pour calculer le coût des prédictions des modèles lors de l'évaluation. Ce risque du coût représente l'argent que l'institution financière risque du perdre avec le modèle.

Ces différentes étapes nous permettent d'avoir une vue sur les techniques d'apprentissage automatique utilisées au sein de la DFF. Nous proposons, dans la suite, une méthodologie pour entraîner un modèle sur des transactions SWIFT en plusieurs étapes : (1) la définition de nouvelles caractéristiques, (2) une méthode d'apprentissage pour entraîner plus rapidement un modèle et (3) une évaluation basée sur une mesure de risque du coût.

3 Méthodologie

Notre méthodologie présentée sur la figure 1 est composée de trois parties : dans la première, nous enrichissons notre jeu de données avec deux types de caractéristiques. Dans la deuxième partie, nous divisons notre jeu de données et entraînons des modèles prédictifs sur chaque partie du jeu de données. Enfin, dans la troisième partie, nous introduisons les mesures utilisées pour l'évaluation de ces modèles avec une mesure de risque du coût utilisée pour choisir le seuil de probabilité à partir duquel une transaction est frauduleuse.

3.1 Données et caractéristiques

Les attributs d'une transaction SWIFT et un label indiquant si la transaction est frauduleuse (Vrai) ou légitime (Faux) sont présentés dans le tableau 1. Il y a 3 acteurs : l'émetteur, l'intermédiaire et le bénéficiaire. La transaction est un transfert d'argent entre l'émetteur et le bénéficiaire, un intermédiaire intervient dans la transaction si l'émetteur et le bénéficiaire ne possèdent pas de relation d'échange établie. Les acteurs sont identifiés à travers un code appelé 'BIC' dont nous pouvons extraire le code du pays de l'acteur avec le quatrième et cinquième

Réduction du risque du coût d'un modèle

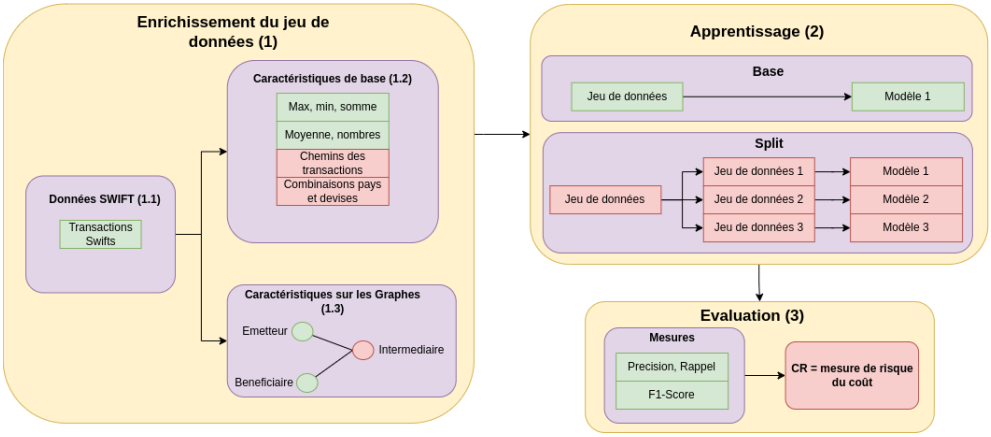


FIG. 1 – Schéma de la méthodologie.

caractère (ex : FR). Le montant représente la valeur de la transaction. D'une manière plus formelle, nous pouvons représenter un jeu de données J avec un ensemble de transactions t^i tel que $i \in [1, n]$ et n son nombre de transactions.

TAB. 1 – Exemples de transactions SWIFT avec un label

Émetteur	Intermédiaire	Bénéficiaire	Date	Devise	Montant	Label
BIC0FR01	BIC0IT01	BIC0FR02	210625	EUR	15006	Faux
BIC0US03	-	BIC0GB01	210625	GBP	33065	Faux
BIC0FR04	BIC0FR06	BIC0FR05	210626	EUR	100325	Vrai

Calcul des caractéristiques de base : à partir d'un jeu de données J, pour chaque acteur et pays, nous calculons des caractéristiques de base présentées dans le tableau 2, que nous ajoutons aux transactions. Ces caractéristiques nous permettent de modéliser leurs comportements (Bhattacharyya et al., 2011). Ainsi, pour chaque transaction t^i, nous calculons 5 caractéristiques sur l'agrégation des historiques de transaction des acteurs ou pays. Dans la suite de cet article, les caractéristiques des acteurs auront pour code $base\_acteurs$ et celles des pays $base\_pays$, ces codes seront utilisés lors de la partie expérimentations.

TAB. 2 – Caractéristiques de base

Caractéristiques	**Formalisations**
Montant maximum	$e_{max} = \underset{1<=i<=n}{max}\ t^i_{montant}$
Montant minimum	$e_{min} = \underset{1<=i<=n}{min}\ t^i_{montant}$
Somme des montants	$e_{sum} = \sum_{i=1}^{n} t^i_{montant}$
Nombre de transactions	$e_{count} = n$
Moyenne des montants	$e_{avg} = e_{sum}/e_{count}$

Calcul des caractéristiques basées sur les graphes : à partir des transactions d'un jeu de données, nous construisons deux graphes orientés et pondérés : le graphe ($G_{acteurs}$) possède comme nœuds les acteurs (émetteurs et bénéficiaires), comme arcs les transactions des émetteurs vers les bénéficiaires et comme poids des arcs le nombre de transactions réalisées entre eux. Le second graphe (G_{pays}) possède comme nœuds les pays (pays émetteurs et pays bénéficiaires), comme arcs les transactions entre les pays émetteurs et pays bénéficiaires et comme poids des arcs le nombre de transactions réalisées entre eux.

À partir de ces deux graphes, nous obtenons deux matrices d'adjacence : $A_{acteurs}$ de dimension $N \times N$ avec N le nombre d'acteurs du jeu de données J ; et A_{pays} de dimension $M \times M$ avec M le nombre de pays de J. Nous construisons également 6 vecteurs, 3 vecteurs de dimensions N avec $E_{acteurs}$ la moyenne des montants des transactions des acteurs, $E^{+}_{acteurs}$ étant la moyenne des montants des transactions des clients en tant qu'émetteur, et $E^{-}_{acteurs}$ étant la moyenne des montants des clients en tant que bénéficiaire. Nous construisons de la même façon les vecteurs E_{pays}, E^{+}_{pays} et E^{-}_{pays} de dimensions M.

À partir des matrices d'adjacences A et des vecteurs E, E^{+} et $E-$, nous calculons les 3 caractéristiques présentées dans le tableau 3, que nous ajoutons aux transactions. Ces caractéristiques permettent de modéliser le comportement des acteurs et pays sur : leur voisinage entier (e_g), leur voisinage en tant qu'émetteur ou pays émetteur (e_{g+}), et de leur voisinage en tant que bénéficiaire ou pays bénéficiaire (e_{g+}) (Huang et al., 2018). Dans la suite de cet article, les caractéristiques des acteurs auront pour code $graphe\_acteurs$ et celles des pays $graphe\_pays$. Ces codes seront utilisés lors de la partie expérimentations.

TAB. 3 – Caractéristiques de graphe.

Caractéristiques	**Formalisations**	
	Acteurs	Pays
Voisinage entier	$e_g = A_{acteurs}.E_{acteurs}$	$e_g = A_{pays}.E_{pays}$
Voisinage émetteur	$e_{g+} = A_{acteurs}.E^{+}_{acteurs}$	$e_{g+} = A_{pays}.E^{+}_{pays}$
Voisinage bénéficiaire	$e_{g-} = A_{acteurs}.E^{-}_{acteurs}$	$e_{g-} = A_{pays}.E^{-}_{pays}$

3.2 Apprentissage des modèles de classification

Les modèles de classification sont entraînés avec des algorithmes d'apprentissage supervisé par le biais d'un jeu de données labellisé. Nous effectuons une tâche de classification avec deux classes *fraude* et *légitime*. L'entraînement d'un modèle peut être long en fonction du nombre de données et de leur dimension. À travers notre méthodologie, nous souhaitons diminuer ce temps tout en conservant la qualité de prédiction de notre modèle. Pour réaliser cela, nous entrainons un modèle prédictif ($modele\_split$) en divisant notre jeu de données afin de réduire les temps d'entraînement et de calcul. Nous évaluons l'efficacité de ce modèle en le comparant à un modèle de base entraîné sur notre jeu de données non divisé ($modele\_base$).

Division du jeu de données : En nous basant sur l'hypothèse de nos experts nous disant que les transactions qui partagent les mêmes devises et pays ont des comportements de fraudes similaires. Nous avons extrait, du jeu de données J, les tuples T^j (devise, pays émetteur et

pays bénéficiaire) avec $j \in [1, N]$, N étant le nombre de combinaisons de tuple du jeu de données. Pour chaque tuple, nous calculons le nombre de transactions T_{count} et le montant moyen des transactions T_{avg}. En utilisant les techniques d'apprentissages non supervisés (k-means), et ces deux caractéristiques, nous avons regroupé les tuples dans k clusters avec des valeurs de T_{count} et T_{avg} proches. Le nombre k est choisi lors des expérimentations. Les transactions avec des tuples T dans les mêmes clusters ont été placés dans les mêmes jeux de données. Ainsi, nous avons divisé notre jeu de données J en k jeux de données J^p avec $p \in [1F, k]$.

Entraînement des modèles : Le $modele\_split$ est un modèle composé de k modèles entraînés sur chaque division du jeu de données. Le $modele\_base$ entraîné sur le jeu de données complet aura plus de données pour son apprentissage. Les modèles de $modele\_split$ possèdent moins de données pour leur apprentissage, cependant leurs données sont plus homogènes, car elles partagent des combinaisons de devises et pays avec des comportements similaires au niveau du nombre (T_{count}) et moyenne des montants des transactions (T_{avg}).

3.3 Évaluation des modèles

TAB. 4 – Matrice de confusion.

		Réalité	
		Fraude	Légitime
Prédiction	Fraude	VP	FP
	Légitime	FN	VN

Pour évaluer les modèles, nous utilisons la matrice de confusion présentée dans le tableau 4. Cette matrice nous donne le nombre de transactions frauduleuses correctement prédites VP, le nombre de transactions frauduleuses incorrectement prédites FP, le nombre de transactions légitimes correctement prédites VN et le nombre de transactions légitimes incorrectement prédites FN. À partir de cette matrice, nous pouvons calculer les mesures suivantes :

$$Precision = \frac{VP}{VP + FP}; \quad Rappel = \frac{VP}{VP + FN}; \quad F1 = \frac{2 * VP}{2 * VP + FP + FN} \quad (1)$$

Lorsqu'un modèle est entraîné, sa prédiction pour une transaction est une probabilité d'appartenance à une classe (frauduleuse, légitime). Par défaut, on attribue à la transaction la classe avec la plus grande probabilité. Dans le cas d'une classification à deux classes, nous pouvons faire varier le seuil de probabilité à partir duquel une transaction est frauduleuse. Ainsi pour chaque seuil, nous pouvons calculer la précision et le rappel du modèle pour tracer une courbe.

L'aire sous courbe de précision et de rappel (AUC-PR) est une mesure utilisée sur les modèles entraînés avec des jeux de données déséquilibrés. Dans le domaine financier, un modèle ayant une précision élevée et un rappel faible est un modèle dont les transactions prédites frauduleuses sont réellement frauduleuses avec peu de fausses alertes, allégeant ainsi le coût du travail des experts. Cependant, un rappel faible indique un faible nombre de transactions frauduleuses détectées, pouvant ainsi exposer les institutions financières à des sanctions par les

régulateurs.

Le coût du risque des prédictions : Bahnsen et al. (2013) ont proposé une matrice de coût du risque des prédictions d'un modèle pour le domaine de la détection de fraude par carte de crédit qui est représenté dans le tableau 5.

TAB. 5 – Matrice de risque du coût de Bahnsen et al. (2013).

		Réalité (t_i)	
		Fraude	Légitime
Prédiction (t_i)	Fraude	C_a	C_a
	Légitime	Amt_i	0

Cette matrice représente le coût d'une prédiction d'une transaction t_i pour une institution financière. D'une part, si la transaction est prédite comme frauduleuse, elle a un coût d'administration C_a représentant l'estimation du coût d'un expert pour l'analyse d'une transaction. D'autre part, si la transaction est prédite légitime alors qu'elle est frauduleuse, le coût est égal au montant (Amt_i) de la transaction.
Cette matrice de coût peut être critiquée sur deux aspects : (i) le coût d'une transaction légitime prédite frauduleuse peut avoir un coût supérieur, si un client voit sa transaction bloquée injustement, cela peut impacter la qualité du service du client et résulter en un coût supérieur ; (ii) l'institution financière ne perd pas le montant d'une transaction frauduleuse prédite légitime. En réalité, elle a un risque plus élevé d'être soumise à une sanction financière par les régulateurs du monde financier. Pour ces deux raisons et pour répondre aux besoins de nos experts, nous proposons une matrice de coût présentée dans le tableau 6.

TAB. 6 – Matrice de risque du coût proposée.

		Réalité (t_i)	
		Fraude	Légitime
Prédiction (t_i)	Fraude	C_a	$C_a + C_c$
	Légitime	C_s	0

C_a représente toujours le coût administratif de l'analyse d'une transaction. Nous introduisons C_c qui est ajouté au coût administratif dans le cas d'une transaction légitime classée frauduleuse, ce coût est lié à l'insatisfaction du client. C_s, dans le cas d'une transaction frauduleuse classée légitime, représente l'évaluation du risque d'être soumis à une sanction. À partir de la matrice proposée, nous définissons la formule de calcul de risque du coût suivante :

$$RC = VP * C_a + FP * (C_a + C_c) + FN * C_s \tag{2}$$

Les coûts C_a, C_c et C_s sont à définir avec des experts en fonction de l'évaluation des risques financiers des coûts.
Dans la suite, nous présentons nos expérimentations, dans lesquelles nous allons : (1) diviser notre jeu de données et calculer les caractéristiques de bases et graphes, (2) entraîner des modèles pour chaque jeu de données, en prenant en compte l'impact des caractéristiques, (3)

TAB. 7 – Jeux de données.

J^p	**Nombre**	**Moyenne des montants**
J	1321125	284624
J^0	331845	278739
J^1	689392	284120
J^2	94377	282366

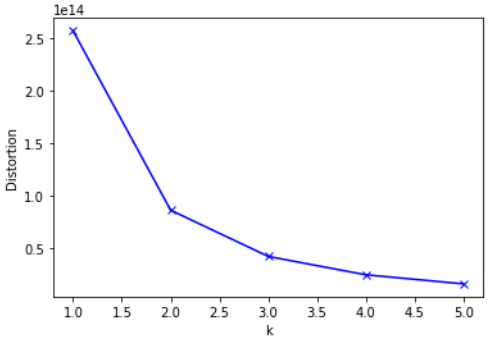


FIG. 2 – Méthode du coude.

choisir le seuil de réduction du risque du coût de notre modèle avec la formule proposée et (4) discuter de nos résultats.

4 Expérimentation

Nous avons réalisé nos expérimentations avec un jeu de données labelisé obtenu grâce à une collaboration avec l'entreprise SKAIZen Group. Ce jeu de données contient 1321125 transactions, 10198 acteurs, 151 devises et 248 pays. 7942 transactions sont labelisées frauduleuses. Les transactions sont réparties sur l'année 2019, nous avons utilisé les transactions de janvier à octobre pour l'entraînement de notre modèle et les transactions de novembre à décembre pour l'évaluation. Nous avons réalisé nos expérimentations en Python avec une machine possédant 16Go de RAM sur un CPU 7-10750H. Nous avons utilisé la librairie Scikit-Learn pour l'apprentissage des modèles.

4.1 Division du jeu de données et calcul des caractéristiques

En analysant le jeu de données, nous avons obtenu 3318 tuples uniques dispersées dans les 1321125 transactions. Pour chaque tuple T nous avons calculé T_{count} et T_{avg} sur lesquels nous avons réalisé un *clustering* avec l'algorithme *k-means*. Pour choisir le nombre de clusters, nous avons utilisé la méthode du coude (*elbow*) permettant d'identifier le nombre de clusters optimal pour notre jeu de données. La figure 2 illustre cette méthode qui calcule pour chaque k la distorsion. Cette dernière correspond à la somme des distances au carré de chaque point avec son centroïde. On définit la valeur de k où la courbe s'infléchit, dans notre cas le k optimal est égal à 3.

Ainsi, nous séparons notre jeu de données J en 3 jeux de données J^0, J^1 et J^2 dont nous présentons le nombre et la moyenne des montants de leurs transactions dans le tableau 7. Pour chaque jeu de données, nous calculons les caractéristiques présentées dans la section 3 : $base\_acteurs$ et $graphe\_acteurs$ pour les émetteurs et bénéficiaires et $base\_pays$ et $graphe\_pays$ pour les pays émetteurs et bénéficiaires. Nous avons ainsi ajouté 32 caractéristiques à chaque transaction en fonction de leurs acteurs et pays.

4.2 Entraînement des modèles

Pour choisir le modèle le plus adapté à notre structure de données et à nos caractéristiques, nous comparons différents algorithmes d'apprentissage supervisé. Les algorithmes les plus performants dans le domaine de la détection de fraude financière sont Random Forest (Xuan et al., 2018), XGBoost (Meng et al., 2020) et CatBoost (Alfaiz et Fati, 2022). Pour chaque algorithme, nous entraînerons deux modèles ($model\_base$ et $model\_split$) que nous évaluons avec les mesures présentées dans la section 3. Pour la mesure du risque du coût RC, nous avons défini, le coût d'administration $C_a = 5$, le coût d'une fausse alerte $C_c = 10$ et le coût d'une transaction frauduleuse prédite légitime $C_p = 20$. Les résultats et le temps d'entraînement (TE) des modèles sont présentés dans le tableau 8. Le choix des hyper-paramètres des algorithmes s'est réalisé avec une recherche exhaustive des paramètres indiqués dans le tableau 9. Pour identifier l'impact des caractéristiques calculées, nous avons conservé le meilleur modèle *XGBoost* qui a obtenu le meilleur F1 (0.78) et AUC-PR(0.66). Nous l'avons entraîné avec les 4 types de caractéristiques pour étudier leur influence sur l'apprentissage. Les résultats sont présentés dans le tableau 10, où nous remarquons une influence importante des caractéristiques liées aux acteurs.

TAB. 8 – Tableau récapitulatif des résultats des algorithmes.

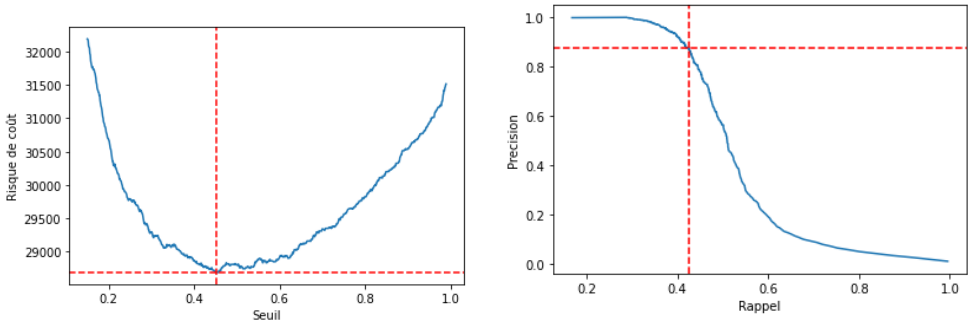
fieur		**Précision**	**Rappel**	**F1**	**AUC-PR**	**RC**	**TE**
RF	BASE	0.98	0.67	0.75	0.65	30170	2mn 57sec
	SPLIT	0.97	0.68	0.76	0.65	30050	2mn
XGBoost	BASE	0.95	0.71	0.78	0.66	28780	1mn 54sec
	SPLIT	0.93	0.70	0.78	0.64	29285	1min 35sec
CatBoost	BASE	0.99	0.66	0.74	0.66	30645	25sec
	SPLIT	0.99	0.66	0.74	0.65	30630	28sec

TAB. 9 – Recherches aléatoires des hyper-paramètres.

Random Forest		**CatBoost**		**XGBoost**	
arbres	5,10,100	*learning rate*	[0.01,0.05,0.1]	*learning rate*	[0.01,0.05,0.1]
profondeur	5,10,None	profondeur	5,10,None	profondeur	5,10,None
critere	gini, entropie	iterations	10,50,100	arbres	5,10,100

TAB. 10 – Tableau comparatif des caractéristiques.

Caractéristiques	**Précision**	**Rappel**	**F1**	**AUR PR**	**RC**	**TE**
$base\_acteurs$	0.94	0.69	0.77	0.64	29645	58sec
$base\_pays$	0.78	0.53	0.55	0.31	39490	1mn 1sec
$graphe\_acteurs$	0.95	0.69	0.76	0.64	29850	55sec
$graphe\_pays$	0.80	0.52	0.54	0.32	39585	1mn 1sec



(a) Risque du coût. (b) Courbe entre la précision et le rappel.

FIG. 3 – Définition du seuil de probabilité d'une transaction frauduleuse.

4.3 Risque du coût de prédiction

Après avoir observé l'impact des caractéristiques, le risque du coût de notre $model\_split$ est minimal avec l'algorithme *XGBoost* de 29280, pour rappel, le seuil probabilité à partir duquel une transaction est frauduleuse est fixé à 0.5 par défaut. À partir des prédictions du modèle et des probabilités d'appartenance à la classe frauduleuse, nous calculons le risque du coût pour chaque seuil de probabilité entre 0 et 1 avec un intervalle de 0.001. Les résultats sont présentés sur la figure 3a. Le coût minimal est obtenu lorsque la probabilité est de 0.45, et le CR associé est égal à 28690. La figure 3b nous indique que le rappel et la précision du modèle, associé à une probabilité de 0.45 pour considérer une transaction frauduleuse, sont respectivement de 0.93 et 0.71 qui nous donne un F1 de 0.78. Nous avons donc réduit le seuil de probabilité à 0.45 pour réduire le risque du coût de notre modèle, tout en gardant la même qualité de prédiction.

4.4 Résultats et discussions

Les résultats de nos expérimentations nous montrent que l'algorithme *XGBoost* est le plus adapté à notre structure de données, car il obtient le meilleur F1 (0.78) et AUC-PR(0.66). Les $model\_split$ et $model\_base$ possèdent sensiblement les mêmes résultats comme le montre le tableau 8 avec une différence de temps de 29 secondes pour l'apprentissage en faveur du $model\_split$, grâce à la division du jeu de données. Le tableau 10, indique le fort impact des caractéristiques : $base\_acteurs$ et $graph\_acteurs$ sur l'apprentissage. Le comportement des acteurs et leurs interactions sont importants pour la détection de fraude. Les caractéristiques sur les pays sont moins impactantes, mais les meilleurs résultats sont obtenus quand elles sont combinées avec celles des acteurs. Enfin, nous avons minimisé le risque du coût (RC) de notre $model\_split$, à 28690 qui était de base à 29285, grâce au seuil de probabilité fixé à 0.45 pour considérer une transaction frauduleuse, et cela, en conservant le même F1 (0.78).

5 Conclusion

Les techniques d'apprentissage automatique peuvent être de réels atouts pour les institutions financières afin de lutter contre les fraudes. Un apprentissage adapté, avec le calcul de caractéristiques, amène des résultats satisfaisants. À travers notre méthode, nous avons détaillé les caractéristiques calculées sur des transactions SWIFT. Ensuite, nous avons divisé notre jeu de données selon les devises et pays des transactions et entraîné un modèle pour chaque division. Ce qui nous a permis d'avoir des résultats similaires avec un temps d'entraînement plus court. Nous avons également étudié l'impact des caractéristiques calculées qui nous informe que celles basées sur les acteurs sont importantes pour l'apprentissage. Nous avons proposé une formule de risque du coût d'un modèle et nous avons réduit ce risque en choisissant un nouveau seuil de probabilité à partir duquel une transaction est considérée frauduleuse. Pour nos travaux futurs, nous souhaitons étudier l'impact des intermédiaires sur les transactions frauduleuses. Nous voulons aussi inclure l'aspect temporel des transactions dans le calcul des caractéristiques.

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.

Alfaiz, N. S. et S. M. Fati (2022). Enhanced credit card fraud detection model using machine learning. *Electronics 11*(4), 662.

Bahnsen, A. C., A. Stojanovic, D. Aouada, et B. Ottersten (2013). Cost sensitive credit card fraud detection using bayes minimum risk. In *2013 12th international conference on machine learning and applications*, Volume 1, pp. 333–338. IEEE.

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.

Borisov, V., T. Leemann, K. Seßler, J. Haug, M. Pawelczyk, et G. Kasneci (2021). Deep neural networks and tabular data : A survey. *arXiv preprint arXiv :2110.01889*.

Breunig, M. M., H.-P. Kriegel, R. T. Ng, et J. Sander (2000). Lof : identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pp. 93–104.

Chergui, H., L. Abrouk, N. Cullot, et N. Cabioch (2022). Détection de fraude financière dans un système de transactions interbancaires. *INFORSID'22*, 141–156.

Huang, D., D. Mu, L. Yang, et X. Cai (2018). Codetect : Financial fraud detection with anomaly feature detection. *IEEE Access 6*, 19161–19174.

John, H. et S. Naaz (2019). Credit card fraud detection using local outlier factor and isolation forest. *Int. J. Comput. Sci. Eng 7*(4), 1060–1064.

Khatri, S., A. Arora, et A. P. Agrawal (2020). Supervised machine learning algorithms for credit card fraud detection : a comparison. In *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 680–683. IEEE.

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

Le Khac, N. A. et M.-T. Kechadi (2010). Application of data mining for anti-money laundering detection : A case study. In *2010 IEEE international conference on data mining workshops*, pp. 577–584. IEEE.

Liu, F. T., K. M. Ting, et Z.-H. Zhou (2008). Isolation forest. In *2008 eighth ieee international conference on data mining*, pp. 413–422. IEEE.

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.

Meng, C., L. Zhou, et B. Liu (2020). A case study in credit fraud detection with smote and xgboost. In *Journal of Physics : Conference Series*, Volume 1601, pp. 052016. IOP Publishing.

Mishra, S. et M. Chawla (2019). A comparative study of local outlier factor algorithms for outliers detection in data streams. In *Emerging Technologies in Data Mining and Information Security*, pp. 347–356. Springer.

Varmedja, D., M. Karanovic, S. Sladojevic, M. Arsenovic, et A. Anderla (2019). Credit card fraud detection-machine learning methods. In *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, pp. 1–5. IEEE.

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.

Xuan, S., G. Liu, Z. Li, L. Zheng, S. Wang, et C. Jiang (2018). Random forest for credit card fraud detection. In *2018 IEEE 15th international conference on networking, sensing and control (ICNSC)*, pp. 1–6. IEEE.

Summary

The fight against financial fraud is a major challenge for financial institutions. In recent years, several approaches based on the analysis of banking transactions have been proposed for fraud detection. In this work, we propose an approach based on machine learning techniques for detecting financial fraud in international and interbank transactions of the SWIFT network. The learning of the model is carried out with new characteristics calculated from the specificities of SWIFT transactions. We define a risk measure of the cost on the predictions of a model we wish to reduce with our methodology. The experiments were conducted on a real dataset and validated by experts in the field.