Une approche hybride pour la détection et l'interprétation de transactions frauduleuses dans le réseau SWIFT.

Hamza Chergui

THÈSE DE DOCTORAT

DE L'ÉTABLISSEMENT UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ

PRÉPARÉE A L'UNIVERSITÉ DE BOURGOGNE

École doctorale n° 37

Sciences Pour l'Ingénieur et Microtechnique

Doctorat d'Informatique

Par

CHERGUI Hamza

UNE APPROCHE HYBRIDE POUR LA DÉTECTION ET L'INTERPRÉTATION DE TRANSACTIONS FRAUDULEUSES DANS LE RÉSEAU SWIFT.

Thèse présentée et soutenue à Dijon, le 07 décembre 2023

Composition du Jury :

BRINGAY Sandra	Professeure, Université Paul-Valéry Montpellier 3	Rapporteur
CABANAC Guillaume	Professeur, Université de Toulouse 3 - Paul Sabatier	Rapporteur
ARDUIN Pierre-Emmanuel	Maître de Conférences, Université de Paris-Dauphine	Examinateur
GROSS-AMBLARD David	Professeur, Université de Rennes 1	Examinateur
SALLABERRY Christian	Maître de Conférences HDR, Université UPPA	Examinateur
ABROUK Lylia	Maître de conférences, Université de Bourgogne	Co-directrice de thèse
CULLOT Nadine	Professeure, Université de Bourgogne	Directrice de thèse

SPIM
école doctorale **sciences pour l'ingénieur et microtechniques**

Titre : Une approche hybride pour la détection et l'interprétation de transactions frauduleuses dans le réseau SWIFT.

Mots clés : Apprentissage automatique, Ontologie, Fraudes financières, Classification

Résumé :
Les institutions financières sont confrontées à des activités frauduleuses de la part de leurs clients et doivent mettre en place des systèmes efficaces de détection et de prévention de la fraude. Dans cette lutte contre la fraude, il est essentiel de comprendre les caractéristiques des transactions afin de distinguer les fraudes des transactions légitimes. Nos travaux se concentrent sur l'analyse et la classification de ces transactions au sein du réseau interbancaire et international SWIFT, tout en examinant les schémas de fraude qui peuvent en découler. Pour ce faire, nous proposons une approche hybride qui combine des algorithmes d'apprentissage automatique avec une ontologie développée pour ce domaine, tout en minimisant les coûts financiers liés aux erreurs de classification.
Notre approche se décline en cinq étapes principales : l'analyse des caractéristiques, le regroupement des transactions frauduleuses, la classification des transactions, l'utilisation de l'ontologie pour la surveillance des transactions suspectes et l'évaluation des résultats. Cette approche améliore la détection des transactions suspectes en combinant les techniques d'apprentissage automatique à l'utilisation de l'ontologie pour une meilleure compréhension et interprétation des données. Dans le cadre de ce projet, nous avons développé l'outil ST-Fraud dédié à la détection et à l'interprétation des transactions frauduleuses. Cet outil s'appuie sur notre approche hybride pour renforcer l'efficacité des systèmes de détection de fraude dans le contexte des transactions SWIFT.

Title : A hybrid approach for detection and interpretation of fraudulent transactions in the SWIFT network.

Keywords : Machine Learning, Ontology, Financial fraud, Classification.

Abstract :
Financial institutions face fraudulent activities from their customers and must establish effective fraud detection and prevention systems. In this fight against fraud, it is essential to understand the characteristics of transactions to distinguish fraud from legitimate transactions. Our work focuses on the analysis and classification of these transactions within the interbank and international SWIFT network while examining potential fraud patterns. To achieve this, we propose a hybrid approach that combines machine learning algorithms with an ontology developed for this domain, all while minimizing the financial costs associated with classification errors.
Our approach consists of five main steps: feature analysis, clustering of fraudulent transactions, transaction classification, the use of ontology for monitoring suspicious transactions, and result evaluation. This approach enhances the detection of suspicious transactions by integrating machine learning techniques with the use of ontology for a better understanding and interpretation of the data. As part of this project, we have developed the ST-Fraud tool dedicated to the detection and interpretation of fraudulent transactions. This tool relies on our hybrid approach to strengthen the efficiency of fraud detection systems in the context of SWIFT transactions.

UBFC UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ
Université Bourgogne Franche-Comté
32, avenue de l'Observatoire
25000 Besanon, France

Remerciements

Cette thèse a représenté un parcours long et parsemé de défis. À l'approche de sa conclusion, je prends pleinement conscience de l'importance des soutiens qui m'ont été indispensables tout au long de ces trois années. Ainsi, je tiens à exprimer ma gratitude envers toutes les personnes impliquées dans notre réussite, et j'insiste sur le choix du terme "nous", car cette réussite n'aurait pas été possible sans l'aide de tous ceux qui m'ont apporté leur soutien.

Tout d'abord, je souhaite exprimer ma reconnaissance à Nadine Cullot pour avoir dirigé cette thèse. Ton implication a été décisive, apportant une aide précieuse aux moments clés. J'ai perfectionné mon travail en m'inspirant de ta rigueur, et j'ai pu apprécier beaucoup de moment en ta compagnie au laboratoire. Un grand merci également à Lylia Abrouk pour avoir codirigé cette thèse, qui a été bien plus qu'un soutien au cours de ces trois années. Ton accompagnement, tant sur le plan humain à travers nos échanges réguliers que sur le plan professionnel, a été déterminant. Sans toi, je n'aurais pas développé les compétences nécessaires à la concrétisation de ce projet. Je suis content d'avoir pu travailler avec une personne avec autant de valeur que je respecte et partage. La fin de cette thèse ne marque pas la fin de notre collaboration, car je suis convaincu que j'ai encore beaucoup à apprendre de toi.

En second lieu, mes remerciements vont à l'entreprise SKAIZen Group, en particulier à Nicolas Cabioch, pour la confiance accordée. Lors de mon recrutement en alternance il y a quatre ans, j'étais ravi de constater que je rejoignais une entreprise prônant des valeurs essentielles axées sur le bien-être de tous. Cette entreprise est désormais comme une famille pour moi, ayant grandi en parallèle avec moi. Un merci également à tous les employés de l'entreprise qui ont contribué, de près ou de loin, à créer un environnement de travail idéal dans lequel j'ai pu réaliser cette thèse.

Je souhaite exprimer ma gratitude envers les membres du jury. C'est un honneur d'avoir été évalué par un jury dont j'apprécie grandement la qualité scientifique et la dimension humaine, reflétant parfaitement l'idée que je me fais de la recherche.

Un remerciement particulier à l'ensemble des membres du laboratoire LIB qui m'ont accueilli, en particulier les doctorants de l'équipe Sciences des données : Hamid Ahaggach, Hiba Abou Jamra, Elio Hbeich, Alexis Guyot, Sébastien Guillemin, et Selsébil Benelaj Sghaier. Votre solidarité a été vitale pour moi, et je suis fier de nos moments partagés. Je vous souhaite à tous le succès dans vos projets.

Enfin, un grand merci à mes parents, ma sœur Anissa, et mon frère Bilel pour leur soutien constant. Je suis convaincu que ma passion pour la recherche provient d'eux, et les valeurs de partage et de dépassement de soi que j'ai acquises sont le fruit de notre vie commune. Je tiens également à exprimer ma gratitude envers mon épouse, Kenza, pour son soutien et sa bienveillance qui ont été essentiels pour mener cette thèse à son terme.

En dédiant ce travail à toute ma famille, je souhaite exprimer ma sincère reconnaissance.

Table des matières

Table des figures

Liste des tableaux

Chapitre 1

Introduction

Sommaire

Les institutions financières sont des entreprises dédiées aux transactions financières et monétaires, incluant les dépôts, les retraits, les prêts, les investissements et l'échange de devises. En 1974, le sociologue Immanuel Wallerstein annonçait « *la montée et la future disparition du système capitaliste mondial* » [Wal74]. Son ouvrage intervient dans une période historique où le capitalisme s'impose mondialement comme le système économique permettant le commerce international, l'industrialisation et l'émergence des grandes entreprises.

Dans notre société, les institutions financières jouent un rôle central à diverses échelles. Les particuliers font appel à elles pour leurs achats quotidiens, les entreprises pour mener leurs activités et les gouvernements pour gérer l'économie des pays. De plus, avec l'émergence de nouvelles méthodes de paiement telles que les paiements instantanés et une meilleure accessibilité aux services bancaires via les banques en ligne, le nombre de transactions traitées par les institutions financières ne cesse d'augmenter.

Cependant, Wallerstein nous a averti d'une « future disparition » faisant référence à notre système capitaliste qui sera confronté à des problèmes, tels que les inégalités économiques, les crises financières ou les évolutions technologiques qui pourraient remettre en cause les fondements du capitalisme. De nos jours, les institutions financières font également face à des problèmes liés à la gestion des flux, à l'adaptation à de nouvelles normes (ISO20022 [1]) ou encore la fraude financière.

L'encyclopédie *Britannica* [oEB23] définit la fraude comme « la représentation délibérée d'un fait dans le but de priver quelqu'un d'une possession précieuse ». La fraude est définie de manière abstraite, car elle représente un concept vaste et il est difficile de délimiter précisément ses frontières. La fraude intervient dans de multiples domaines tels que la fraude scientifique [Cab22] avec des publications contenant des actes non éthiques tels que le plagiat ou des données fabriquées. De même, la fraude agricole englobe des pratiques illégales telles que la vente de produits périmés ou la manipulation des normes de sécurité [Man16].

1.1 Contexte

Dans le domaine de la finance, la fraude englobe les actes frauduleux commis dans le secteur financier, tels que la manipulation des comptes, la falsification de documents financiers, le détournement de fonds, l'escroquerie ainsi que le blanchiment d'argent. Son objectif est d'obtenir illégalement des avantages financiers en trompant les investisseurs, les actionnaires ou les institutions financières. La fraude financière peut entraîner des conséquences graves, notamment d'importantes pertes financières, la ruine des investisseurs et la déstabilisation des marchés. Les autorités réglementaires et les organismes de réglementation financière s'emploient à détecter et à prévenir la fraude financière, tout en imposant des sanctions sévères aux contrevenants pour dissuader ces pratiques frauduleuses.

Les institutions financières, souvent impliquées dans des activités de fraude, se retrouvent dans une position délicate, agissant comme intermédiaires pour faire circuler des transactions liées à des activités frauduleuses. Les institutions sont tenues de lutter contre la fraude financière en

1. https://www.iso20022.org/

mettant en place des systèmes de détection et de prévention de la fraude. Cependant, les systèmes actuels montrent des limites en termes d'efficacité, car seulement un faible pourcentage de transactions frauduleuses est détecté. Selon Europol [Kno20], l'agence européenne spécialisée dans la répression de la criminalité, 98,9% de ces transactions frauduleuses échappent aux mesures de contrôle.

Pour expliquer ces difficultés, Jensen [Jen97] met en avant plusieurs facteurs pour expliquer ces difficultés liées à la surveillance des transactions financières. Les transactions frauduleuses sont peu nombreuses par rapport au grand nombre de transactions légitimes, elles sont dissimulées parmi des millions de transactions, ce qui rend leur recherche plus complexe. Il est difficile de faire la distinction entre une transaction légitime et une transaction illégitime. En outre, les criminels adoptent des comportements intelligents en s'adaptant aux nouvelles règles mises en place par les institutions financières. Ils développent en permanence de nouvelles stratégies sophistiquées, rendant leur identification encore plus délicate.

Actuellement, les institutions financières s'appuient principalement sur des systèmes de règles pour lutter contre la fraude financière [CTN+18]. Toutefois, ces systèmes présentent des limites importantes. Les règles sont écrites manuellement par des experts qui doivent fixer des seuils sur les montants de transactions, sur le nombre de transactions autorisées par un client ou vérifier s'il se trouve dans un pays à risque. Lorsqu'une transaction déclenche une règle, une alerte est générée, ce qui nécessite alors un contrôle manuel par un expert pour libérer ou bloquer la transaction. Les systèmes de règles présentent l'avantage de pouvoir être robustes aux grands volumes de données, mais ont l'inconvénient d'être statiques. Il devient de plus en plus facile pour les fraudeurs d'identifier ces règles, et d'adopter alors de nouveaux comportements qui rendent difficile la détection de leurs transactions liées à des activités frauduleuses.

Il n'existe pas de manière absolue pour décrire une transaction frauduleuse et une transaction légitime. Pour évaluer si une transaction est frauduleuse, il est nécessaire qu'un expert examine une transaction bloquée par le système et mène une enquête sur les différentes parties impliquées. Une fois cette étape réalisée, l'expert peut déterminer si la transaction est frauduleuse ou légitime. Deux transactions peuvent avoir les mêmes valeurs en termes de date et de montant, mais selon le contexte dans lequel elles ont été réalisées, l'une peut être frauduleuse et l'autre légitime.

Dans ce contexte, la société SKAIZen Group se positionne comme une entreprise de conseil en gestion spécialisée dans la transformation numérique et la gestion des risques financiers. Les travaux de cette thèse CIFRE s'inscrivent dans le cadre d'une collaboration avec cette entreprise, visant à élaborer des solutions pour relever les défis auxquels sont confrontés ses clients, principalement des institutions financières. Actuellement, la lutte contre la fraude financière représente un enjeu majeur pour ces institutions.

SWIFT (Society for Worldwide Interbank Financial Telecommunication) est un réseau sécurisé et standardisé qui propose différents services permettant le transfert d'argent entre des comptes bancaires se trouvant dans des pays différents. Ce réseau permet de réaliser des transactions financières entre plus de 11000 organismes bancaires à travers près de 200 pays. La fraude financière au sein des transactions SWIFT est un problème majeur pour les institutions financières. Elles sont confrontées à la nécessité de contrôler des transactions complexes impliquant de multiples acteurs et intermédiaires dans plusieurs pays, avec différentes devises.

1.2 Problématique

Il est essentiel de comprendre le rôle des acteurs impliqués dans les transactions financières, notamment lorsqu'ils effectuent des échanges d'argent en étant situés dans des pays différents. Le réseau SWIFT permet et facilite ces échanges internationaux. Toutefois, si les institutions financières de ces acteurs ne disposent pas de relations d'échange établies, d'autres institutions financières interviennent en tant qu'intermédiaires pour faciliter la transaction. Selon la nature des relations entre les institutions financières, des changements de devises peuvent survenir lors de l'échange d'argent. Toutes ces informations sont comprises dans une transaction SWIFT, et chaque information ou relation entre les acteurs complexifie l'analyse de ces transactions. Néanmoins, la nature confidentielle des transactions SWIFT limite leur accessibilité en matière d'analyse.

Dans la lutte contre la fraude, il est essentiel de comprendre les caractéristiques d'une transaction frauduleuse. Les informations permettant de distinguer une transaction légitime d'une transaction frauduleuse sont liées à son contexte, comprenant les acteurs impliqués, les pays concernés et les devises utilisées. En général, une transaction frauduleuse suit un schéma spécifique, qui représente un mode opératoire utilisé pour commettre une fraude. Par exemple, l'un des schémas de fraude les plus couramment rencontrés est la dissimulation de grandes sommes d'argent en effectuant plusieurs transactions de petits montants. Dans le réseau SWIFT, ces schémas peuvent être plus complexes et impliquer des paramètres supplémentaires tels que des acteurs intermédiaires, des pays et des devises différents.

Cependant, les systèmes actuels, basés sur des règles prédéfinies, s'adaptent mal aux nouveaux schémas de fraude de plus en plus complexes. Par conséquent, ces systèmes détectent un faible nombre de transactions frauduleuses et génèrent un grand nombre de fausses alertes, également appelées faux positifs.

Dans ce travail, nous avons identifié plusieurs verrous techniques et scientifiques dans le domaine de la détection de fraude :

1. Verrous techniques :
   - Accessibilité et volume de données : Les données financières sont souvent privées et difficiles à obtenir, notamment lorsqu'il s'agit de données spécifiques telles que les transactions SWIFT. Cette limitation entrave considérablement les travaux de recherche et rend difficile la comparaison entre eux. Les recherches sont généralement menées en collaboration avec les institutions financières, ce qui limite le partage des données et des implémentations. De plus, le grand volume de données constitue un défi supplémentaire, car les méthodes doivent être capables de traiter toutes les données sans ralentir les systèmes.
   - Protection des données : L'accès aux données clients est souvent aussi difficile que pour les transactions financières. La surveillance des clients est d'une part obligatoire pour les institutions financières lors de l'ouverture et du suivi de leurs comptes bancaires, mais d'autre part, leur utilisation pour la détection de fraude est très réglementée d'un point de vue légal et éthique.
2. Verrous scientifiques :
   - Sélection de caractéristiques : La sélection de caractéristiques pertinentes est primordiale dans la détection de fraude. Il est souvent difficile de définir quelles sont

les caractéristiques les plus importantes pour identifier les fraudes avec la complexité des schémas de fraude.

— Complexité des schémas de fraude et nécessité d'un système adaptable : Les schémas de fraude sont de plus en plus complexes, et les fraudeurs s'adaptent en permanence aux systèmes en utilisant des méthodes de plus en plus sophistiquées. Les systèmes actuels ne s'adaptent pas efficacement aux nouveaux schémas de fraude et souffrent d'un manque d'efficacité, ce qui représente un coût financier considérable pour les institutions. De plus, certaines transactions frauduleuses sont plus facilement repérables que d'autres selon la complexité des schémas. Il est donc nécessaire de classifier les transactions suspectes par ordre de gravité, de manière à traiter en priorité les plus suspectes.

— Évaluation et mesure de performance : Il est important de développer des métriques adaptées au domaine de la finance afin de mesurer l'efficacité des techniques de détection de fraude en prenant en compte le coût financier. Les faux négatifs correspondent aux transactions frauduleuses qui ne sont pas détectées, et il est difficile d'estimer leur nombre réel, car elles ne génèrent pas d'alerte. Le nombre de transactions frauduleuses détectées est insuffisant, exposant ainsi les institutions financières à d'importantes amendes réglementaires. Les faux positifs, quant à eux, correspondent à des transactions légitimes qui génèrent des alertes, et ces alertes sont contrôlées par des experts qui décident de bloquer ou de libérer la transaction. Le grand nombre de faux positifs entraîne des coûts financiers significatifs pour les institutions financières liés aux services des experts.

— Interprétabilité des modèles : Les modèles de détection de fraude basés sur des techniques d'apprentissage automatique peuvent être très performants, mais ils sont souvent considérés comme des boîtes noires en termes d'interprétation. Dans le domaine de la finance, il est essentiel de développer des méthodes pour rendre les modèles interprétables pour les experts. Une transaction détectée frauduleuse doit être contrôlée par un expert. Il n'existe pas de consensus sur les informations à apporter à cet expert pour guider son enquête.

La problématique de cette thèse peut alors se résumer par la question suivante :

Comment analyser et interpréter efficacement les transactions dans un réseau financier international impliquant plusieurs acteurs, tout en identifiant les schémas de fraude et en minimisant les coûts financiers associés à un système de lutte contre la fraude ?

1.3 Méthodologie

Pour répondre à cette problématique, nous proposons une approche permettant d'analyser les transactions SWIFT, en prenant en compte les défis identifiés précédemment. En nous basant sur la littérature et en fonction de notre positionnement, nous avons développé une méthodologie pour l'analyse des transactions et la détection de transactions frauduleuses tout en minimisant les coûts associés, comme illustrée dans la figure 1.1. Cette méthodologie s'applique à un jeu de données contenant des transactions qui ont déclenché des alertes, certaines étant légitimes et d'autres frauduleuses. Cependant, nous ne disposons pas d'informations sur les raisons pour lesquelles elles ont été considérées frauduleuses. La méthode que nous proposons repose sur les principes suivants :

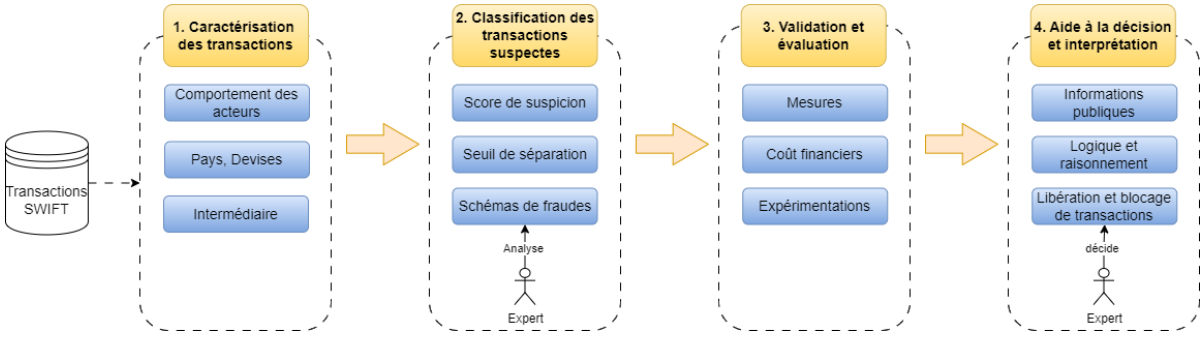

FIGURE 1.1 – Méthodologie générale

— Caractérisation des transactions : Nous identifions les caractéristiques clés qui peuvent être utilisées pour distinguer les transactions frauduleuses des transactions légitimes dans un jeu de données de transactions SWIFT. Cela peut inclure des informations telles que les pays impliqués, les montants, les acteurs et d'autres informations pertinentes.
— Classification des transactions suspectes : Nous calculons un score de suspicion pour chaque transaction en fonction de son appartenance à des schémas de fraude potentiels.
— Validation et évaluation : Nous évaluons les performances du modèle à l'aide de mesures classiques du domaine, telles que la précision, le rappel et le F1-score. Cela nous permet d'estimer l'efficacité du modèle dans la détection des transactions frauduleuses. De plus, nous estimons le coût financier associé aux erreurs de classification, à la fois en termes de faux positifs et de faux négatifs.
— Aide à la décision et interprétation : Nous définissons une ontologie du domaine qui représente les concepts des transactions et des acteurs du système financier. Nous peuplons cette ontologie avec les transactions suspectes prédites par le modèle. Les experts peuvent alors utiliser ces informations combinées à des données externes pour prendre des décisions sur le blocage ou la libération de ces transactions. L'utilisation du raisonnement et des inférences permet d'apporter une aide supplémentaire dans le processus de décision.

Notre méthode repose sur une approche hybride combinant des techniques d'apprentissage automatique et sémantiques afin de détecter les transactions suspectes liées à des schémas de fraude. Les modèles d'apprentissage automatique identifient les transactions suspectes, puis nous appliquons des techniques sémantiques pour filtrer les résultats et faciliter la prise de décision des experts responsables de ces transactions. Cette approche hybride permet de combiner l'efficacité des modèles de l'apprentissage automatique et exploiter les informations contextuelles pour améliorer la détection des transactions frauduleuses tout en réduisant les coûts liés aux fausses alertes.

1.4 Contributions

Nous présentons dans ce qui suit les principales contributions de cette thèse.

Définition des caractéristiques associées aux transactions SWIFT : Nous avons modélisé les comportements des acteurs et les chemins des transactions en prenant en compte les pays

et les devises. Dans un premier temps, nous avons utilisé un modèle basé sur les arbres de décision pour sélectionner les caractéristiques les plus pertinentes, en évaluant l'importance de chaque caractéristique dans les prédictions du modèle.

Analyse et identification des schémas de fraude : Une analyse du jeu de données en collaboration avec un expert a permis d'estimer le nombre de schémas de fraude. Les transactions frauduleuses ont été regroupées en clusters à l'aide de techniques d'apprentissage non supervisé. Par la suite, nous avons entraîné un modèle pour prédire si une transaction appartient à la classe légitime ou à l'une des classes associées aux schémas de fraude. Des techniques d'interprétation et de visualisation du modèle ont été utilisées pour attribuer chaque cluster à un schéma de fraude spécifique.

Calcul d'un score de suspicion pour les transactions : Nous avons entraîné un modèle d'apprentissage supervisé pour classifier les transactions, en utilisant les classes correspondant aux schémas de fraude ainsi qu'une classe pour les transactions légitimes. À partir des prédictions du modèle, nous avons développé une méthode de calcul du score de suspicion. Ce score permet d'établir un classement des transactions suspectes en fonction de leur degré de suspicion, facilitant ainsi le processus de contrôle.

Minimisation des coûts : Afin de déterminer le seuil de suspicion à partir duquel une transaction est considérée comme suspecte, nous avons proposé une évaluation quantitative des coûts financiers associés aux prédictions du modèle. En combinant ces coûts avec les mesures d'évaluation, nous avons cherché à minimiser les coûts de notre modèle tout en préservant son efficacité.

Création d'une ontologie pour la détection et l'analyse des transactions suspectes : Nous avons défini une ontologie spécifique au domaine de la détection et de l'analyse des transactions suspectes, permettant ainsi une meilleure interprétation des prédictions du modèle d'apprentissage automatique. Les concepts clés de cette ontologie sont les transactions et les acteurs. Nous avons peuplé l'ontologie avec des transactions suspectes et des informations provenant de sources externes. Cette ontologie joue un rôle d'aide à la prise de décision pour les transactions suspectes, orientant les actions des experts en contrôle.

L'outil ST-FRAUD : Nous avons développé un outil pour valider les propositions de cette thèse. Cet outil est composé de cinq modules :

1. Le premier module calcule les caractéristiques sélectionnées par l'expert et identifie les plus pertinentes.
2. Le deuxième module propose des outils de visualisation pour identifier le nombre potentiel de schémas de fraude dans le jeu de données.
3. Le troisième module utilise des techniques d'apprentissage non supervisé pour regrouper les transactions frauduleuses en clusters selon les schémas de fraude.
4. Le quatrième module entraîne un modèle pour classer les transactions dans la classe légitime ou l'une des classes frauduleuses.
5. Enfin, le dernier module utilise l'ontologie pour analyser les transactions suspectes détectées par le modèle.

1.5 Plan de la thèse

Cette thèse est organisée comme suit : le chapitre 2 présente l'état de l'art des techniques de détection des transactions frauduleuses, en mettant en évidence les approches basées sur l'apprentissage automatique. Les approches basées sur les ontologies sont également présentées. Le chapitre 3 présente notre approche hybride pour la détection et l'analyse des transactions frauduleuses. Le chapitre 4 détaille les expérimentations sur un jeu de données SWIFT et présente l'outil ST-Fraud, développé pour la détection et l'interprétation des transactions frauduleuses. Enfin, le chapitre 5 conclut cette thèse et aborde les perspectives futures de recherche dans le domaine de la détection des fraudes.

Chapitre 2

État de l'art : les techniques de la détection de transactions frauduleuses

Sommaire

2.1 Introduction

La fraude financière constitue une préoccupation majeure dans le secteur financier, impliquant des actes frauduleux visant à obtenir illégalement des avantages financiers. Les institutions financières font face au défi de détecter et de prévenir ces fraudes, mais les systèmes actuels montrent des limites importantes en termes d'efficacité. En effet, seule une faible proportion de transactions frauduleuses est détectée, exposant ainsi les institutions financières à des pertes financières importantes.

La surveillance des transactions financières est complexe en raison de plusieurs facteurs. Les transactions frauduleuses se dissimulent parmi des millions de transactions légitimes, rendant leur identification difficile. Les fraudeurs adoptent également des stratégies sophistiquées pour contourner les règles et les mécanismes de détection, rendant la tâche encore plus ardue.

Actuellement, les institutions financières utilisent principalement des systèmes de règles pour détecter la fraude financière. Ces systèmes sont basés sur des seuils préétablis et des règles manuellement définies par des experts. Cependant, ces approches statiques présentent des limitations importantes. Les fraudeurs peuvent rapidement identifier ces règles et ajuster leurs comportements en conséquence, échappant ainsi aux mécanismes de détection de fraude. De plus, ces systèmes génèrent souvent un grand nombre de fausses alertes, nécessitant un contrôle manuel intensif et entraînant des inefficacités opérationnelles.

Les techniques d'apprentissage automatique sont largement utilisées dans le domaine de la détection des transactions frauduleuses. L'apprentissage automatique, en particulier l'apprentissage supervisé, a montré une grande efficacité dans la détection de fraude. Grâce à des algorithmes sophistiqués, il est capable d'apprendre à partir de données historiques et de détecter des schémas de fraude complexes. Les modèles d'apprentissage automatique peuvent identifier des relations non linéaires entre les caractéristiques des transactions et les schémas de fraude, permettant ainsi de détecter des activités suspectes avec une précision élevée. De plus, l'apprentissage automatique peut être appliqué à grande échelle, en traitant de vastes volumes de données rapidement et efficacement.

Les approches sémantiques, telles que l'utilisation d'ontologies et de raisonnement, apportent des avantages distincts. Les ontologies permettent de représenter les connaissances dans un domaine spécifique, facilitant ainsi la modélisation des concepts et des relations entre les entités. Le raisonnement peut être utilisé pour inférer de nouvelles informations à partir des données disponibles, en exploitant les règles logiques et les axiomes définis dans l'ontologie.

Dans la section suivante, nous présentons les méthodes d'apprentissage automatique dans le domaine de la détection de fraude.

2.2 Approches basées sur l'apprentissage automatique

L'apprentissage automatique (*machine learning*) est une branche de l'intelligence artificielle qui se concentre sur la construction de systèmes informatiques capables d'apprendre et de s'améliorer à partir de données, sans être explicitement programmés pour chaque tâche spécifique. Une définition couramment utilisée est celle proposée par Tom Mitchell en 1997

[M+07] : « *Un programme informatique apprend à partir de l'expérience E pour une tâche T et une mesure de performance P, s'il améliore sa performance P pour la tâche T à partir de son expérience E.* ». Autrement dit, l'objectif de l'apprentissage automatique est de découvrir des structures dans les données pour résoudre des problèmes, plutôt que de les programmer explicitement.
Dans le contexte de la détection de fraude, l'apprentissage automatique est utilisé pour développer des modèles prédictifs capables d'identifier les transactions frauduleuses en se basant sur des caractéristiques et des schémas cachés dans les données financières. Les modèles d'apprentissage automatique peuvent être formés à partir d'un ensemble de données historiques comprenant à la fois des transactions frauduleuses et légitimes.
Lorsqu'il est utilisé pour la détection de fraude, l'apprentissage automatique permet de créer des modèles qui apprennent à reconnaître les schémas et les comportements associés aux activités frauduleuses. Ces modèles peuvent ensuite être utilisés pour évaluer de nouvelles transactions et estimer leur probabilité d'être frauduleuses. L'apprentissage automatique comprend généralement 5 étapes représentées sur la figure 2.1 :

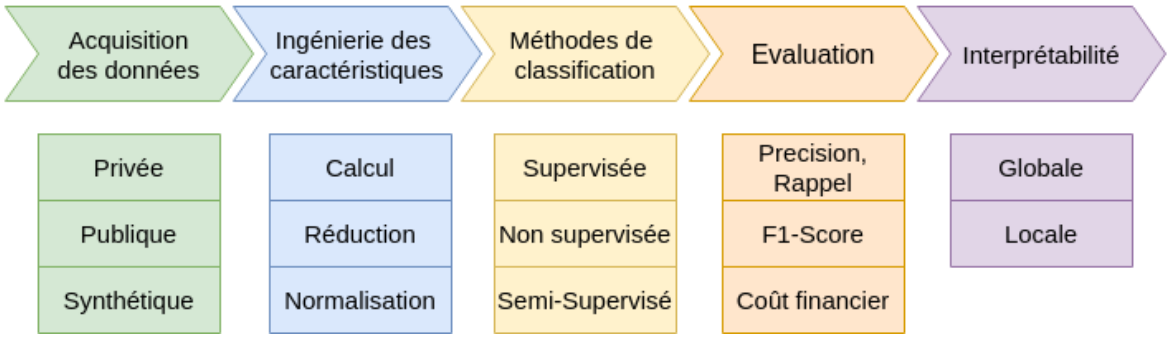

FIGURE 2.1 – Les 5 étapes des techniques d'apprentissage automatique

— Acquisition et préparation des données : La collecte des données est indispensable pour l'entraînement et la validation des modèles d'apprentissage automatique. Dans des domaines comme la finance ou la médecine où les données sont confidentielles, cette étape peut se révéler complexe. Certains travaux recourent à des données synthétiques pour mener leurs expérimentations.
— Ingénierie des caractéristiques : Le calcul, la sélection ou réduction des caractéristiques des données permet d'améliorer la performance des modèles.
— Le choix des algorithmes d'apprentissage et de leurs hyperparamètres : Le choix du modèle d'apprentissage automatique dépend du problème et des données disponibles. L'apprentissage peut être supervisé, non supervisé ou semi-supervisé en fonction du jeu de données et des prédictions souhaitées. Le modèle sélectionné est ensuite entraîné en utilisant les données d'entraînement, étape durant laquelle il apprend à identifier les schémas et à effectuer des prédictions. Cette étape revêt une importance pour l'objectif de détection de la fraude, car elle vise à proposer des méthodes permettant de séparer les transactions frauduleuses des transactions légitimes. Les hyperparamètres du modèle sont ajustés pour optimiser ses performances sur les données de validation.
— Évaluation du modèle : L'évaluation des modèles est effectuée en utilisant des données de validation. Cela permet de valider les performances du modèle. Diverses mesures

d'évaluation peuvent être utilisées, telles que le rappel et la précision.
— Interprétabilité : L'interprétation des résultats est importante, en particulier dans des domaines sensibles tels que la détection de fraude financière. Ceci est essentiel pour comprendre les raisons sous-jacentes aux décisions prises par les modèles.

Les différentes étapes sont détaillées dans la suite de ce chapitre. Dans la section suivante, nous abordons les défis liés aux données dans le domaine de l'apprentissage automatique, en mettant particulièrement l'accent sur leurs applications dans le domaine financier.

2.2.1 Acquisition des données

Les données jouent un rôle important dans le processus d'apprentissage automatique. En effet, le modèle est entraîné en se basant sur les données disponibles. La qualité ainsi que la quantité de ces données ont un impact direct sur la qualité des prédictions et des résultats obtenus. Le choix des données à utiliser dépend du problème à résoudre. Il est essentiel de les adapter en fonction de l'objectif à atteindre.

L'étape de collecte de données est importante pour l'entraînement d'un modèle. Un choix inapproprié des données peut entraîner des erreurs de prédictions. Deux concepts essentiels à connaître lors d'entraînement de modèle, également appelée le dilemme biais-variance, sont les suivants :
— La variance : Elle mesure la sensibilité d'un modèle aux variations des données d'entraînement. Un modèle présentant une variance élevée a tendance à être excessivement complexe et à sur-ajuster les données d'entraînement. Cela peut se produire si le modèle est trop complexe en lui-même, soit lorsqu'il dispose de trop peu de données d'entraînement pour soutenir un modèle complexe.
— Le biais : Il fait référence aux erreurs de prédiction d'un modèle par rapport à la vérité terrain. Cela se produit lorsque le modèle est trop simplifié et ne parvient pas à capturer la complexité des données.

Le dilemme entre le biais et la variance survient lorsque nous cherchons à construire un modèle capable de généraliser les corrélations des données d'entraînement tout en étant efficace avec de nouvelles données. Si un modèle est trop simple, il présente un biais élevé et ne pourra pas capturer la complexité des données. Si un modèle est trop complexe, il présente une variance élevée et devient trop sensible aux variations des données d'entraînement.

L'apprentissage automatique fait face à plusieurs difficultés liées aux données. Les principales problématiques sont les suivantes [SM22] :

— Volume de données insuffisant : Un nombre insuffisant de données empêche le modèle d'apprendre efficacement à partir des données et limite sa capacité à généraliser correctement les prédictions.
— Données non représentatives : Les données utilisées doivent être représentatives du sujet traité afin de permettre une généralisation précise des prédictions. Si les données sont biaisées ou non représentatives de l'ensemble de données, le modèle peut produire des prédictions erronées.
— Mauvaise qualité : Les données de mauvaise qualité, telles que des données manquantes, ou des valeurs aberrantes, peuvent également entraîner des erreurs et des biais dans le modèle.

— Caractéristiques non pertinentes : La présence de caractéristiques redondantes ou sans impact sur les résultats peut affecter l'efficacité du modèle et conduire à un surapprentissage du modèle.
— Surapprentissage des données d'entraînement : Il se produit lorsque le modèle ou l'algorithme s'adapte trop étroitement aux données, capturant le bruit plutôt que les relations réelles. Cela se traduit par une variance élevée mais un faible biais.
— Sous-apprentissage des données d'entraînement : Il se produit lorsque le modèle ou l'algorithme ne parvient pas à capturer la tendance sous-jacente des données, généralement en raison de sa simplicité excessive. Cela se traduit par une faible variance, mais un fort biais.

Dans le domaine financier, les travaux sur la détection de fraude se basent sur des données transactionnelles. Ces transactions possèdent des caractéristiques de base telles que le montant, la date et les acteurs [KDT20, PM18, CLBC+21]. Selon le contexte, ces transactions peuvent prendre différents formats, voici quelques exemples :
— Les transactions par carte de crédit impliquent un client, un commerçant et des attributs de base [RSM+18].
— Les transactions au sein des institutions bancaires d'investissement englobent un client, un type de transaction (retrait ou dépôt), un montant et une date qui peuvent refléter des investissements ou des retraits d'argent [LKK10].
— Les transactions liées à des services d'exportation possèdent des informations telles que le volume et la valeur des biens exportés [PLCM16].

Lors de l'analyse des données utilisées dans les expérimentations de la littérature, nous avons observé une grande hétérogénéité en termes de volume de données, allant de milliers [WD09] à des millions [TL20] de transactions. Une transaction peut être qualifiée de frauduleuse si elle est associée à une activité illégale. Les transactions légitimes sont celles effectuées conformément aux lois et aux réglementations. Une caractéristique commune des ensembles de données de fraude financière est le déséquilibre de classe entre le nombre de transactions frauduleuses et légitimes, avec un ratio frauduleux généralement autour de 0,1% [AJA20].
La validation des approches de détection de fraude nécessite des données étiquetées indiquant la vérité terrain (fraude ou légitime). En utilisant ces ensembles de données labellisés, il devient possible d'évaluer la capacité des modèles à détecter efficacement les transactions frauduleuses tout en minimisant les faux positifs (transactions légitimes prédites comme frauduleuses) et les faux négatifs (transactions frauduleuses prédites comme légitimes). Les ensembles de données labellisés jouent un rôle central dans la validation et l'amélioration des méthodes de détection de fraude dans le contexte des transactions financières.

Le déséquilibre entre le nombre de transactions frauduleuses et de transactions légitimes peut être résolu en utilisant des techniques d'augmentation de données sur l'ensemble d'entraînement, telles que l'algorithme SMOTE [CBHK02]. Cet algorithme génère des données synthétiques de la classe minoritaire en interpolant des données existantes, ce qui améliore l'apprentissage des modèles de détection de fraude. De plus, les techniques de réduction de données peuvent être utilisées pour diminuer la taille des données d'entrée tout en conservant des informations importantes pour la détection de fraude, ce qui permet d'améliorer la vitesse de traitement et la scalabilité des modèles.

Dans le domaine financier, l'obtention de données publiques contenant des informations pri-

vées et sensibles sur les clients peut s'avérer difficile. Cependant, les chercheurs ont exploré la possibilité de générer des données financières synthétiques, qui sont des transactions fictives créées en se basant sur des règles statistiques et des profils de clients simulés synthétiquement. Lopez et al. [LRA12] ont développé *PaySim*, un outil de simulation de paiement mobile avec des transactions frauduleuses. Ils ont mis à disposition un jeu de données sur *Kaggle* [1]. Toutefois, ce jeu de données présente certaines limitations. En l'étudiant, nous trouvons que les transactions frauduleuses sont faciles à identifier, car elles correspondent à la liquidation totale du compte de la personne à l'origine de la transaction, le scénario de fraude est facilement identifiable. Des travaux utilisent ce jeu de données pour conduire leurs expérimentations comme Feng et al. [FLW+19] pour développer un système de détection et de catégorisation de comportements suspects basé sur les données de transactions financières collectées ou Stojanovic et al. [SBHS+21] pour évaluer les méthodes de détection d'anomalies.

En ce qui concerne la génération et l'utilisation de données synthétiques, Michalak et al. [MK11] détaillent leur méthode de génération de transactions, où ils utilisent une distribution gaussienne pour générer des réseaux de transactions entre les employés d'entreprises. Lv et al. [LJZ08] utilisent des données réelles combinées à des données frauduleuses générées artificiellement, mais ne détaillent pas le processus de génération. Cependant, la plupart des approches étudiées pour la génération de données ne fournissent pas d'explication sur le processus utilisé pour obtenir ces nouvelles données [TY05, KT11].

Un grand nombre d'approches de la littérature utilisent l'ensemble de données public provenant de *Kaggle* [2] pour développer et valider leurs méthodes de détection de transactions frauduleuses. Cet ensemble de données renferme des transactions réalisées avec des cartes de crédit impliquées dans des fraudes. Il comporte trois attributs principaux : la date, le montant, la classe de transaction (frauduleuse ou légitime), ainsi que 28 autres attributs anonymisés. L'exploitation d'ensembles de données comportant des attributs anonymisés présente une certaine complexité, car la signification de chaque attribut doit être connue pour proposer des approches adaptées aux données. Par exemple, Varmedja et al. [VKS+19] utilisent cet ensemble de données pour analyser l'impact de l'algorithme SMOTE sur ce jeu de données et pour comparer les résultats avec des algorithmes de classification.

L'acquisition des données dans le domaine de la finance présente des défis majeurs qui entravent les avancées de la recherche. Les ensembles de données publics présentent des lacunes, telles qu'un nombre restreint de scénarios de fraude ou la présence d'attributs anonymisés. Les ensembles de données financières se caractérisent par des attributs de base tels que le montant, les intervenants et la date des transactions. Une fois que l'ensemble de données est acquis, l'étape suivante concerne l'ingénierie des caractéristiques. Cette phase consiste à sélectionner et à transformer les variables ou les caractéristiques qui seront utilisées dans un modèle d'apprentissage automatique. Son objectif est d'optimiser les performances du modèle en améliorant la qualité des données d'entrée.

1. https://www.kaggle.com/datasets/ealaxi/paysim1
2. https://www.kaggle.com/mlg-ulb/creditcardfraud

2.2.2 Ingénierie des caractéristiques

L'ingénierie des caractéristiques joue un rôle important dans l'apprentissage automatique, consistant à transformer des données brutes en caractéristiques pertinentes pour les modèles d'apprentissage. Dans le contexte spécifique de la détection de fraude, l'enjeu est encore plus important en raison du volume des données souvent accompagné d'un nombre limité d'attributs pertinents. Dans ce domaine, l'ingénierie des caractéristiques vise essentiellement à sélectionner les caractéristiques les plus pertinentes pour les modèles d'apprentissage automatique. En sélectionnant judicieusement ces caractéristiques, on renforce la capacité des modèles à identifier efficacement les transactions frauduleuses, minimisant ainsi les risques de faux positifs et de faux négatifs.

Plusieurs techniques sont utilisées dans l'ingénierie des caractéristiques. Dans la suite de cette section, nous présentons les techniques les plus pertinentes et les plus utilisées pour nos travaux

2.2.2.1 Les types de caractéristiques

Les caractéristiques, également appelées variables ou attributs, sont les entrées utilisées pour entraîner un modèle à prédire une variable cible. Elles peuvent être classées en plusieurs catégories selon leur nature et leur format. Voici les principales catégories de caractéristiques [Rum11] :

- Caractéristiques numériques : Ce sont des variables quantitatives qui peuvent être continues ou discrètes, telles que le montant d'une transaction.
- Caractéristiques catégorielles : Ce sont des variables qualitatives qui prennent des valeurs dans un ensemble fini de catégories, comme la devise d'une transaction.
- Caractéristiques binaires : Il s'agit d'un type particulier de caractéristiques catégorielles qui ne prennent que deux valeurs possibles, généralement codées comme 0 ou 1, telles que la présence ou l'absence d'un intermédiaire dans une transaction.
- Caractéristiques ordinales : Ce sont des variables catégorielles qui ont un ordre naturel entre les catégories, telles que le niveau de risque associé à une transaction pouvant aller de faible à élevé.
- Caractéristiques temporelles : Il s'agit de variables qui mesurent un temps ou une durée, telles que la date de la transaction.

Un jeu de données de transactions financières peut contenir différentes catégories de caractéristiques telles que les montants des transactions, les devises, les dates et heures de transaction, ainsi que d'autres informations pertinentes. La présence de ces différentes catégories de caractéristiques dans un même jeu de données peut rendre la tâche d'analyse plus complexe, en particulier pour certains algorithmes qui se basent sur des mesures de distance. Les distances calculées servent à évaluer la similarité entre les données : plus la distance entre deux données est réduite, plus elles sont similaires. Certains algorithmes de regroupement ou de classification peuvent rencontrer des difficultés à traiter des données hétérogènes, car ils supposent souvent que les caractéristiques sont de même nature et donc similaires. Par exemple, lorsqu'un algorithme de regroupement est utilisé pour grouper des transactions semblables, il peut éprouver des difficultés à former des groupes homogènes lorsque les catégories de caractéristiques diffèrent considérablement les unes des autres.

2.2.2.2 L'encodage des caractéristiques catégorielles

La plupart des techniques d'apprentissage automatique utilisent des caractéristiques numériques. Ainsi, des techniques d'encodage des caractéristiques catégorielles sont employées pour convertir les variables catégorielles en variables numériques, afin de les utiliser dans les modèles d'apprentissage automatique. Les algorithmes de machine learning requièrent généralement une représentation numérique des données pour fonctionner correctement, ce qui fait de l'encodage des variables catégorielles une étape essentielle du prétraitement des données pour l'apprentissage automatique [Fla12]. Les principales techniques d'encodage des caractéristiques catégorielles sont les suivantes [DJ21] :

- *One-Hot Encoding* : Cette technique consiste à créer une nouvelle colonne pour chaque valeur unique dans la caractéristique catégorielle. Chaque colonne contient des valeurs binaires qui indiquent si la valeur appartient ou non à cette catégorie. Cette technique est souvent utilisée pour les caractéristiques catégorielles avec un petit nombre de valeurs uniques, telles que le type de paiement d'une transaction.
- *Label Encoding* : Cette technique consiste à remplacer chaque valeur unique dans la caractéristique catégorielle par un entier. Les entiers sont généralement assignés en se basant sur l'ordre des valeurs. On recourt fréquemment à cette méthode pour les caractéristiques catégorielles comportant un grand nombre de valeurs uniques, telles que le pays associé à un intervenant d'une transaction.
- *Binary Encoding* : Cette technique consiste à convertir chaque valeur unique dans la caractéristique catégorielle en une représentation binaire. Chaque colonne contient un seul bit qui indique si la valeur appartient ou non à cette catégorie. Cette technique est souvent utilisée pour les caractéristiques catégorielles avec un nombre modéré de valeurs uniques, telles que la devise d'une transaction.
- *Hashing Encoding* : Cette technique consiste à convertir chaque valeur unique dans la caractéristique catégorielle en un nombre entier en utilisant une fonction de hachage. Le nombre entier est ensuite utilisé comme caractéristique numérique. L'utilisation de cette technique est fréquente pour les caractéristiques catégorielles comportant un grand nombre de valeurs uniques, et elle peut se révéler utile pour réduire la dimensionnalité des caractéristiques, comme dans le cas de l'identifiant d'un acteur d'une transaction.

En résumé, le choix de la technique d'encodage dépend de la nature de la caractéristique catégorielle et de la quantité de données disponibles. Il est important de prendre en compte ces facteurs lors du choix de la méthode d'encodage pour garantir une représentation efficace et précise des données catégorielles dans le cadre de l'apprentissage automatique.

2.2.2.3 Création de caractéristiques

Bolton et Hand [BH02] ont souligné la nécessité de créer de nouvelles caractéristiques sur les transactions pour distinguer les transactions frauduleuses des transactions légitimes. En effet, les attributs de base des transactions ne sont pas suffisants pour réaliser cette distinction.

La création de nouvelles caractéristiques dans le domaine de la détection de fraude par carte de crédit repose sur le modèle *RFM (Recency - Frequency - Monetary)* [BKN+08]. Dans le domaine de la détection de transactions frauduleuses, la **récence** est caractérisée par la différence de temps entre une transaction courante et une transaction la précédant, la **fréquence** par

le nombre de transactions dans une période, et la **valeur monétaire** décrit le montant dépensé par l'acteur [KFMBEN21, VVBC$^+$15].

Concernant les fraudes via transactions par carte de crédit, Whittrow et al. [WHJ$^+$09] ont élaboré une méthode qui repose sur l'agrégation du comportement antérieur d'un acteur sur une période définie. Lors de la détermination des caractéristiques d'une transaction x_i, celles-ci sont définies en fonction des informations intrinsèques à la transaction ainsi que des données relatives au client initiateur. À partir de l'ensemble $S = \{s_1, ..., s_i, s_n\}$ des transactions de l'historique d'un client, Whittrow et al. estiment la somme totale (sum) des montants (x_{amt}) et le volume de transactions ($count$) présents dans S.

Bhattacharyya et al. [BJTW11] soulignent le problème des transactions financières qui ne possèdent généralement que le montant comme caractéristique numérique, les autres caractéristiques étant catégorielles ou temporelles. À partir des caractéristiques de base d'une transaction (appelées caractéristiques primaires), ils créent 16 caractéristiques dérivées. Pour une transaction x_i, ils calculent trois types de caractéristiques : la somme (sum_i), la fréquence ($count_i$) et la moyenne ($\frac{sum_i}{count_i}$) sur différents ensembles de transactions. Ils construisent des ensembles de transactions avec les transactions passées de l'acteur, les transactions passées du même type d'acteur que x_i et les transactions passées du même pays que les acteurs de x_i. De plus, ces ensembles sont construits sur trois périodes : le jour, le mois et les trois mois précédant la date de la transaction x_i. Ces nouvelles caractéristiques servent d'indicateurs permettant de mieux caractériser les acteurs.

Les transactions SWIFT présentent des particularités distinctes, telles que l'intervention d'intermédiaires et la participation de plusieurs pays. À notre connaissance, la littérature existante n'a pas exploré en profondeur la création de caractéristiques spécifiques à ce type de transactions. Cette lacune représente une opportunité de recherche pour développer des méthodes d'analyse adaptées aux particularités des transactions SWIFT. Une exploration approfondie de ces caractéristiques pourrait contribuer au développement d'un modèle efficace pour la détection de fraude dans le domaine des transactions internationales et interbancaires.

2.2.2.4 Réduction de caractéristiques

Chaque caractéristique créée ajoute une dimension supplémentaire au jeu de données. Lorsque des algorithmes de fouille de données et d'apprentissage automatique sont appliqués à des données de haute dimension, un problème critique connu sous le nom de "malédiction de la dimensionnalité" [HTF01] se pose. Ce terme, inventé par Bellman en 1961 [Bel61], fait référence aux problèmes associés à l'analyse de données multivariées lorsque la dimensionnalité augmente. En outre, les données de haute dimensionnalité augmentent considérablement les exigences de stockage en mémoire et les coûts de calcul nécessaires pour analyser les données. Par conséquent, la réduction de la dimensionnalité est devenue une étape nécessaire pour rendre l'analyse plus gérable et extraire des connaissances utiles. La réduction de la dimensionnalité peut être réalisée de deux manières différentes [Fos19] : par la sélection de caractéristiques ou par l'extraction de caractéristiques.

La sélection de caractéristiques consiste à choisir un ensemble significatif de caractéristiques à partir du jeu de données initial. Cela permet de supprimer les caractéristiques non

pertinentes (qui n'ont pas d'impact sur l'apprentissage du modèle) et les caractéristiques redondantes [GE03]. On peut distinguer trois types de méthodes de sélection de caractéristiques [JBB15] :

1. Les méthodes de filtrage : Ces méthodes évaluent l'importance des caractéristiques de manière indépendante de l'algorithme de modélisation. Elles utilisent des mesures statistiques pour évaluer la pertinence de chaque caractéristique par rapport à la variable cible (par exemple, la corrélation entre la caractéristique et la variable cible). Les caractéristiques sont ensuite classées en fonction de leur pertinence, et un sous-ensemble de caractéristiques est sélectionné pour la modélisation (par exemple, en utilisant le coefficient de corrélation de Pearson [CHC+09] qui mesure la relation linéaire entre deux variables quantitatives à l'aide d'un coefficient variant entre -1 (relation linéaire négative parfaite) et 1 (relation linéaire positive parfaite).
2. Les méthodes d'enveloppe : Ces méthodes évaluent l'importance des caractéristiques en utilisant un algorithme de modélisation spécifique. Elles utilisent une procédure itérative pour sélectionner un sous-ensemble de caractéristiques qui maximise les performances du modèle. Cette méthode peut être mise en œuvre de manière *greedy*, en partant soit d'un ensemble vide (*forward*) et en ajoutant des caractéristiques, soit en partant d'un ensemble contenant toutes les caractéristiques (*backward*) et en supprimant des caractéristiques jusqu'à obtenir un ensemble adéquat (par exemple, avec l'algorithme de *Recursive Feature Elimination* [GWBV02]).
3. Les méthodes intégrées : Ces méthodes fusionnent la sélection de caractéristiques avec l'entraînement du modèle en une seule étape. À titre d'exemple, l'algorithme Lasso [Tib97] attribue un coefficient à chaque caractéristique, et celles ayant un coefficient non nul sont généralement retenues. Dans le cas des algorithmes basés sur les arbres de décision, les caractéristiques obtiennent un score d'importance durant la construction des arbres. Les caractéristiques avec les scores d'importance les plus élevés sont alors privilégiées pour la sélection.

L'extraction de caractéristiques consiste à projeter l'espace des caractéristiques dans un espace de dimension inférieure [WP03]. Le nouvel espace peut résulter d'une combinaison linéaire ou non linéaire de l'espace d'origine. Parmi les algorithmes les plus couramment utilisés, on peut citer :

— L'Analyse en Composantes Principales (ACP) [MR93] est une technique de transformation linéaire qui projette les données dans un espace de dimension inférieure tout en maximisant la variance des données projetées. Les composantes principales obtenues à partir de l'ACP peuvent être utilisées comme caractéristiques pour la classification.
— L'Analyse Discriminante Linéaire (ADL) [TGIH17] est une technique de transformation linéaire similaire à l'ACP, qui cherche à maximiser la séparation entre les classes plutôt que la variance des données projetées. Les vecteurs propres obtenus à partir de l'ADL peuvent être utilisés comme caractéristiques pour la classification.
— L'Analyse de Corrélation Canonique (ACC) [Tho84] est une technique de transformation linéaire qui cherche à maximiser la corrélation entre deux ensembles de variables. L'ACC peut être utilisée pour extraire des caractéristiques à partir de deux ensembles de données qui sont corrélés.

D'autres techniques d'extraction de caractéristiques incluent l'analyse en composantes indé-

pendantes (ICA), qui cherche à extraire des sources indépendantes de données à partir de mélanges linéaires, et les autoencodeurs (*autoencoders*) basés sur des réseaux de neurones [BKG20]. Ces derniers apprennent à compresser les données dans un espace de dimension inférieure tout en préservant les informations pertinentes pour la reconstruction des données originales.

L'extraction de caractéristiques génère de nouveaux attributs qui, une fois transformés, peuvent perdre leur signification initiale. En revanche, la sélection de caractéristiques conserve les attributs d'origine, ce qui garantit une meilleure interprétabilité du modèle. Dans les problèmes de classification à haute dimension, la réduction des caractéristiques est reconnue pour améliorer la précision prédictive tout en fournissant des modèles à la fois plus rapides et plus interprétables.

2.2.2.5 Prétraitement des caractéristiques

La dernière étape de l'ingénierie des caractéristiques est le pré-traitement des données, qui vise à transformer les valeurs des caractéristiques pour faciliter l'apprentissage des modèles. La normalisation et la standardisation sont deux techniques couramment utilisées pour prétraiter les données dans le domaine de l'apprentissage automatique. Bien qu'elles aient toutes deux pour objectif de mettre les données sur une échelle commune, elles diffèrent dans leur fonctionnement [Bon17] :

- La normalisation redimensionne les données de manière à ce qu'elles se situent dans une plage spécifique, généralement entre 0 et 1. Cela s'effectue en soustrayant la valeur minimale de l'ensemble de données et en divisant par la plage (différence entre la valeur maximale et la valeur minimale).
- La standardisation transforme les données de sorte qu'elles aient une moyenne de 0 et un écart type de 1. Cela se fait en soustrayant la moyenne de l'ensemble de données, puis en divisant par l'écart type.

Il existe plusieurs techniques couramment utilisées en ingénierie des caractéristiques dans le domaine de la détection de transactions frauduleuses, telles que la normalisation, la transformation des données et la création de nouvelles caractéristiques à partir de combinaisons d'attributs existants. Cependant, malgré l'importance de cette étape dans la détection des fraudes, il n'existe pas de travaux spécifiques pour les transactions internationales et interbancaires. Dans la prochaine section, nous présentons les méthodes de classification des transactions frauduleuses.

2.2.3 Méthodes de classification de transactions frauduleuses

Dans cette troisième étape, nous abordons les méthodes de détection des transactions frauduleuses. Les systèmes de détection de fraude actuels reposent sur des règles pour repérer des comportements suspects. Cependant, ces règles sont généralement basées sur des données historiques et peuvent rapidement devenir obsolètes dès lors qu'elles sont identifiées par les fraudeurs. En effet, ces derniers peuvent s'adapter en évitant les comportements qui déclenchent des alertes de fraude, exploitant ainsi les faiblesses du système de détection. Par conséquent, pour rester efficaces, les systèmes de détection de fraude doivent constamment être mis à jour et améliorés.

Dans cette section, nous présentons les méthodes de détection de transactions frauduleuses basées sur les techniques d'apprentissage automatique. Nous commençons par introduire l'apprentissage ensembliste, qui consiste à combiner les résultats de plusieurs modèles. Ensuite, nous abordons respectivement les trois approches permettant de former un modèle de classification en utilisant les techniques d'apprentissage supervisé, non supervisé et semi-supervisé.

2.2.3.1 Apprentissage ensembliste

Les techniques d'apprentissage ensembliste, également connues sous le nom de méthodes d'ensemble, sont des approches qui combinent les prédictions de plusieurs modèles d'apprentissage pour améliorer la précision globale des prédictions. Les ensembles peuvent être utilisés avec n'importe quel algorithme d'apprentissage automatique, mais ils sont généralement utilisés avec des arbres de décision en raison de leur variance élevée. Les différentes catégories de techniques d'apprentissage ensembliste sont les suivantes :

- Le *Bagging* [Bre96] : Le bagging est une méthode qui consiste à entraîner plusieurs modèles sur des sous-ensembles d'échantillons choisis aléatoirement. Les prédictions de chaque modèle sont combinées par moyenne ou vote pour obtenir une prédiction finale plus robuste et généralisable.
- Le *Boosting* [Fre95] : Le boosting est une technique qui consiste à entraîner une série de modèles de manière séquentielle, en donnant plus de poids aux exemples mal classés. Chaque modèle est entraîné sur un sous-ensemble différent des données d'entraînement pondérées.
- Le *Stacking* [DZ04] : Le stacking est une technique où plusieurs modèles sont formés sur les données d'entraînement, et leurs prédictions servent d'entrées à un modèle « métamodèle », qui effectue la prédiction finale.

Les algorithmes basés sur l'apprentissage ensembliste ont démontré leur efficacité dans la littérature. Des travaux tels que ceux de Lorenz et al. [LSA+20], Zareapoor et al. [ZS+15], et Varmedja et al. [VKS+19] ont mis en évidence l'efficacité de l'algorithme *Random Forest* pour la détection de fraude dans les transactions financières, en le comparant à d'autres algorithmes d'apprentissage supervisé tels que *SVM*, *Naive Bayes* et les réseaux de neurones. Il a été observé que les réseaux de neurones n'obtiennent pas les meilleurs résultats. Borisov et al. [BLS+22] expliquent cela en identifiant plusieurs raisons possibles, notamment la qualité des données (données manquantes, données aberrantes), l'absence de dépendances spatiales ou de corrélations complexes et irrégulières dans les ensembles de données tabulaires, la dépendance à l'étape de pré-traitement des données et l'importance élevée de certaines caractéristiques.

Plus récemment, des algorithmes basés sur le *boosting* sont apparus et ont démontré des résultats supérieurs à *Random Forest*, notamment avec les algorithmes *XGBoost* [CHB+15], *LightGBM* [KMF+17] et *CatBoost* [PGV+18]. Une étude comparative détaillée des algorithmes d'apprentissage supervisé a été présentée par Alfaiz et Fatih en 2022 [AF22], où l'impact des techniques d'augmentation et de réduction des données a été examiné. Leur expérimentation a été réalisée en utilisant le jeu de données *Kaggle* [3]. Ils ont obtenu des meilleurs résultats en termes de mesures telles que l'*accuracy*, la *precision*, le *rappel* et le *f1-score* avec les trois algorithmes de *boosting* mentionnés précédemment. Le modèle le plus performant était celui de *CatBoost*.

3. https://www.kaggle.com/mlg-ulb/creditcardfraud

2.2.3.2 Méthodes d'apprentissage supervisé

Dans le domaine de la détection de transactions frauduleuses, il existe des méthodes qui visent à améliorer les algorithmes d'apprentissage supervisé. Ces méthodes se concentrent souvent sur la sélection des hyperparamètres des algorithmes. Par exemple, en 2005, Tang et al. [TY05] proposent une nouvelle fonction de noyau pour l'algorithme *Support Vector Machines (SVM)* [HDO+98] basée sur la distance *HDVM* [WM97] qui sert à quantifier la divergence entre deux distributions de probabilité, permettant ainsi de déterminer à quel point elles sont similaires ou différentes, ce qui est essentiel dans des applications telles que la détection d'anomalies, la classification ou l'évaluation de modèles probabilistes. Toujours sur l'algorithme SVM, Keyan et Yu [KT11] propose une expérimentation d'une sélection croisée des hyperparamètres pour une tâche de classification sur des transactions frauduleuses.

En 2018, Metzler et al. [MBB+18] ont proposé une version modifiée de l'algorithme des arbres de décisions avec une nouvelle stratégie de séparation basée sur les coûts des transactions incorrectement prédites, spécifiquement conçue pour les institutions financières. Ils ont optimisé leur modèle en tenant compte le coût des erreurs de prédiction afin de minimiser les pertes pour les institutions financières. Leur méthodologie est intéressante, car, dans un premier temps, ils quantifient les coûts financiers des prédictions, puis adaptent l'algorithme des arbres de décisions qui est efficace dans le domaine de la détection de transactions frauduleuses. En réalisant cela, ils améliorent leurs résultats en termes de f1-score tout en réduisant la perte de coûts financier. Cependant, les résultats de leur f1-score restent relativement faibles avoisinant les 0.15 comparés aux autres approches qui ont des f1-score autours des 0.90 [PR19].

En 2019, Carta et al. [CFRS19] ont proposé une approche basée sur l'apprentissage ensembliste pour la détection de fraude. Leur méthode considère une transaction comme légitime si la probabilité de légitimité prédite par un modèle est supérieure à la moyenne des probabilités de légitimité de tous les modèles. Ensuite, un vote majoritaire est utilisé entre les modèles pour classer la transaction comme légitime ou frauduleuse. Ils ont expérimenté leur approche sur le jeu de données de transactions par carte de crédit [4]. Leurs mesures de sensibilité, de taux de fausses alarmes, d'AUC, de spécificité et de taux d'erreur ont montré de meilleurs résultats par rapport aux algorithmes tels que *Random Forest*, *Naive Bayes*, *Adaboost*, *Gradient Boosting* et *Multi-Layer Perceptron*.

En 2020, Rtayli et Enneya [RE20] ont proposé une méthode utilisant l'algorithme *Random Forest* pour la sélection de caractéristiques et l'algorithme *SVM* pour la tâche de classification, sur le même jeu de données de transactions par carte de crédit. Leurs résultats ont montré une amélioration significative par rapport aux algorithmes tels que *Local Outlier Factor*, *Isolation Forest* et *Decision Tree*. Cependant, ils n'ont pas comparé leurs résultats avec d'autres algorithmes d'apprentissage supervisé, à l'exception de *Decision Tree*, ce qui limite la généralisation de leurs conclusions.

Ces différentes méthodes contribuent à l'amélioration des techniques d'apprentissage supervisé. Tang et al. [TY05] ont optimisé l'algorithme SVM en utilisant la distance HDVM pour quantifier la divergence entre distributions. Metzler et al. [MBB+18] ont adapté les arbres de décision en tenant compte des coûts financiers des prédictions incorrectes. De plus, des travaux subséquents menés par Carta et al. [CFRS19] ainsi que par Rtayli et Enneya [RE20]

4. https://www.kaggle.com/mlg-ulb/creditcardfraud

ont exploré l'apprentissage ensembliste et la combinaison de Random Forest avec SVM pour améliorer la détection des transactions frauduleuses.

2.2.3.3 Méthodes d'apprentissage non supervisé

Les méthodes d'apprentissage non supervisé consistent à fournir à un algorithme des données non labellisées, afin qu'il puisse découvrir des structures et des relations cachées entre les données. Le clustering est un algorithme qui regroupe des données similaires en fonction de leurs caractéristiques communes.

En 2011, Larik et al. [LH11] ont introduit une méthode s'appuyant sur le clustering. Utilisant l'algorithme EART (*Euclidean ART Neural Networks*) [KC08] comme base, ils ont élaboré une variante nommée TEART (*Transformed EART*), qui fusionne les petits clusters en un seul. En appliquant TEART et k-means sur un ensemble de $8,2$ millions de transactions, caractérisées par leur fréquence et montant, ils ont développé un indice nommé AICAF (*Anomaly Index Computation based on Amount and Frequency*). Cet indice mesure la distance entre une transaction et le centroïde de son cluster : une plus grande distance implique un indice plus élevé. Les transactions présentant un indice AICAF élevé pour les deux méthodes sont alors considérées comme suspectes.

Récemment, de nombreux chercheurs ont exploré l'utilisation des réseaux de neurones, notamment les auto-encodeurs [BMT06], pour la réduction de dimensionnalité des données. Ces auto-encodeurs apprennent une représentation latente qui est une version réduite et abstraite des données d'origine. Lors de la reconstruction des données, ils parviennent à identifier les données qui diffèrent du modèle normal, les considérant ainsi comme des anomalies potentielles. Kazemi et Zarrami [KZ17] ont formé un auto-encodeur pour extraire l'espace latent des transactions, suivi d'une couche basée sur une fonction softmax, fréquemment employée comme dernière couche en apprentissage profond, pour classifier les transactions en légitimes ou frauduleuses. Deux autres études [PL18, SHK18] ont basé la classification des transactions sur leurs erreurs de reconstruction, associant une erreur élevée à un risque élevé de fraude. Junyi et al. [ZZJ19] ainsi que Misra et al. [MTGS20] ont élaboré des méthodes analogues, en utilisant d'abord un auto-encodeur pour obtenir une représentation latente, puis en employant un réseau neuronal pour classifier les transactions.

Cependant, il est important de noter que l'utilisation récente de réseaux de neurones n'est pas toujours l'option optimale, notamment dans le domaine de la détection de transactions frauduleuses. Dans la majorité des travaux, les résultats ne sont pas comparés avec d'autres algorithmes plus anciens tels que les *Random Forest* ou des plus récents comme CatBoost ou XGBoost. En effet, sur des données tabulaires, les réseaux de neurones n'offrent pas forcément de meilleures performances que ces méthodes. L'apprentissage profond ou par réseaux de neurones est performant lorsqu'il est utilisé pour représenter des données non structurées comme c'est le cas des images ou des textes. Mais, dans des données structurées comme les données tabulaires, les données n'ont pas besoin d'une nouvelle représentation. Shwartz-Ziv et Armon [SZA22] expliquent cela dans leur travail, en affirmant que *« les données tabulaires : l'apprentissage profond n'est pas tout ce qu'il vous faut »*. Grinsztajn et al. [GOV22] répondent dans leur travail à la question *« pourquoi les modèles basées sur les arbres de décision surclassent l'apprentissage profond sur des données tabulaires »*. Ils soulignent également le fait que les

modèles entraînés avec un apprentissage profond sont plus performants sur des données textuelles, des images ou de l'audio que sur des données tabulaires. Ils énoncent trois éléments de réponse : les modèles entraînés avec les algorithmes d'apprentissage profond ont plus de difficulté à prédire des fonctions irrégulières ; les caractéristiques non pertinentes affectent leur apprentissage ; et les données tabulaires ne varient pas avec une rotation des caractéristiques. De plus, un autre point freinant l'utilisation des réseaux de neurones dans le domaine de la détection de fraude financière est l'interprétabilité des modèles. Les modèles basés sur les réseaux de neurones sont souvent plus difficiles à expliquer. Par conséquent, dans notre domaine, l'utilisation de l'apprentissage profond n'est pas la solution à privilégier.

2.2.3.4 Méthodes d'apprentissage semi-supervisé

Les méthodes d'apprentissage semi-supervisé combinent des techniques d'apprentissage supervisé et non supervisé. Plusieurs travaux [LKK10, RH11, CLBC+21, POKB20] ont proposé des méthodes utilisant ces approches pour la détection de transactions frauduleuses. Certains de ces travaux suivent une démarche en deux étapes : d'abord une étape de regroupement des transactions à l'aide de techniques d'apprentissage non supervisé, puis une étape de labellisation des transactions est effectuée en définissant des clusters légitimes et frauduleux. Ensuite, les techniques d'apprentissage supervisé classiques sont utilisées pour classer les transactions. Cette méthode est intéressante, car elle combine des techniques d'apprentissage non supervisé pour labelliser les transactions, suivie de l'utilisation de techniques d'apprentissage supervisé. Pour appliquer cette approche, les auteurs tirent parti des points forts des deux techniques. Cependant, la précision de la labellisation des données dépend largement de la phase non supervisée. Par conséquent, il est essentiel que cette étape soit évaluée par un expert compétent.

Desrousseaux et al. [DBM21] ont présenté une approche pour classifier les activités de blanchiment d'argent à partir de transactions frauduleuses. Ils ont utilisé l'algorithme SOM (Self-Organizing Map) sur un jeu de données de transactions non labellisées pour projeter ces transactions dans une matrice à deux dimensions. L'algorithme SOM est utilisé comme algorithme non supervisé pour regrouper et visualiser les données sur une carte à deux dimensions. En utilisant les valeurs des nœuds de la carte associées à chaque transaction, ils ont attribué une nouvelle représentation à chaque transaction. Ensuite, ils ont utilisé cette représentation pour entraîner un réseau de neurones appelé Fuzzy ART, qui forme des groupes avec les transactions. Ils ont ensuite combiné ces groupes avec la carte de l'algorithme SOM pour obtenir une carte avec différentes régions en fonction de la valeur du nœud associé à chaque groupe. Pour interpréter les résultats, ils ont utilisé deux méthodes : (1) le vecteur pondéré du modèle Fuzzy Art pour chaque type de blanchiment d'argent et la distribution des caractéristiques pour identifier les caractéristiques les plus importantes, et (2) l'analyse des transactions du même type en étudiant la distribution des caractéristiques. Cette approche permet l'identification de différents types de fraude. Cependant, des explications détaillées sur le choix des algorithmes ne sont pas fournies, et en outre, le manque de précision concernant l'implémentation rend difficile la reproduction de l'approche afin de comparer les méthodes.

Ces approches semi-supervisées permettent de tirer parti des avantages des approches supervisées et non-supervisées pour la détection de transactions frauduleuses. Cependant, le choix des algorithmes et des méthodes spécifiques peut varier en fonction des caractéristiques des données et des objectifs de détection de fraude.

2.2.4 Évaluation des modèles

Après l'application des modèles d'apprentissage, la quatrième étape consiste à évaluer la performance de ces modèles. Cette étape permet de mesurer l'efficacité des algorithmes de détection de fraude.

Lors de l'évaluation des modèles, il est important de prendre en compte le déséquilibre des classes, fréquent dans les jeux de données de fraude financières, où les transactions frauduleuses sont généralement rares par rapport aux transactions légitimes. Ce déséquilibre crée des défis lors de l'évaluation d'une tâche de classification binaire, car se concentrer uniquement sur l'exactitude globale peut générer une mauvaise évaluation.

Pour comprendre et évaluer correctement les performances d'un modèle, il est utile de se familiariser avec les concepts suivants [Zhe15] :

— **Vrais positifs (VP)** : Le nombre de prédictions correctes pour les transactions frauduleuses.
— **Faux négatifs (FN)** : Le nombre de prédictions incorrectes pour les transactions frauduleuses prédites comme légitimes.
— **Faux positifs (FP)** : Le nombre de prédictions incorrectes pour les transactions légitimes prédites comme frauduleuses.
— **Vrais négatifs (VN)** : Le nombre de prédictions correctes pour les transactions légitimes.

TABLE 2.1 – Matrice de confusion.

		Réalité	
		Fraude	Légitime
Prédiction	Fraude	VP	FP
	Légitime	FN	VN

Ces valeurs sont représentées dans une matrice de confusion, comme présenté dans le tableau 2.1. Cette matrice est utilisée pour évaluer la précision du modèle en comparant les prédictions faites par le modèle avec les valeurs réelles de l'échantillon de données utilisé pour les tests.

Pour chaque valeur, on peut calculer son taux de la manière suivante :

— **Taux de vrais positifs** (TVP) :

$$TVP = \frac{VP}{VP + FN} \tag{2.1}$$

— **Taux de faux négatifs** (TFN) :

$$TFN = \frac{FN}{VP + FN} \tag{2.2}$$

— **Taux de faux positifs** (TFP) :

$$TFP = \frac{FP}{VN + FP} \tag{2.3}$$

— **Taux de vrais négatifs** (TVN) :

$$TVN = \frac{VN}{VN + FP} \tag{2.4}$$

L'*accuracy* (exactitude) est une mesure qui représente le nombre total de prédictions correctes (vrais positifs et vrais négatifs) divisé par le nombre total d'exemples dans l'ensemble de test. Elle mesure la proportion d'exemples correctement classés par le modèle. Cependant, cette mesure n'est pas adaptée aux jeux de données déséquilibrés, où il y a une grande différence entre le nombre de prédictions correctes et le nombre total d'exemples. Sa formule est la suivante :

$$\textit{accuracy} = \frac{VP + VN}{VP + FP + VN + FN} \tag{2.5}$$

Dans le cas des jeux de données déséquilibrés, la mesure de *précision* est plus pertinente. La précision met en avant la classe positive, c'est-à-dire les cas que l'on souhaite détecter. Dans notre domaine, cela représente le nombre de transactions réellement frauduleuses parmi celles prédites comme frauduleuses par le modèle. Une faible précision entraîne un nombre élevé de faux positifs, ce qui signifie que le modèle prédit à tort des transactions comme frauduleuses, ce qui peut entraîner une charge de travail supplémentaire pour les experts chargés de vérifier ces transactions. Sa formule est la suivante :

$$\textit{précision} = \frac{VP}{VP + FP} \tag{2.6}$$

La mesure de *rappel* (*recall*) met également en valeur la classe positive. Elle représente le nombre de transactions frauduleuses correctement prédites parmi toutes les transactions frauduleuses réelles. En d'autres termes, le rappel mesure la capacité du modèle à détecter les cas positifs. Une valeur élevée de rappel indique que le modèle est capable de trouver un grand nombre de transactions frauduleuses. Sa formule est la suivante :

$$\textit{rappel} = \frac{VP}{VP + FN} \tag{2.7}$$

Le *f1-score* est une mesure qui combine la précision et le rappel en une seule valeur. Il est particulièrement utile pour évaluer la performance d'un modèle sur un jeu de données déséquilibré, où la distribution des classes est inégale.
Le f1-score est calculé en prenant la moyenne harmonique de la précision et du rappel, ce qui permet de prendre en compte à la fois les vrais positifs, les faux positifs et les faux négatifs. Il est souvent utilisé lorsque la précision et le rappel ont une importance équilibrée dans le problème considéré. Sa formule est la suivante :

$$\textit{f1-score} = \frac{2 \times (\textit{précision} \times \textit{rappel})}{\textit{précision} + \textit{rappel}} \tag{2.8}$$

D'autres mesures spécifiques aux modèles basés sur l'apprentissage supervisé ont été développées. Dans notre contexte, un modèle de classification attribue des probabilités d'appartenance

aux classes légitimes et frauduleuses pour chaque transaction. La classe associée à la transaction est déterminée en comparant la probabilité respective avec un seuil fixé à 0.5 par défaut. Cependant, en ajustant ce seuil, il devient possible de calculer les mesures mentionnées précédemment pour différentes valeurs de seuil.

La courbe ROC (Receiver Operating Characteristic) est un graphique représentant le taux de vrais positifs et de vrais négatifs pour différents seuils de classification. Une mesure couramment utilisée pour évaluer les performances d'un modèle de classification est l'aire sous cette courbe, également appelée AUC (Area Under the Curve).

Les modèles de détection de fraude sont conçus pour s'intégrer dans les systèmes bancaires dans le but de réduire les pertes financières. Il est essentiel d'évaluer les coûts liés à l'implémentation de modèles d'apprentissage automatique dans ces systèmes. Ces modèles engendrent des coûts dans deux situations : lorsque le modèle détecte une transaction comme frauduleuse et qu'elle doit être vérifiée par des experts, ainsi que lorsque le modèle ne parvient pas à détecter une transaction frauduleuse.

Dans la littérature, plusieurs mesures de coût ont été introduites dans le domaine de la détection de fraude financière. En 2001, Elkan [Elk01] a proposé une première matrice de coût, qui a été reprise et améliorée par Bahnsen et al. [BSAO13]. Cette matrice de coût, présentée dans le tableau 2.2, associe un coût administratif (C_a) à une transaction prédite comme frauduleuse, afin d'estimer le coût d'un expert pour vérifier si la transaction est réellement frauduleuse. Elle attribue également le montant (Amt_i) de la transaction en tant que coût si une transaction frauduleuse n'a pas été détectée par le modèle. Ces mesures de coût ont été utilisées pour améliorer les algorithmes d'arbres de décision dans des travaux tels que [CAO15] et [MBB+18]. Elles permettent de prendre en compte les conséquences financières réelles associées à la détection de la fraude et aux erreurs de classification.

TABLE 2.2 – Matrice de risque du coût de [BSAO13].

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

Dans le domaine de l'apprentissage automatique, l'évaluation des performances d'un modèle se fait généralement en utilisant des jeux de données labellisées. Pour la détection de transactions frauduleuses, les mesures d'évaluation classiques telles que le rappel, la précision et le f1-score sont couramment utilisées. Cependant, étant donné que la détection de fraude financière peut entraîner des conséquences financières importantes, il est essentiel d'optimiser le coût financier associé au modèle pour les institutions financières. Ainsi, des mesures de coût ont été proposées pour prendre en compte les coûts liés aux fausses détections ou aux détections manquées de transactions frauduleuses. Les mesures de coût sont une extension importante des mesures d'évaluation classiques dans le domaine de la détection de fraude financière.

2.2.5 Interprétabilité des modèles

L'interprétation des modèles d'apprentissage automatique est devenue de plus en plus importante, étant donné leur utilisation croissante dans la prise de décisions critiques dans de nombreux domaines tels que la finance [CPC19] et la médecine [Vel20]. Ces modèles sont complexes et opaques, rendant difficile la compréhension de leur processus décisionnel. L'interprétation des modèles permet aux utilisateurs de comprendre comment ces modèles prennent des décisions, ce qui facilite l'identification des biais ou des erreurs dans les prédictions et contribue à accroître la transparence des modèles. En nous basant sur les travaux de [CPC19], nous avons retenu deux critères que nous allons présenter. Le premier est l'agnosticisme du modèle par rapport à la spécificité du modèle. Le deuxième est la portée d'interprétation du modèle.

2.2.5.1 Interprétation agnostique vs spécifique

Méthode agnostique Les méthodes agnostiques d'interprétabilité des modèles en apprentissage automatique sont des approches génériques qui peuvent être appliquées à différents types de modèles, indépendamment de leur structure interne. Ces méthodes se concentrent principalement sur les relations entre les caractéristiques d'entrée et les prédictions du modèle, plutôt que sur les détails spécifiques du modèle lui-même. Les méthodes agnostiques cherchent à fournir des explications globales et générales sur le fonctionnement du modèle, en se basant sur des mesures d'importance des caractéristiques, des perturbations des données d'entrée ou la construction de modèles approximatifs plus simples. Cette approche permet de mettre en évidence les caractéristiques les plus influentes pour les prédictions du modèle et de comprendre les relations entre ces caractéristiques et les sorties du modèle, sans se soucier spécifiquement de la structure interne du modèle.

Méthode spécifique Les méthodes spécifiques d'interprétabilité sont conçues pour un type de modèle d'apprentissage automatique spécifique. Elles exploitent les propriétés spécifiques du modèle pour fournir des explications détaillées et spécifiques à ce modèle particulier. Les méthodes spécifiques tirent parti des caractéristiques et des mécanismes internes du modèle, tels que les coefficients de régression, les poids des connexions neuronales ou les activations des neurones, pour comprendre comment les entrées sont transformées en sorties et comment les décisions sont prises. Ces techniques permettent de plonger plus profondément dans le fonctionnement du modèle et de fournir des explications spécifiques, mais elles sont souvent limitées à un seul type de modèle et ne peuvent pas être facilement généralisées à d'autres architectures.

2.2.5.2 La portée d'interprétation

Il existe deux catégories de portée d'interprétation des modèles. Les méthodes globales d'interprétation des modèles sont utilisées pour comprendre le comportement global d'un modèle d'apprentissage automatique. Les méthodes locales d'interprétation de modèles sont utilisées pour comprendre les prédictions individuelles du modèle. Ces deux catégories de méthodes de l'interprétation des modèles ont été détaillées par Christoph Molnar dans son livre *« Interpretable Machine Learning. »* [Mol19].

Les méthodes globales Elles reposent sur des techniques visant à mesurer l'importance des caractéristiques du modèle en fonction de leur contribution à la prédiction finale. Les méthodes globales permettent d'identifier les caractéristiques les plus importantes pour le modèle, à comprendre les relations entre les caractéristiques et de détecter les éventuels biais dans les prédictions. Parmi les méthodes globales les plus couramment utilisées, on trouve les graphiques de dépendance partielle. Ces graphiques permettent d'observer comment la prédiction varie en fonction de chaque caractéristique individuellement, tout en marginalisant les autres caractéristiques. Les graphiques d'effet de caractéristiques accumulé sont également utilisés pour visualiser l'effet cumulatif des caractéristiques sur la prédiction. D'autres méthodes globales comprennent la décomposition fonctionnelle, qui décompose la prédiction en termes de contributions de caractéristiques individuelles, et l'importance des caractéristiques par permutation, qui évalue l'importance de chaque caractéristique en échangeant les valeurs de la caractéristique avec des valeurs aléatoires et en mesurant l'impact sur la prédiction. Les méthodes globales sont particulièrement utiles pour comprendre le fonctionnement global du modèle et pour identifier les relations entre les caractéristiques du modèle.

Les méthodes locales Elles fournissent des explications ciblées sur les caractéristiques spécifiques qui ont contribué à ces prédictions. Elles permettent d'identifier les caractéristiques qui ont eu le plus d'influence sur la prédiction et de comprendre comment ces caractéristiques ont contribué à celle-ci. Parmi les méthodes locales les plus couramment utilisées, on retrouve les courbes d'espérance conditionnelle individuelle. Ces courbes décrivent comment la prédiction évolue lorsque la valeur d'une caractéristique est modifiée. Les modèles de substitution locaux sont également utilisés. Ils remplacent le modèle d'apprentissage automatique par un modèle plus simple et interprétable afin d'expliquer la prédiction. Les règles ancrées, qui sont des règles qui décrivent les valeurs de caractéristiques spécifiques qui influencent une prédiction donnée. D'autre part, les explications contrefactuelles décrivent comment les caractéristiques d'une instance devraient être modifiées pour obtenir une prédiction différente. Les méthodes locales, telles que les valeurs Shapley [MT20] qui attribuent équitablement la contribution de chaque caractéristique pour une prédiction, sont particulièrement utiles pour comprendre les contributions spécifiques de chaque caractéristique à une prédiction individuelle.

2.2.6 Synthèse

La détection des transactions frauduleuses est essentielle dans le domaine financier, car les méthodes existantes ne répondent pas toujours aux exigences d'efficacité des institutions financières. Elles doivent faire face à deux problèmes majeurs : un taux élevé de faux positifs, les obligeant à solliciter des experts pour vérifier les transactions détectées, et un faible taux de détection des transactions frauduleuses, ce qui les expose à d'importantes pertes financières.

L'acquisition de données dans ce domaine est difficile en raison de la confidentialité des données et des réglementations strictes. Les jeux de données financiers disponibles présentent souvent des lacunes, telles que le manque des scénarios de fraude ou des caractéristiques anonymes. Ces jeux de données financiers se caractérisent par des attributs de base tels que le montant, les acteurs et la date, qui peuvent être utilisés pour détecter les fraudes [KDT20, PM18].

Les données financières avec leurs attributs de base ne suffisent pas à identifier des transactions frauduleuses. L'ingénierie des caractéristiques est une étape importante dans la détection

de fraude, car elle permet de mieux comprendre les données et de détecter les schémas de fraude cachés. Les techniques courantes d'ingénierie des caractéristiques incluent la normalisation, la transformation des données et la création de nouvelles caractéristiques à partir de combinaisons d'attributs existants.

À notre connaissance, il n'existe pas encore de travaux spécifiques pour les transactions SWIFT, qui sont utilisées pour les transferts internationaux et interbancaires de fonds entre institutions financières. Des techniques d'ingénierie des caractéristiques spécifiques aux transactions SWIFT permettraient de mieux utiliser les modèles d'apprentissage automatique.

Les méthodes basées sur les techniques d'apprentissage automatique ont fait leurs preuves dans la détection de transactions frauduleuses [LSA+20, ZS+15, VKS+19]. L'utilisation de techniques d'apprentissage supervisé permet d'apprendre des schémas de fraude dans des jeux de données labellisés pour les détecter sur de nouvelles données. Les techniques d'apprentissage non supervisé, initialement utilisées pour regrouper des données partageant des caractéristiques similaires, peuvent également être utilisées pour la labellisation de données, puis pour appliquer des techniques d'apprentissage supervisé. La littérature propose des méthodes pour détecter différents types de fraude dans des ensembles de données non étiquetés, ouvrant ainsi de nouvelles perspectives pour l'étude des transactions frauduleuses. Dans ce contexte, les modèles sont généralement testés et validés en utilisant des ensembles de données pré-étiquetés, où chaque échantillon est associé à une étiquette ou à une classe. Dans la détection de transactions frauduleuses, plusieurs métriques de performance sont privilégiées pour évaluer la justesse des prédictions. Parmi ces métriques, le rappel (qui mesure la capacité du modèle à identifier correctement les fraudes), la précision (qui évalue la pertinence des transactions identifiées comme frauduleuses) et le f1-score (une moyenne harmonique de la précision et du rappel) sont fréquemment employés pour obtenir une évaluation complète des performances du modèle.

Des mesures de coûts ont été proposées afin d'optimiser le modèle afin de minimiser son coût financier pour les institutions financières [BSAO13]. Il est important de noter que la performance des modèles d'apprentissage automatique peut varier en fonction des jeux de données utilisés, de la qualité des données et des techniques d'ingénierie des caractéristiques utilisées.

Après avoir étudié les diverses techniques d'apprentissage automatique utilisées pour détecter les transactions frauduleuses, ainsi que les métriques pertinentes pour évaluer leur performance, il est important de se pencher sur d'autres approches qui peuvent compléter ou améliorer ces méthodes. Parmi ces alternatives, les approches basées sur les ontologies se distinguent par leur capacité à modéliser les connaissances dans le domaine de la détection de fraude. Dans la section qui suit, nous étudions les approches basées sur les ontologies dans le contexte de la détection de transactions frauduleuses.

2.3 Approches basées sur les ontologies

Dans cette section, nous abordons en premier lieu l'utilisation des ontologies pour décrire le domaine de la détection de fraude financière. Ensuite, nous présentons des travaux axés sur le peuplement des ontologies. Enfin, nous étudions les approches qui utilisent des règles pour détecter les fraudes, en exploitant la structure et les contraintes de l'ontologie.

Ces deux dernières décennies, l'explosion du volume de données sur le Web a posé de nouveaux défis en termes de recherche, de traitement et d'interprétation de ces données. Initialement constitué principalement de pages Web statiques destinées à la consultation humaine, le Web s'est rapidement diversifié avec l'apparition de nombreuses sources d'information, ce qui a rendu indispensable la compréhension des documents par les machines.

Le besoin de structurer les informations et de les rendre compréhensibles par les systèmes informatiques a conduit à l'émergence du Web sémantique. Ce dernier offre une approche permettant de structurer les informations et de faciliter leur interprétation par les systèmes informatiques. Le Web sémantique vise à enrichir le contenu du Web en ajoutant des métadonnées sémantiques aux documents, ce qui permet aux machines de comprendre le sens et la signification des informations qu'ils renferment. Ces métadonnées sémantiques décrivent les propriétés et les relations des ressources présentes sur le Web, favorisant ainsi des recherches plus précises, des inférences et une interconnexion des connaissances dispersées.

Dans le contexte du Web sémantique, les ontologies sont des structures de données formelles qui décrivent de manière explicite les concepts, les relations et les propriétés d'un domaine spécifique. Elles servent à représenter et à organiser les connaissances de manière cohérente et unifiée, facilitant ainsi l'interprétation et l'échange de données entre les systèmes informatiques. Les ontologies décrivent les termes et les relations entre ces termes à l'aide d'un langage formel, tel que le langage de représentation des connaissances OWL (Web Ontology Language).

Une ontologie est constituée de concepts, qui représentent des entités abstraites du domaine, ainsi que de relations qui définissent les liens entre ces concepts. Ces relations peuvent définir des hiérarchies ou des relations sémantiques (comme la relation « est associé à »). Deux éléments fondamentaux des ontologies sont les axiomes et les règles de raisonnement. Nous allons d'abord définir ces deux notions, puis expliquer leurs distinctions.

Les axiomes sont des déclarations formelles qui jouent un rôle essentiel dans la structuration du domaine de connaissance, en permettant de poser des contraintes sur les concepts (entités, classes) et d'établir des relations claires entre ces concepts. Les axiomes sont également utilisés pour créer de nouveaux concepts à partir de concepts existants, en formulant des conditions précises qui doivent être satisfaites. Par exemple, on pourrait formaliser un concept d'« Acteur Frauduleux » comme étant un individu (ou une entité) dont au moins une des transactions a été identifiée comme frauduleuse. Cet axiome peut être exprimé en logique de description comme suit :

$$ActeurFrauduleux \equiv \exists aEffectue.TransactionFrauduleuse$$

Les règles permettent d'exprimer des déductions logiques qui peuvent être effectuées à partir des connaissances déjà définies dans cette ontologie. En utilisant un format de condition-action

exprimée dans des langages comme SWRL (Semantic Web Rule Language), elles permettent de déclarer explicitement les inférences qui doivent être faites sous certaines conditions. Par exemple, une règle peut définir que si une transaction est réalisée par un acteur qui est client d'une banque suspecte alors cette transaction est bloquée, ce qui peut être exprimé comme suit :

$Transaction(?t) \wedge aActeur(?t, ?a) \wedge clientDe(?a, ?b) \wedge BanqueSuspecte(?b) \Rightarrow TransactionBloquee(?t)$

Un autre aspect important des ontologies concerne le requêtage couramment utilisé avec le langage SPARQL, il permet de formuler des requêtes expressives et précises afin d'interroger les informations de l'ontologie. Cela facilite la recherche, l'exploration et l'extraction des connaissances représentées dans l'ontologie.

2.3.1 Modélisation du domaine avec les ontologies

Les ontologies fournissent une représentation structurée des connaissances d'un domaine, permettant aux machines de raisonner, de rechercher, de déduire de nouvelles informations à partir des données disponibles. Elles sont essentielles dans la gestion des connaissances, la résolution de problèmes complexes et la facilitation de l'interopérabilité des systèmes d'information. Dans le domaine de la détection de fraude financières, des travaux ont été réalisés pour représenter ce domaine en utilisant des ontologies.

En 2004, Kingston et al. [KSV04] ont proposé une approche visant à créer une ontologie intégrant à la fois des aspects juridiques et financiers. Pour ce faire, ils ont combiné deux ontologies financières : l'ontologie SUMO (Suggested Upper Merged Ontology) [NP01] et l'ontologie REA (Resources-Events-Agents) de Geerts et McCarthy [GM02]. L'ontologie proposée décrit les activités commerciales d'une entreprise en termes de ressources, d'événements et d'agents. Elle est utilisée dans le domaine de la comptabilité et de la gestion des entreprises pour formaliser les règles comptables et fiscales, ainsi que pour garantir la conformité réglementaire [CS08].

En 2022, Hussaini et al. [HGL22] ont présenté leur travail portant sur l'élaboration d'une ontologie décrivant neuf types de fraude financières. Chaque type de fraude est présenté successivement, en détaillant les concepts et propriétés créés. Bien que les détails spécifiques de l'ontologie développée dans cette étude soient privés et non divulgués, cette approche illustre comment les ontologies peuvent être employées pour décrire de manière ciblée le domaine particulier de la détection de fraude financières.

2.3.2 Peuplement des ontologies

Le peuplement d'une ontologie est le processus d'ajout d'instances dans cette ontologie. Le peuplement peut être réalisé de différentes manières. L'une des approches couramment utilisées est l'extraction d'informations à partir de sources de données structurées ou non structurées telles que des bases de données, des fichiers textuels, des documents en ligne, des flux de données en temps réel. Les données extraites sont ensuite transformées et alignées sur les concepts et les propriétés de l'ontologie, garantissant ainsi une intégration cohérente.

Dans le domaine de la détection de fraude financière, des travaux ont été proposés pour peupler les ontologies avec des connaissances provenant de sources externes, notamment des articles de journaux financiers. Deux travaux [AMPK18, SSS18] se concentrent sur l'analyse et l'identification des termes significatifs impliqués dans les fraudes. La première approche [AMPK18] met l'accent sur la nécessité d'un système bancaire plus sécurisé en soulignant l'importance de comprendre le domaine de la fraude et de créer une base de connaissances à cet égard. Des techniques d'analyse de texte, telles que la méthode de pondération TF-IDF, sont utilisées pour identifier les termes spécifiques à des types de fraude.

La deuxième approche [SSS18] se concentre sur la criminologie financière. Les auteurs proposent une représentation ontologique de la criminologie financière afin de capturer les termes et les sujets fréquemment discutés dans la recherche. L'utilisation d'un outil d'analyse de texte permet d'extraire et d'identifier les termes couramment utilisés. La validité de la représentation ontologique est confirmée par des experts en criminologie financière et des auditeurs. Les résultats de la recherche mettent en évidence les classes identifiées et fournissent un aperçu des termes et des sujets clés dans le domaine de la criminologie financière.

Ces deux approches permettent de peupler une ontologie qui modélise le domaine de la détection de fraude en identifiant les termes les plus utilisés pour des types de fraude. Cependant, les auteurs ne précisent pas comment cette identification peut être utilisée avec des outils de détection de fraude.

Dans le cadre de notre collaboration avec l'entreprise SKAIZen Group, en 2020 nous avons créé une base de connaissance à l'aide d'une ontologie ayant pour but de surveiller les activités financières des clients des institutions financières [JSZC20]. Les experts ont défini une liste de concepts et propriétés permettant de modéliser un client ainsi que ses informations pouvant être utilisées pour une institution financière, telle que son lieu de résidence ou l'entreprise où il travaille. L'ontologie a été peuplée à partir d'un module basé sur les techniques de langage naturel pour extraire les informations de journaux financiers. Cette base de connaissances a pour but de faciliter la collection d'informations sur les clients lors du contrôle de transactions frauduleuses. Il est intéressant de se baser sur des sources externes

2.3.3 Les règles d'inférence

Les règles dans l'ontologie constituent un aspect fondamental dans l'exploitation des ontologies. Les ontologies offrent une représentation formelle des connaissances et des relations entre les concepts, les propriétés et les instances d'un domaine spécifique. Cependant, il est nécessaire d'effectuer un raisonnement logique sur les connaissances contenues dans l'ontologie pour différentes raisons, telles que la vérification de la cohérence des connaissances ou la déduction de nouvelles connaissances et de faire des inférences basées sur des règles et des axiomes.

L'un des aspects importants du raisonnement consiste à inférer des connaissances implicites à partir des connaissances explicites représentées dans l'ontologie. Par exemple, si une transaction suspecte est liée à un compte bancaire frauduleux, l'ontologie peut contenir des règles d'inférence pour déduire que d'autres transactions liées à ce compte sont également suspectes.

Les règles d'inférence peuvent être exprimées dans différents langages, et l'un des langages utilisés pour écrire des règles est SWRL (Semantic Web Rule Language). SWRL est un langage de règles qui permet d'exprimer des connaissances sous forme de règles logiques.

Les règles SWRL se composent généralement d'une précondition et d'une conséquence. La précondition spécifie les conditions nécessaires pour que la règle s'applique, tandis que la conséquence spécifie les nouvelles connaissances qui peuvent être inférées lorsque la règle est satisfaite. Par exemple, une règle SWRL peut être formulée comme suit :

$$TransactionSuspecte(?t) \wedge aActeur(?t, ?a) \wedge CompteFrauduleux(?a) \Rightarrow TransactionBloquee(?t)$$

Cette règle indique que si une transaction suspecte est liée à un compte frauduleux, alors la transaction est bloquée.

Les moteurs d'inférence, tels que Pellet [SPG+07], HermiT [SMH08] ou Racer[HM01], sont largement utilisés pour effectuer le raisonnement sur les ontologies. Ces moteurs utilisent des algorithmes d'inférence pour parcourir la structure de l'ontologie, appliquer les règles de raisonnement définies et générer de nouvelles connaissances implicites. Ainsi, les règles SWRL permettent d'automatiser le processus de détection de fraude en déduisant des informations supplémentaires à partir des connaissances déjà présentes dans l'ontologie.

Certains travaux ont exploré l'utilisation des règles et du raisonnement des ontologies pour détecter des transactions frauduleuses [EOBA18, RKLH14, CXD20]. En se basant sur des caractéristiques des transactions ou des acteurs, ils réalisent des règles avec des seuils pour définir si une transaction est frauduleuse. Cependant, cette utilisation des ontologies ne répond pas de manière adéquate aux problèmes des systèmes de détection de fraude traditionnels qui sont basés sur des règles également. En se limitant à l'utilisation de règles basées sur les caractéristiques de base des transactions, les ontologies n'apportent aucune valeur ajoutée significative à la détection des fraudes financières.

Ahmed et al. [AAM+21] ont proposé une méthode en trois étapes pour la détection de fraude basée sur les ontologies. Leur approche commence par l'extraction de données à partir d'une base de données relationnelle, puis ces données sont prétraitées et enregistrées dans une base de connaissances. Dans la deuxième étape, un seuil est calculé pour chaque compte et utilisé par le moteur d'inférence lors de l'application des règles à chaque transaction à ce compte. Les règles utilisées sont définies dans l'ontologie et permettent de déterminer les critères du niveau de gravité d'une alerte suspecte. Les alertes générées présentent différents niveaux de gravité, allant de simples suspicions à des cas de fraude avérée.

2.3.4 Synthèse

Les ontologies sont des outils importants pour la représentation structurée des connaissances d'un domaine, la détection de fraude financières. Différents travaux ont été réalisés pour créer des ontologies spécifiques à ce domaine. Par exemple, une ontologie intégrant des aspects juridiques et financiers a été proposée en 2004 [KSV04], décrivant les activités commerciales d'une entreprise. Une autre ontologie, élaborée en 2022, décrit neuf types de fraude financière [HGL22].

Le peuplement d'une ontologie consiste à ajouter des données à une ontologie existante. Différentes approches peuvent être utilisées, telles que l'extraction d'informations à partir de sources de données structurées ou non structurées. Dans le domaine de la détection de fraude financières, des travaux ont été réalisés pour peupler les ontologies avec des connaissances provenant de sources externes, notamment des articles de journaux financiers [AMPK18, SSS18]. Ces travaux identifient les termes significatifs liés aux fraudes et élaborent une représentation ontologique de la criminologie financière.

Le raisonnement sur les ontologies est essentiel pour exploiter les connaissances contenues dans une ontologie ou déduite par les règles d'inférence, exprimées notamment dans le langage SWRL. Les moteurs d'inférence, tels que Pellet, HermiT ou Racer, sont utilisés pour effectuer le raisonnement sur les ontologies de détection de fraude.

Certains travaux ont exploré l'utilisation des règles et du raisonnement des ontologies pour détecter des transactions frauduleuses, en se basant sur des caractéristiques des transactions ou des acteurs. Cependant, ces approches ne fournissent pas de valeur ajoutée significative par rapport aux systèmes de détection de fraude traditionnels basés sur des règles [EOBA18, RKLH14, CXD20]. Une méthode en trois étapes a été proposée pour la détection de fraude basée sur les ontologies, impliquant l'extraction de données, le calcul de seuils et l'application des règles d'inférence pour générer des alertes de différentes gravités [AAM+21].

En conclusion, les ontologies servent à la détection des fraudes financières en représentant le domaine concerné et en facilitant le raisonnement logique basé sur ces connaissances. Toutefois, il n'existe à ce jour aucune ontologie complète et publique dédiée à la détection de transactions frauduleuses. Les approches fondées sur des règles, quant à elles, semblent ne pas apporter de valeur aux systèmes en place. Ainsi, dans la section suivante, nous étudions les approches hybrides qui combinent les techniques d'apprentissage automatique et les ontologies, afin d'examiner comment ils peuvent s'associer pour renforcer les systèmes actuels de détection de fraude.

2.4 Approches hybrides

Afin d'améliorer l'efficacité des systèmes de détection de fraude, les chercheurs et les praticiens explorent des approches hybrides qui combinent l'apprentissage automatique et les ontologies. Cette combinaison permet de tirer parti des avantages des deux approches : l'apprentissage automatique offre la capacité d'analyser de vastes volumes de données et de détecter des schémas complexes, tandis que les ontologies fournissent un cadre de représentation sémantique formel pour modéliser les connaissances et les règles spécifiques au domaine des fraudes financières.

Les premiers travaux de la littérature sur la détection de fraude financière en utilisant les méthodes hybrides sont récents datant de 2018. Deux approches [TLYW18] et [EOB19] similaires ont été proposées pour combiner les ontologies et l'apprentissage automatique. Les approches partagent un objectif commun qui consiste à intégrer des règles de raisonnement issues d'un modèle entraîné avec l'algorithme des arbres de décision.
Les deux méthodes utilisent les chemins des arbres de décision, qui représentent une séquence de nœuds parcourus depuis la racine jusqu'à une feuille, en suivant les règles de décision ba-

sées sur les caractéristiques des transactions. Dans ces travaux, les règles de décision menant à une transaction frauduleuse sont extraites. Ensuite, ces règles sont converties au format SWRL et intégrées à l'ontologie correspondante.
Par la suite, les ontologies sont peuplées avec des transactions, et un moteur de raisonnement est utilisé pour classifier les transactions frauduleuses associées à un chemin frauduleux dans l'arbre de décision. Cependant, cette méthodologie présente certaines limites et l'utilisation des ontologies n'est pas justifiée. En effet, l'utilisation de l'ontologie permet d'inférer à partir des règles pour classifier les transactions, mais le modèle entraîné par l'arbre de décision peut également réaliser cette tâche, d'autant plus que le raisonnement sur de grandes quantités de données peut s'avérer gourmand en ressources.

Une méthode de détection d'anomalies a été présentée par [SDPH$^+$21] fonctionnant en trois phases. Dans la première phase, des techniques basées sur l'apprentissage et les connaissances sont utilisées pour détecter et interpréter les anomalies à partir des flux de données et des données contextuelles. Les anomalies détectées sont ensuite affichées dans un tableau de bord où l'utilisateur peut fournir des commentaires. Dans la troisième phase, les informations relatives aux anomalies détectées, les commentaires de l'utilisateur et les méta-informations contextuelles sont utilisées pour améliorer les techniques de détection basées sur les données et les connaissances. De nouvelles connaissances sont extraites du système de stockage des connaissances et utilisées pour mettre à jour automatiquement les outils de détection. L'approche présentée dans cet article offre des avantages significatifs dans la détection d'anomalies et de défauts, ainsi que dans la vérification de leurs causes. En combinant à la fois des techniques basées sur les données et des techniques basées sur la connaissance, les auteurs parviennent à exploiter les points forts de chaque approche. Les techniques basées sur les données sont capables de s'adapter aux données brutes, tandis que les techniques basées sur la connaissance fournissent des explications sur les causes des défauts détectés. En intégrant des connaissances sémantiques dans leur technique d'apprentissage automatique, ils améliorent l'expressivité du système, lui permettant de mieux comprendre et représenter des informations complexes. L'adaptabilité du système est également renforcée grâce aux méthodes d'extraction de règles sémantiques à partir des retours sur les défauts et les anomalies. Cependant, il convient de noter que cette approche présente également certaines limites. La mise en place des techniques basées sur la connaissance demande souvent un effort humain considérable pour définir les règles et les connaissances spécifiques au domaine. La méthodologie proposée optimise les techniques en fonction des retours d'utilisateurs, ce qui peut restreindre la diversité et la représentativité des informations disponibles. De plus, les évaluations réalisées se limitent à un cas d'étude spécifique dans le domaine de la maintenance prédictive des trains, ce qui soulève des questions quant à la généralisation des résultats à d'autres domaines ou applications.

En résumé, les travaux de recherche se concentrant sur l'utilisation des approches hybrides demeurent limités. Il existe un manque de ressources et de méthodologies spécifiques pour guider les institutions financières dans la mise en œuvre de telles approches dans un contexte de détection de fraude financières.

2.5 Synthèse générale

La détection de transactions frauduleuses dans le domaine de la finance est un enjeu majeur, mais les institutions financières sont confrontées à des systèmes de détection de fraude peu efficaces. Deux problèmes principaux se posent : un taux élevé de faux positifs et un faible taux de transactions frauduleuses détectées [Kno20]. Ces problèmes exposent les institutions financières à des coûts élevés, à la fois en termes de rémunération d'experts pour vérifier les transactions suspectes et en termes de lourdes amendes en cas de transactions frauduleuses non détectées.

Pour développer de nouvelles approches de détection de fraude, l'utilisation de données financières est essentielle. Cependant, l'acquisition de telles données présente des défis en raison de la confidentialité et des réglementations strictes. Les jeux de données financiers disponibles sont souvent lacunaires, avec un manque de scénarios de fraude et des caractéristiques anonymes. Ces jeux de données se caractérisent par des attributs de base tels que le montant, les acteurs et la date, qui peuvent être utilisés pour détecter les fraudes [KDT20, PM18]. Cependant, ces données de base ne suffisent pas à identifier les transactions frauduleuses. Il est donc nécessaire d'utiliser des techniques d'ingénierie des caractéristiques pour mieux comprendre les données et détecter des schémas de fraude cachés. Ces techniques comprennent la normalisation, la transformation des données et la création de nouvelles caractéristiques à partir de combinaisons d'attributs existants.

Un défi spécifique dans la détection de fraude financière concerne les transactions SWIFT, qui sont utilisées pour les transferts internationaux de fonds entre institutions financières. À ce jour, il n'existe pas de travaux spécifiques sur l'ingénierie des caractéristiques pour les transactions internationales et interbancaires. La recherche future dans ce domaine pourrait permettre de mieux comprendre les fraudes circulant à travers ce réseau et d'améliorer les systèmes de lutte contre les fraudes financières internationales et interbancaires.

Les méthodes basées sur l'apprentissage automatique ont montré leur efficacité dans la détection de transactions frauduleuses. L'utilisation de techniques d'apprentissage supervisé permet d'apprendre les schémas de fraude à partir de jeux de données labellisés, tandis que les techniques d'apprentissage non supervisé peuvent être utilisées pour la labellisation de données, puis pour appliquer des techniques d'apprentissage supervisé. Les travaux de recherche ont proposé différentes approches pour identifier les types de fraude dans des jeux de données non labellisés, ouvrant ainsi des perspectives intéressantes pour l'analyse des transactions frauduleuses. Les algorithmes d'apprentissage supervisé basé sur les arbres de décisions ont prouvé leur efficacité dans la tâche de classification de transactions frauduleuses et légitimes [LSA+20, ZS+15, VKS+19].

L'évaluation des modèles d'apprentissage automatique se fait à l'aide de jeux de données labellisés. Les mesures classiques utilisées dans le domaine de la détection de fraude comprennent le rappel, la précision et le f1-score. Le rappel mesure la proportion de vrais positifs détectés par rapport à tous les vrais positifs, tandis que la précision mesure la proportion de vrais positifs détectés par rapport à tous les éléments détectés comme positifs par le modèle. Le f1-score est une mesure combinée de rappel et de précision. Cependant, ces mesures ne prennent pas en compte les coûts réels associés à la détection de fraude. Dans les systèmes de détection

de fraude, des mesures de coûts spécifiques peuvent être utilisées pour optimiser les modèles et minimiser leur impact financier. Ces mesures prennent en compte le coût de la vérification des transactions suspectes, le coût des erreurs de classification et les amendes potentielles en cas de transactions frauduleuses non détectées. Une méthode pour estimer les coûts de prédiction d'un modèle pour une institution financière a été proposée avec le coût du contrôle d'une transaction ayant généré une alerte et avec le coût des transactions frauduleuses non détectées [BSAO13]. Ces travaux permettent d'ouvrir le sujet sur la quantification des coûts financier d'un modèle pour les institutions financière, qui ont pour but de maximiser l'efficacité des systèmes de détection de fraude tout en minimisant leur coût financier. Il existe un manque de connaissances exploitées sur les approches basées sur les techniques d'apprentissage automatique pour des raisons à la fois légales, avec la difficulté d'intégrer des informations privées sur les clients dans les modèles, et avec les problématiques de données manquantes.

C'est dans ce cadre que nous avons étudié les ontologies qui sont des outils importants pour représenter les connaissances dans le domaine de la détection de fraude financières. Elles permettent de modéliser les relations entre les concepts et d'organiser les connaissances de manière structurée. Le peuplement des ontologies peut être réalisé en extrayant des informations à partir de sources de données structurées ou non structurées. Par exemple, des techniques de fouille de textes peuvent être utilisées pour extraire des informations à partir de journaux financiers [JSZC20]. Le raisonnement sur les ontologies permet d'exploiter les connaissances contenues dans une ontologie et de déduire de nouvelles connaissances à partir des connaissances explicites. Cependant, des travaux supplémentaires sont nécessaires pour améliorer l'efficacité de la détection de fraude basée sur les ontologies, notamment en ce qui concerne l'incorporation de règles de raisonnement ou d'axiomes spécifiques à la détection de fraude.

Enfin, des approches hybrides combinant l'apprentissage automatique et les ontologies sont également explorées pour améliorer l'efficacité des systèmes de détection de fraude. Ces approches ne sont pas assez convaincantes, et l'utilisation des ontologies n'est pas toujours justifiée [TLYW18, EOB19]. Cependant, ces approches offrent un potentiel intéressant pour améliorer la performance des systèmes de détection de fraude en exploitant les avantages des deux approches. Les approches hybrides nécessitent davantage de recherches et de ressources pour être mises en œuvre dans le contexte de la détection de fraude financières.

En conclusion, la détection de transactions frauduleuses dans le domaine de la finance est un défi complexe. Les défis comprennent l'acquisition de données financières complètes et anonymes, l'ingénierie des caractéristiques pour la détection de modèles frauduleux, l'utilisation d'approches d'apprentissage automatique pour la détection de fraude, l'évaluation des modèles en tenant compte des coûts réels et l'exploitation des ontologies et des approches hybrides pour améliorer la détection de fraude. Ces domaines nécessitent des recherches continues pour développer des solutions plus efficaces dans la lutte contre les fraudes financières.

C'est dans ce contexte que nous proposons notre approche hybride de détection de fraude dans le réseau SWIFT. Nous utilisons les techniques d'apprentissage automatique pour classifier les transactions faisant partie de schémas de fraude identifiés, tout en nous intéressant à l'analyse de ces schémas et à la minimisation des coûts financiers des modèles d'apprentissage automatique. Dans un second temps, nous faisons appel aux ontologies pour apporter une couche sémantique aux transactions suspectes détectées par notre modèle. De cette manière, le

volume des transactions à intégrer dans notre ontologie est réduit, ce qui facilite les raisonnements et les inférences. Nous utilisons le raisonnement sémantique comme aide à la décision concernant les transactions suspectes. Nous pouvons ainsi décider de libérer ou de bloquer ces transactions, tout en tirant profit d'autres connaissances provenant de sources externes.

Nous présentons dans le chapitre suivant notre approche de détection de fraude financière dans le réseau international et interbancaire SWIFT.

Chapitre 3

Approche hybride pour la détection de transactions frauduleuses et leur analyse

Sommaire

3.1 Introduction

Les systèmes actuels de détection de transactions frauduleuses sont confrontés à des limitations en raison de leur dépendance à des règles simples. Ces règles sont souvent contournées par les fraudeurs, ce qui entraîne à la fois une détection limitée des transactions frauduleuses et un taux élevé de fausses alertes.
Pour remédier à ces problèmes, différentes approches ont été proposées, notamment celles basées sur les techniques d'apprentissage automatique, l'utilisation des ontologies, ainsi que les approches hybrides combinant les techniques précédentes.

Cependant, après une analyse approfondie de la littérature, nous avons constaté que les approches basées sur les techniques d'apprentissage automatique nécessitent une adaptation aux transactions SWIFT, en particulier en ce qui concerne le calcul et la sélection des caractéristiques. D'autre part, les approches basées sur les ontologies du domaine reposent sur le même principe que les systèmes actuels, c'est-à-dire l'utilisation des règles, basées sur les caractéristiques des transactions. Nous n'avons pas identifié d'approches hybrides adaptées et efficaces intégrant des connaissances avec des techniques d'apprentissage automatique pour la détection de transactions frauduleuses.

Nous présentons notre approche qui vise à résoudre ces problèmes et à atteindre les objectifs suivants :

- Classer les transactions suspectes en fonction de schémas de fraude spécifiques.
- Identifier et analyser les schémas de fraude au sein des transactions classifiées comme frauduleuses.
- Réduire le coût financier de notre modèle
- Intégrer les données des transactions considérées comme suspectes avec des informations sur les acteurs impliqués.

Pour atteindre ces objectifs, nous proposons une approche hybride combinant les techniques d'apprentissage automatique et l'utilisation des ontologies. La méthode de notre approche est illustrée dans la figure 3.1, et elle comprend cinq étapes :

1. **Analyse du jeu de données** : Cette étape est une analyse exploratoire où le jeu de données est présenté et analysé afin de comprendre les transactions qu'il contient.
2. **Extraction des caractéristiques** : Un ensemble de caractéristiques est calculé à partir des caractéristiques de base des transactions, telles que les acteurs, les pays et les devises. Ensuite, nous utilisons une méthode de sélection de caractéristiques pour ne conserver que les plus pertinentes qui permet de distinguer les transactions légitimes des frauduleuses.
3. **Clustering** : Nous regroupons les transactions frauduleuses qui partagent des valeurs de caractéristiques similaires.
4. **Classification** : Dans cette étape, nous entraînons un modèle de classification pour classifier les transactions dans les classes légitimes et frauduleuses. Ensuite, nous proposons une méthode de calcul d'un score de suspicion en utilisant les prédictions du modèle. Nous sélectionnons ensuite un seuil de suspicion en combinant les mesures d'évaluation et les considérations de coût. Enfin, nous utilisons des techniques d'interprétation de modèles pour identifier et analyser les schémas de fraude.

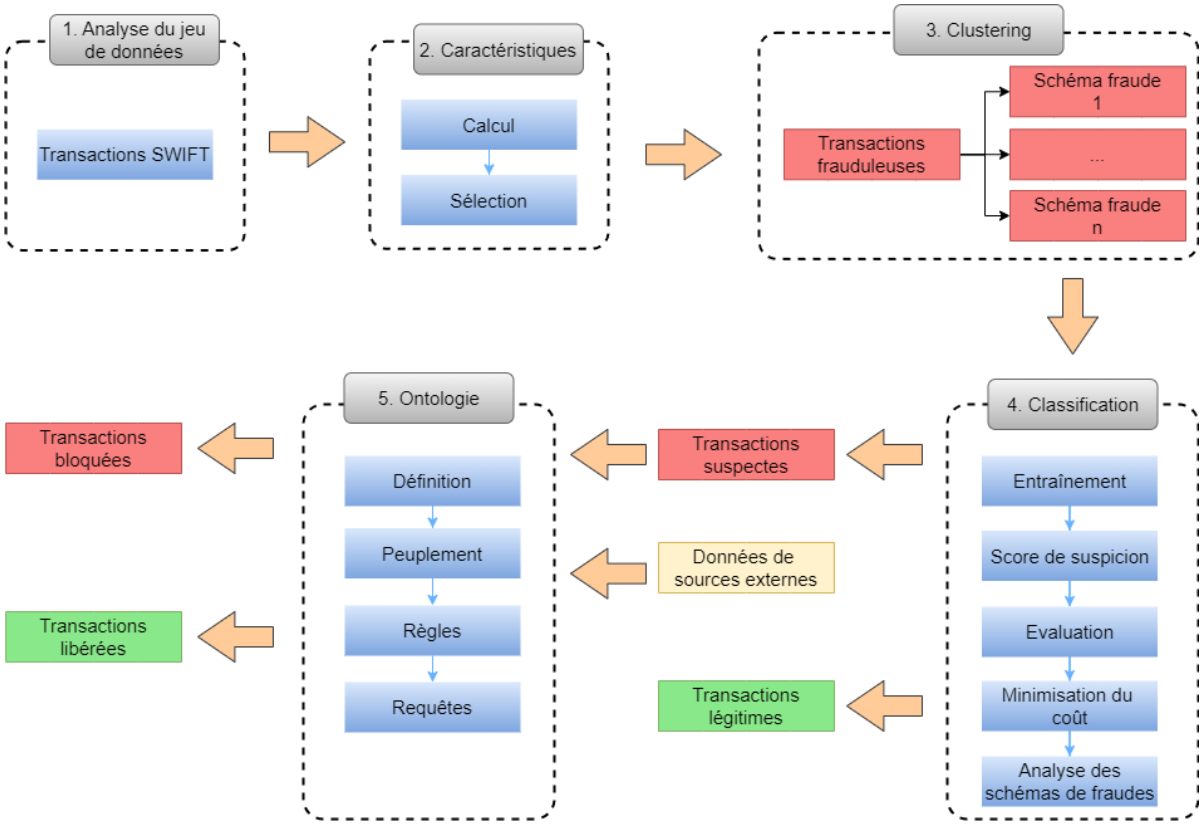

FIGURE 3.1 – Méthodologie de l'approche

5. **Utilisation de l'ontologie** : La dernière étape consiste à analyser les transactions suspectes prédites par le modèle de classification. Pour ce faire, nous utilisons les informations du modèle (score de suspicion, schémas de fraude) ainsi que des informations provenant de sources externes (journaux financiers, réseaux sociaux). L'ontologie, qui est une base de connaissances sur les acteurs et les transactions, facilite la décision de l'expert, fournissant davantage d'informations. Nous utilisons des règles qui sont utilisées pour décider de libérer ou de bloquer automatiquement certaines transactions suspectes, tandis que d'autres transactions font l'objet à une étude plus approfondie par les experts.

3.2 L'analyse du jeu de données

Dans cette section, nous examinons en détail les données essentielles à notre approche, à savoir les transactions SWIFT. Nous porterons une attention particulière aux messages MT103, qui sont le type de messages les plus fréquemment utilisés dans les transferts internationaux de fonds.

3.2.1 Présentation des messages MT103

SWIFT est un réseau interbancaire qui permet aux banques d'échanger des transactions. Il existe différentes catégories de transactions qui sont représentées par des messages. Les mes-

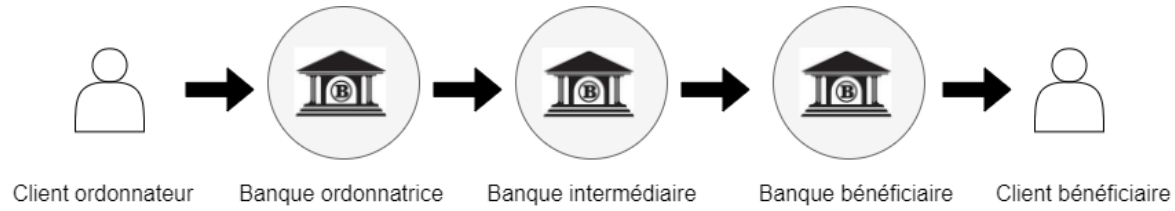

FIGURE 3.2 – Message MT103

sages MT103, qui sont les plus couramment utilisés sont illustrés dans la figure 3.2 et dont les champs sont présentés dans le tableau 3.1. Dans un MT103, un acteur (client ordonnateur) envoie de l'argent à un autre acteur (client bénéficiaire). Ces deux acteurs clients peuvent être des personnes ou des entreprises et sont représentés respectivement par les champs *:50* et *:59*. Le transfert d'argent s'effectue depuis des acteurs bancaires : la banque du client ordonnateur (champ *:52*) et la banque du client bénéficiaire (champ *:57*). Cependant, la transaction peut impliquer des acteurs bancaires intermédiaires dans le cas où les banques des clients ordonnateurs et bénéficiaires ne possèdent pas de relation d'échange établie. Le tableau 3.1 décrit les différents champs d'un message MT103. Chacun de ces champs contient des informations spécifiques sur la transaction, incluant le numéro de référence de la transaction, le code opération bancaire, la date de valeur, la devise, le montant, les détails du payeur et du bénéficiaire, les informations sur les correspondants bancaires, les frais applicables, les informations de remise, les détails de l'expéditeur et du destinataire, et les rapports réglementaires. Il est important de noter que certains de ces champs sont obligatoires pour tous les messages MT103, tandis que d'autres sont facultatifs ou peuvent varier en fonction des exigences spécifiques de la banque ou des parties impliquées dans la transaction.

TABLE 3.1 – Les champs d'un message MT103

Champ	**Nom du champ**
:20	Numéro de référence de transaction
:23	Code opération bancaire
:32	Date de valeur / Devise / Règlement interbancaire
:33	Devise / Montant ordonné original
:50	Client ordonnateur (Payeur)
:52	Institution ordonnatrice (Banque du payeur)
:53	Correspondant de l'expéditeur (Banque)
:54	Correspondant du bénéficiaire (Banque)
:56	Intermédiaire (Banque)
:57	Compte avec l'institution (Banque du bénéficiaire)
:59	Bénéficiaire
:70	Informations de remise (Référence de paiement)
:71	Détails des frais (BEN / OUR / SHA)
:72	Informations de l'expéditeur au destinataire
:77	Reporting réglementaire

Lors du développement de notre approche, les experts ont fait le choix de conserver un sous-

ensemble de champs des messages MT103 [1]. Ils incluent trois acteurs bancaires : la **banque ordonnatrice**, qui est la banque du client ordonnateur, la **banque intermédiaire** et la **banque bénéficiaire**, qui est la banque du client bénéficiaire. Nous ne disposons pas des champs concernant les acteurs clients de la transaction, mais seulement des acteurs bancaires impliqués dans celle-ci. Chaque transaction correspond au transfert d'un **montant** d'argent dans une **devise**, effectué à une **date** précise. Le cheminement des transactions SWIFT dépend de la relation entre la banque ordonnatrice et la banque bénéficiaire. Si ces deux entités n'entretiennent pas de relation directe, alors la transaction implique une banque intermédiaire qui assure la connexion entre les deux banques. Dans le cas contraire, le parcours ne comprend que la banque ordonnatrice et la banque bénéficiaire. Chaque banque est identifiée par un code BIC, dont les quatre premières lettres représentent le code de la banque, les deux lettres suivantes indiquent le code de son pays, et les deux dernières lettres précisent la localisation de son siège social.

TABLE 3.2 – Exemple de messages MT103 du jeu de données.

Bq ordonnatrice	Bq intermédiaire	Bq bénéficiaire	Date	Devise	Montant	Classe
BIC0FR01	BIC0IT01	BIC0FR02	210625	EUR	15006	L
BIC0US03	-	BIC0GB01	210625	GBP	33065	L
BIC0FR04	BIC0FR06	BIC0FR05	210626	EUR	100325	F

La classe de la transaction indique si celle-ci est frauduleuse (F) ou légitime (L). Ce label est attribué par les experts chargés de contrôler les transactions bloquées par des systèmes de lutte contre la fraude financière.

En conséquence, pour modéliser une transaction, nous utilisons la variable x qui comporte six caractéristiques dont 4 catégorielles, 1 numérique et 1 temporelle. De plus, il y a une variable cible qui est la classe de la transaction :

— catégorielles : $x_{ordonnateur}$, $x_{intermediaire}$, $x_{beneficiaire}$, x_{devise}
— Numérique : $x_{montant}$
— Temporelle : x_{date}

3.2.2 Analyse des données

Dans un jeu de données contenant des transactions labellisées comme légitimes et frauduleuses, il est important d'identifier les transactions associées à des schémas de fraude. Pour cela, une première étape importante réside dans la compréhension globale du jeu de données. Une analyse exploratoire approfondie permet d'étudier les différences entre les transactions légitimes et frauduleuses en se basant sur les caractéristiques de base des transactions.

L'analyse exploratoire peut inclure la comparaison des distributions de montants en fonction des pays, des devises utilisées ou des acteurs impliqués. En examinant ces différentes dimensions, il est possible repérer des tendances significatives qui mettent en évidence des schémas de fraude potentiels.

1. https://www.moneymover.com/about/faqs/what-mt103/

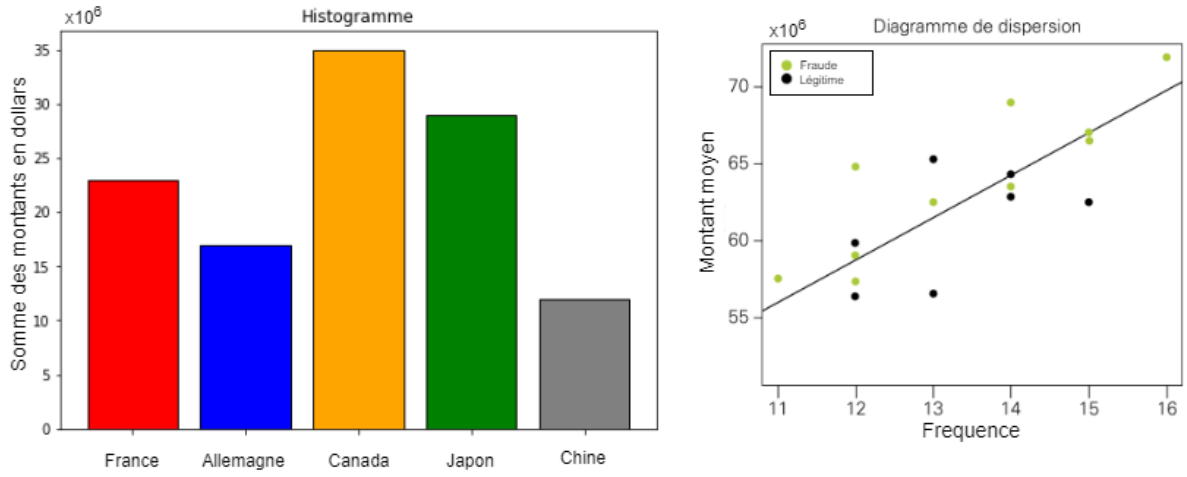

FIGURE 3.3 – Résultats de visualisation des données

Pour mener à bien cette analyse exploratoire, il est pertinent d'exploiter des méthodes de visualisation de données. Les graphiques et les diagrammes peuvent être utiles pour représenter les relations entre les variables et identifier d'éventuelles corrélations ou anomalies au sein des données. Par exemple, un diagramme de dispersion ou un histogramme des montants de transactions en fonction des pays peut révéler des regroupements ou des écarts significatifs, pouvant indiquer des schémas de fraude spécifiques à certains pays.

Notre analyse exploratoire nous permet d'identifier des schémas de fraude en utilisant des méthodes de visualisation adaptées. Ces informations seront utiles lors de la troisième étape, où nous devrons évaluer le nombre potentiel de schémas de fraude. À titre d'exemple, dans la figure 3.3, nous présentons un histogramme représentant la somme des montants des transactions par pays, ce qui nous aide à déterminer si certains pays sont plus difficiles à surveiller que d'autres. Généralement, plus le nombre de transactions est élevé, plus la surveillance devient complexe. Nous avons également un diagramme de dispersion, dans lequel chaque point représente un acteur selon la moyenne des montants de ses transactions et le nombre de transactions effectuées (fréquence). Nous avons des acteurs impliqués dans une fraude (classe « Fraude ») et des acteurs légitimes (classe « Légitime »). Les deux classes peuvent être distinguées avec peu de caractéristiques. Cependant, dans ce cas, les acteurs des deux classes ne peuvent être séparés de manière distincte. Ces diagrammes nous permettent d'obtenir des informations, surtout lorsque nous étudions des classes séparées afin d'identifier les tendances qui permettraient de distinguer les classes, et ainsi nous orienter sur les caractéristiques à calculer.

3.3 Calcul et sélection des caractéristiques

Dans cette section, nous abordons la phase de calcul et sélection des caractéristiques de notre jeu de données. Cette étape est importante, car ces caractéristiques joueront un rôle crucial dans la distinction entre les transactions légitimes et frauduleuses. Nous entamons ce processus en enrichissant notre jeu de données avec les caractéristiques spécifiques aux transactions SWIFT que nous avons calculées dans le cadre de notre démarche. Par la suite, nous décrirons la méthode que nous avons choisie pour effectuer la sélection de caractéristiques.

3.3.1 Enrichissement du jeu de données

L'enrichissement des caractéristiques du jeu de données est une étape cruciale dans la détection de transactions frauduleuses dans les données SWIFT. En effet, les données de transactions SWIFT peuvent être complexes et hétérogènes, comprenant des informations, telles que les codes BIC, les dates, les montants et les devises. Afin de distinguer les transactions frauduleuses des transactions légitimes, il est important de définir des caractéristiques pertinentes et d'appliquer des techniques de sélection de caractéristiques pour choisir les plus significatives, améliorant ainsi les performances des modèles de classification. Les caractéristiques sélectionnées ont un impact significatif sur la précision des résultats de la détection de fraude. Par conséquent, l'ajout de nouvelles caractéristiques est une étape importante qui peut aider à améliorer la précision de la détection de fraude.

Nous commençons par définir des caractéristiques basées sur les acteurs identifiés par des codes BIC. Le quatrième et le cinquième caractère du code BIC représentent le code pays de chaque acteur (par exemple : « FR » pour « France »). Ainsi, nous obtenons les caractéristiques suivantes : $x_{pays\_emetteur}$, $x_{pays\_intermediaire}$, $x_{pays\_beneficiaire}$ qui sont essentielles pour prendre en compte la dimension internationale des transactions SWIFT, qui représente un indicateur potentiel de fraude dans nos transactions.

Pour modéliser le comportement des acteurs, nous nous appuyons sur les principes du modèle RFM [VVBC$^+$15] (récence, fréquence et somme) qui est largement adopté dans la littérature de la détection de transactions frauduleuses, tout en définissant de nouvelles caractéristiques spécifiques au réseau SWIFT sélectionnées en collaboration avec des experts financiers.
Pour chaque transaction x impliquant un acteur x_{acteur} et en se basant sur son historique de transactions X, nous définissons les caractéristiques avec leur description dans le tableau 3.3.

TABLE 3.3 – Descriptifs des caractéristiques

Caractéristique	**Description**
Récence	Temps écoulé en secondes depuis la transaction x.
Fréquence	Nombre de transactions réalisées sur une période donnée.
Somme	Somme des montants des transactions sur une période.
Min	Montant minimum des transactions de X.
Max	Montant maximum des transactions de X.
Avg	Montant moyen réalisé des transactions de X.
Nombre d'acteurs différents	Nombre d'acteurs différents dans X.
Nombre de pays différents	Nombre de pays différents dans X.
Nombre de devises différentes	Nombre de devises différentes dans X.
Nombre de transactions avec intermédiaire	Nombre de transactions impliquant des acteurs intermédiaires dans X.

Les caractéristiques « Fréquence » et « Somme » sont calculées en fonction d'une période spécifique. En effet, la temporalité est à prendre en compte pour caractériser un acteur. Par

exemple, un acteur effectuant 100 transactions en une journée n'est pas comparable à un acteur réalisant ce même nombre sur une année. Ainsi, nous utilisons la caractéristique x_{date} pour pouvoir calculer ces deux caractéristiques sur différentes périodes. Nous modélisons le comportement de l'acteur en utilisant les 10 caractéristiques définies ci-dessus, dont 2 peuvent être calculées à plusieurs reprises sur différentes périodes.

Les messages MT103 peuvent impliquer jusqu'à 3 acteurs, ce qui signifie que ces 10 caractéristiques sont calculées 2 ou 3 fois en fonction de la présence d'intermédiaires. Pour chaque transaction, ces caractéristiques sont calculées à partir de l'historique de transactions de chaque acteur impliqué. En outre, nous pouvons également exploiter les dimensions internationales des transactions SWIFT en calculant des caractéristiques sur les pays et les devises. Par exemple, il est utile d'avoir une caractéristique représentant la moyenne des transactions dans la devise de la transaction. Nous calculons des caractéristiques basées sur des ensembles de transactions regroupées sur la devise ou le pays de la transaction.

Nous avons synthétisé dans le tableau 3.4 les caractéristiques calculées pour les acteurs (A.), les pays (P.) et les devises (D.). Nous calculons les caractéristiques présentes dans le tableau 3.3 sur des historiques de transactions. Nous avons adopté une approche exhaustive en explorant toutes les relations possibles entre les acteurs, les pays et les devises impliqués dans les transactions. Cette approche met ainsi en évidence les liens et les interactions à différentes échelles. Par exemple, pour une transaction donnée, nous pouvons calculer des caractéristiques sur les transactions effectuées par un acteur impliqué dans la transaction sur la même devise que la transaction (A.→ D.). Nous pouvons également examiner les transactions réalisées entre deux acteurs pour déterminer si le montant de la transaction x diffère de la moyenne des transactions qu'ils échangent habituellement (A.→A.). En outre, nous avons calculé des caractéristiques pour les relations entre l'acteur et les pays impliqués dans la transaction (A.→P.), entre les pays impliqués (P.→P.), et entre les pays impliqués et la devise (P.→D.) utilisée.

Caractéristiques	**Acteur**	**Pays**	**Devise**	**Relations**				
	A.	P.	D.	A.→A.	A.→P.	P.→P.	A.→D.	P.→D.
Min	✓	✓	✓	✓	✓	✓	✓	✓
Max	✓	✓	✓	✓	✓	✓	✓	✓
Moyenne	✓	✓	✓	✓	✓	✓	✓	✓
Somme	✓	✓	✓	✓	✓	✓	✓	✓
Récence	✓			✓	✓		✓	
Nb de P. différent	✓			✓				
Nb de A. différent	✓							
Nb de D. différent	✓							
Nb transactions avec Intermédiaire	✓	✓	✓	✓	✓	✓	✓	✓

TABLE 3.4 – Les caractéristiques définies sur un ensemble de transactions

Toutes les transactions ne comportent pas systématiquement d'intermédiaire. Par conséquent, pour ces transactions, les dimensions impliquant un intermédiaire ne seront pas considérées. Ce travail a été publié dans les conférences nationales EGC et Inforsid [CAA+22, CACC22].

Après cette étape, un grand nombre de caractéristiques sont calculées, parmi lesquelles certaines sont pertinentes tandis que d'autres se révèlent redondantes ou non pertinentes. C'est

pourquoi il est essentiel de mettre en place une méthode de réduction de caractéristiques. Dans la suite, nous présentons notre méthode de réduction de caractéristiques.

3.3.2 Réduction de dimension et algorithmes basés sur les arbres de décisions

La réduction de la dimensionnalité des jeux de données est essentielle pour améliorer les performances des modèles en réduisant leur complexité et en évitant le surapprentissage. Cela permet également de réduire le temps de calcul nécessaire à l'entraînement et à la prédiction. En sélectionnant les caractéristiques les plus importantes, il est possible de considérablement réduire la taille du jeu de données tout en conservant une grande partie de l'information pertinente, ce qui permet d'améliorer l'efficacité sans compromettre la précision du modèle. Dans la suite, nous allons présenter le choix de l'algorithme de sélection des caractéristiques.

3.3.2.1 Choix de l'algorithme de réduction de caractéristiques

Dans un premier temps, nous rappelons les méthodes utilisées pour réaliser une sélection de caractéristiques. Dans un second temps, nous présentons l'algorithme retenu pour notre approche : les arbres de décision.

Rappel des méthodes de sélection de caractéristiques Dans notre revue de la littérature, nous avons identifié deux approches courantes pour la réduction de dimensionnalité : l'extraction de caractéristiques et la sélection de caractéristiques. Nous avons opté pour la méthode de sélection de caractéristiques, car elle nous permet de conserver un sous-ensemble des caractéristiques originales, facilitant ainsi une interprétation et une analyse ultérieure. Contrairement à l'extraction de caractéristiques, qui peut rendre les caractéristiques moins significatives ou plus complexes à interpréter.

La sélection de caractéristiques peut s'opérer selon trois méthodes [JBB15] :

- Intégrée : Les méthodes intégrées effectuent la sélection de caractéristiques pendant le processus d'apprentissage du modèle. Cela signifie que la sélection des caractéristiques est incorporée directement dans l'algorithme d'apprentissage, ce qui permet une interaction étroite entre la sélection des caractéristiques et la construction du modèle. Les méthodes intégrées recherchent le sous-ensemble optimal de caractéristiques en utilisant des critères tels que la performance du modèle, la complexité du modèle et la pénalité du surapprentissage. Ces critères permettent de trouver un équilibre entre la performance prédictive et la complexité du modèle, en évitant le surapprentissage. Les méthodes intégrées sont souvent plus puissantes que les méthodes de filtrage en termes de performance prédictive, car elles tiennent compte des interactions entre les caractéristiques
- Enveloppe : Les méthodes enveloppe sont une approche de sélection de caractéristiques qui commencent soit avec un ensemble de caractéristiques vide, soit avec l'ensemble complet. Ensuite, elles effectuent une recherche itérative en ajoutant ou en supprimant des caractéristiques, puis en entraînant un modèle de classification. L'évaluation du modèle est utilisée comme critère pour sélectionner les caractéristiques ajoutées ou supprimées. Si l'évaluation du modèle est satisfaisante, les caractéristiques sont conservées dans le sous-ensemble sélectionné. En revanche, si l'évaluation est insatisfaisante,

les caractéristiques sont rejetées. Cette approche itérative se poursuit jusqu'à ce qu'un critère d'arrêt prédéfini soit atteint, tel que l'obtention d'une performance suffisante ou l'épuisement de toutes les combinaisons possibles. Les méthodes enveloppe utilisent donc l'entraînement et l'évaluation du modèle de classification comme mécanisme pour guider la sélection des caractéristiques les plus pertinentes.

— Filtrage : Contrairement aux méthodes de sélection de caractéristiques intégrées et d'enveloppes, les méthodes de filtrage sont indépendantes de l'algorithme de classification. Cette indépendance réduit le surapprentissage en évitant le biais de l'algorithme de classification. Cependant, cette indépendance signifie également qu'il n'y a pas d'interaction avec le classifieur lors de la sélection des caractéristiques, ce qui rend l'ensemble de caractéristiques sélectionné plus général et moins spécifique à un classifieur particulier. Le filtrage univarié consiste à évaluer les caractéristiques indépendamment les unes des autres, en utilisant des mesures statistiques telles que le test du χ^2 (khi carré) ou la corrélation de Pearson. D'un côté, le filtrage univarié est facile à comprendre et à mettre en œuvre. Il implique l'application de critères simples à chaque variable, ce qui permet de détecter rapidement les caractéristiques potentiellement intéressantes. Il examine chaque variable indépendamment, le filtrage univarié peut être efficace d'un point de vue computationnel. Enfin, le processus de sélection est interprétable avec le choix des critères et l'analyse des valeurs des caractéristiques. D'un autre côté, le filtrage univarié présente quelques faiblesses. Tout d'abord, il ignore les *interactions* entre les caractéristiques. En se concentrant uniquement sur les relations entre chaque variable et la variable cible de manière isolée, il peut manquer des relations complexes qui ne sont visibles qu'en combinant plusieurs variables. Il est sensible *aux données bruitées*, si les données contiennent du bruit ou des valeurs aberrantes, le filtrage univarié peut être influencé.

Nous avons opté pour l'utilisation de techniques de sélection de caractéristiques **intégrée** dans notre méthode pour plusieurs raisons. Tout d'abord, ces techniques se sont avérées plus performantes que les méthodes par filtrage, car elles exploitent des algorithmes d'apprentissage automatique capables de détecter les corrélations entre les caractéristiques.

Les techniques de sélection de caractéristiques peuvent être mises en œuvre avec l'algorithme des arbres de décisions. Dans la littérature, nous avons identifié que les algorithmes fondés sur les arbres de décisions sont efficaces pour une tâche de classification dans le domaine de la détection de fraude. Ces algorithmes présentent plusieurs avantages, tels que leur interprétabilité, leur efficacité et la possibilité de les combiner avec d'autres approches d'apprentissage ensembliste.

Ainsi, les arbres de décisions permettent d'attribuer un score d'importance aux caractéristiques pendant leur entraînement pour une tâche de classification. Dans la suite, nous rappelons le fonctionnement de cet algorithme pour comprendre comment il peut effectuer cette tâche de sélection de caractéristiques.

L'algorithme des arbres de décisions [Qui86] est un modèle d'apprentissage automatique utilisé pour la classification et la régression. Il s'agit d'un modèle de prédiction simple et interprétable qui permet de prendre des décisions en se basant sur une suite de règles conditionnelles.

Le principe de l'algorithme consiste à construire un arbre à partir du jeu de données d'entrée. Nous présentons un exemple simplifié d'un arbre de décision dans la figure 3.4. Chaque nœud de l'arbre représente une question sur une caractéristique du jeu de données, et les branches de l'arbre représentent les différentes réponses possibles à cette question. Les feuilles de l'arbre représentent les décisions finales, qui peuvent être des classes pour la classification ou des valeurs pour la régression.

L'algorithme de construction de l'arbre consiste à diviser le jeu de données en sous-ensembles plus petits et plus homogènes à chaque nœud de l'arbre. Pour cela, l'algorithme utilise des critères d'*impureté* pour mesurer l'homogénéité des sous-ensembles. L'impureté est une mesure utilisée dans les algorithmes d'arbres de décision pour évaluer la qualité des différentes divisions possibles d'un ensemble de données. Elle quantifie à quel point les classes sont mélangées dans les sous-ensembles : une impureté de 0 signifie que le sous-ensemble est parfaitement homogène, c'est-à-dire qu'il contient des données d'une seule classe, tandis qu'une impureté plus élevée indique un mélange de classes. Les deux mesures d'impureté couramment utilisées sont le coefficient de Gini et l'entropie. Le coefficient de Gini est une mesure de la probabilité que deux éléments choisis au hasard dans le sous-ensemble appartiennent à des classes différentes, tandis que l'entropie mesure le désordre ou l'incertitude associée au sous-ensemble. L'objectif est de trouver les caractéristiques qui maximisent la séparation entre les différentes classes ou valeurs. Une fois que l'arbre est construit, il peut être utilisé pour prédire la classe ou la valeur d'un nouvel exemple en le faisant traverser l'arbre en suivant les règles conditionnelles jusqu'à atteindre une feuille.

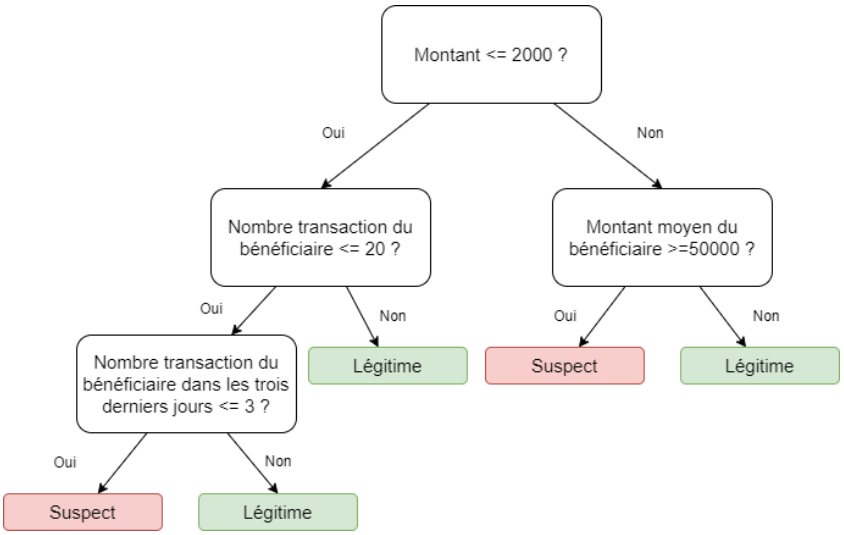

FIGURE 3.4 – Exemple d'arbre de décision

Cependant, l'algorithme d'arbre de décision peut souffrir de surapprentissage, ce qui se produit lorsque l'arbre est trop complexe et s'adapte trop étroitement au jeu de données d'entraînement. Pour éviter cela, des paramètres tels que la taille maximale de l'arbre, la taille minimale des feuilles ou le pruning peuvent être utilisés.
De plus, cet algorithme permet d'estimer l'importance de chaque caractéristique. Chaque nœud (n_i) est associé à une caractéristique k ($n_{i,k}$). Nous pouvons calculer l'importance du nœud d'un arbre de décision avec l'équation suivante :

$$\text{Importance}(n_{i,k}) = \text{Impurete}(n_{i,k}) - (\text{Proportion}(S_{left}) \text{ *Impurete}(S_{left}) + \text{Proportion}(S_{right}) * \text{Impurete}(S_{right})) \tag{3.1}$$

La proportion de S_{left} et la proportion de S_{right} représentent respectivement les proportions d'instances dans la branche gauche et la branche droite.

Pour calculer, l'importance d'une caractéristique k dans un arbre de décision, il faut diviser la somme de l'importance des nœuds de l'arbre qui se divise avec la caractéristique k par la somme de l'importance de tous les nœuds de l'arbre (n_j). Cela peut être exprimé par l'équation suivante :

$$\text{Importance}(k) = \frac{\sum_i \text{Importance}(n_{i,k})}{\sum_j \text{Importance}(n_j)} \tag{3.2}$$

En disposant des importances de toutes les caractéristiques, nous pouvons sélectionner celles qui ont les importances les plus élevées pour réduire la dimensionnalité du jeu de données. De plus, l'algorithme des arbres de décisions présente l'avantage de ne pas nécessiter de prétraitement des données supplémentaires. Les techniques de normalisation et de mise à l'échelle n'ont pas d'impact sur cet algorithme.

Dans la section suivante, nous abordons la tâche de clustering réalisée pour regrouper les transactions frauduleuses par schéma de fraude.

3.4 Clustering des transactions frauduleuses par schémas de fraude

Dans cette section, nous effectuons un regroupement des transactions frauduleuses en clusters pour rassembler les transactions similaires. Ainsi, nous sommes en mesure de mettre à jour le label de ces transactions en les associant à leur cluster correspondant.

Cette étape nous permet d'obtenir une compréhension plus approfondie des raisons pour lesquelles ces transactions ont été classifiées suspectes par le modèle. Il existe différentes méthodes de clustering pour regrouper des données similaires [RM05] :

- Clustering basé sur la densité : regroupent les données en fonction des zones de forte concentration de points entourées de zones de faible concentration de points. Les clusters peuvent prendre n'importe quelle forme et les valeurs aberrantes sont ignorées.
- Clustering basé sur la distribution : Dans cette approche, tous les points de données sont considérés comme faisant partie d'un cluster en fonction de la probabilité qu'ils appartiennent à un cluster donné. Les points de données sont attribués à un cluster en fonction de leur distance par rapport au centre.
- Clustering basé sur le centroïde : séparent les points de données en fonction de plusieurs centroïdes dans les données. Les points de données sont attribués à un cluster en fonction de sa distance euclidienne par rapport au centroïde.

— Clustering hiérarchique : utilisé pour les données hiérarchiques, comme celles que l'on trouve dans une base de données d'entreprise ou dans des taxonomies. Il construit un arbre de clusters pour organiser toutes les données de haut en bas.

Lors de l'analyse exploratoire des données, nous avons identifié un nombre potentiel n de schémas de fraude distincts présents dans le jeu de données. Les experts métier nous ont permis de rattacher des comportements identifiés avec des schémas de fraude. Notre approche vise à identifier et analyser en profondeur ces schémas de fraude spécifiques. Dans cette optique, nous formons des clusters regroupant des transactions similaires. Ainsi, un cluster donné peut être associé à un schéma de fraude particulier. Il est également possible que plusieurs clusters distincts soient liés au même schéma de fraude.

Dans notre cas, nous avons choisi d'utiliser une méthode de clustering hiérarchique, principalement pour sa capacité à gérer la complexité des schémas de fraude. En effet, lorsque différents clusters peuvent correspondre au même schéma de fraude, il est important d'avoir une méthode qui puisse identifier et regrouper ces liens sous-jacents. Le clustering hiérarchique est particulièrement adapté à cette tâche car il peut regrouper ces clusters similaires sous la même branche de l'arbre, permettant une meilleure visualisation et interprétation des relations entre clusters. De plus, cette méthode offre l'avantage de déterminer le nombre de clusters de manière visuelle.

La classification ascendante hiérarchique (CAH) [MC12] regroupe des éléments en fonction de leur similarité. La CAH commence par considérer chaque élément comme un cluster indépendant, puis elle les fusionne progressivement pour former des groupes plus larges.

Pour fusionner deux clusters, il faut déterminer leur distance. Cette distance peut être calculée de différentes manières, telles que le lien simple (distance minimale entre deux points des clusters), le lien complet (distance maximale entre deux points des clusters) ou encore la distance centroïdale (distance entre les centroïdes des deux clusters).

Le choix de la méthode de calcul de la distance dépend des données et de l'objectif du clustering. Ensuite, la CAH fusionne les clusters les plus proches, jusqu'à obtenir un seul groupe contenant tous les éléments. Le nombre de clusters à obtenir est souvent déterminé à l'avance, mais il peut également être déterminé de manière visuelle en analysant la courbe de dendrogramme, qui représente les étapes de fusion des clusters.

Le dendrogramme est un outil de visualisation couramment utilisé dans le cadre de la classification ascendante hiérarchique, nous en présentons un exemple dans la figure 3.5. Il permet de représenter graphiquement la structure hiérarchique des regroupements ou des similarités entre des objets (ex. transactions). Il est constitué de nœuds qui représentent les objets ou les regroupements d'objets. Ces nœuds sont reliés par des branches qui indiquent les liens de similarité entre les groupes. La hauteur des branches indique la similarité ou la distance entre les groupes.
En observant le dendrogramme, il est possible de déterminer le nombre optimal de clusters à partir duquel on souhaite faire des groupements. Graphiquement, la ligne de coupure horizontale est choisie en identifiant le point où une ligne horizontale traverse le plus grand nombre de branches verticales, permettant ainsi de délimiter des clusters distincts.

En appliquant l'algorithme de Clustering Hiérarchique Agglomératif (CAH) uniquement sur les transactions frauduleuses de notre jeu de données, nous parvenons à identifier n classes

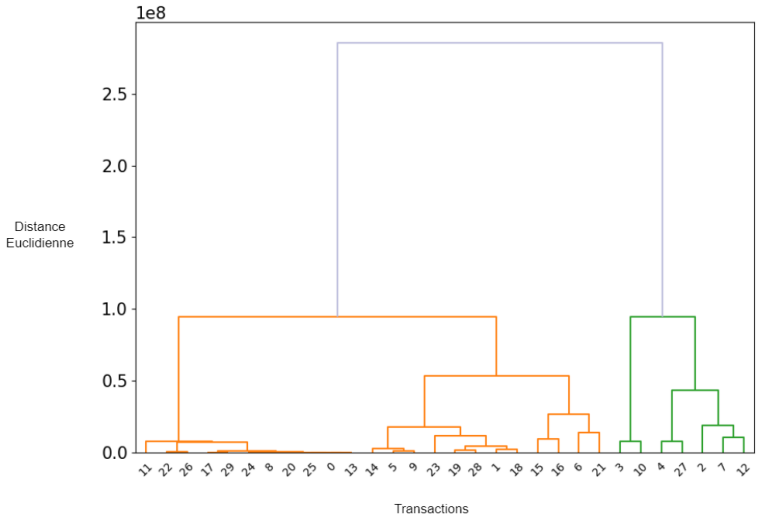

FIGURE 3.5 – Exemple de dendrogramme

distinctes de transactions frauduleuses. Par ailleurs, notre jeu de données comprend également des transactions légitimes, constituant ainsi une classe supplémentaire. Ainsi, au total, nous disposons de $n + 1$ classes de transactions, avec n classes correspondant à des schémas de fraude et 1 classe regroupant les transactions légitimes. À ce stade de notre démarche, l'association précise entre ces clusters de fraude et les schémas de fraude spécifiques n'est pas encore déterminée. Cette étape d'identification sera réalisée ultérieurement à travers notre modèle de classification, enrichi par des méthodes d'interprétation des modèles.

Après avoir identifié les clusters potentiels de fraude, nous abordons maintenant l'étape de classification. Durant cette phase, notre objectif est d'entraîner un modèle capable de discriminer efficacement entre les transactions légitimes et celles appartenant à différents schémas de fraude.

3.5 Classification des transactions suspectes

Dans cette section, nous détaillons notre méthode pour entraîner un modèle prédictif à partir de notre jeu de données. Dans un premier temps, nous adoptons une stratégie d'apprentissage ensembliste pour gérer la classification multi-classes. Dans un second temps, nous présentons notre méthode de calcul d'un score de suspicion basé sur les prédictions fournies par le modèle.

3.5.1 Rappel des méthodes d'apprentissage ensembliste

Pour optimiser l'utilisation des arbres de décision, nous adoptons des méthodes d'apprentissage ensembliste qui forment plusieurs modèles et fusionnent leurs prédictions selon des stratégies définies. Quatre algorithmes de classification ont été privilégiés : le Random Forest et trois variantes de l'algorithme *Gradient Boosting Tree*. Ces algorithmes ont démontré leur efficacité dans la littérature [BCMM21]. Dans la suite, nous détaillons les principes des algorithmes Random Forest et *Gradient Boosting Tree*. Un comparatif de ces algorithmes est

présenté dans le tableau 3.5.

Algorithme	Description	Type d'apprentissage
Random Forest [Bre01]	Random Forest est un algorithme de classification et de régression qui construit plusieurs arbres de décision aléatoires, puis agrège leurs prédictions pour obtenir une prédiction finale plus robuste et moins susceptible au surapprentissage.	*Bagging*
XGBoost [CHB+15]	XGBoost est un algorithme de boosting de gradient tree basé sur l'optimisation d'une fonction de perte régularisée. Il utilise également des techniques d'élagage pour améliorer l'efficacité du modèle.	*Gradient Boosting Tree*
CatBoost [PGV+18]	CatBoost est un algorithme de boosting qui utilise un encodage catégoriel avancé pour gérer automatiquement les caractéristiques catégorielles dans les données d'entrée, tout en utilisant une méthode de régularisation innovante pour éviter le surapprentissage.	*Gradient Boosting Tree*
LightGBM [KMF+17]	LightGBM est un autre algorithme de boosting de gradient tree qui utilise une technique de division de feuille verticale pour réduire le coût de la recherche de la meilleure division, ainsi qu'une stratégie de binning pour gérer les caractéristiques numériques.	*Gradient Boosting Tree*

TABLE 3.5 – Tableau comparatif des algorithmes ensemblistes

Les algorithmes de Random Forest et les algorithmes de Gradient Boosting Tree offrent des stratégies sophistiquées d'apprentissage ensembliste. Chacun possède ses particularités et forces, offrant des solutions adaptées à différents scénarios et types de données. Après avoir compris leurs mécanismes, nous nous tournons maintenant vers le choix du modèle en présentant comment évaluer leur performance.

3.5.2 Le choix du modèle et le calcul du score de suspicion

Selon les besoins des institutions financières, le choix de l'algorithme approprié peut varier. Par exemple, l'algorithme Catboost est recommandé pour gérer les données avec des caractéristiques catégorielles en raison de sa gestion automatique de ces dernières. En revanche, LightGBM est idéal pour l'entraînement rapide de modèles grâce à ses techniques optimisées pour traiter de grands ensembles de données rapidement. Cependant, il n'y a pas de méthode universelle pour déterminer le meilleur algorithme en fonction des données. Il est courant d'entraîner des modèles avec plusieurs algorithmes. Chaque modèle est évalué selon des mesures, et le choix de l'algorithme s'effectue en choisissant le modèle avec la meilleure évaluation.

Une fois l'algorithme choisi, un modèle est entraîné avec le jeu d'entraînement obtenu après les étapes de :

— calcul de caractéristiques
— sélection de caractéristiques
— mise à jour des labels

Une fois le modèle entraîné, il est prêt à effectuer des prédictions sur des transactions non labellisées. Dans notre approche, nous avons également proposé une méthode de calcul du score de suspicion pour optimiser les résultats de notre modèle.

La prédiction d'un modèle de classification fournit des probabilités d'appartenance à différentes classes pour chaque transaction. Dans notre contexte, nous disposons de $n + 1$ classes, où n classes correspondent à différents schémas de fraude et une classe représente les transactions légitimes.

La probabilité d'appartenance indique la probabilité qu'une transaction soit classée dans une certaine catégorie, tandis que le score de suspicion combine les probabilités de tous les schémas de fraude pour donner une vision globale du risque. En d'autres termes, alors que la probabilité d'appartenance répartit le risque entre différents schémas de fraude, le score de suspicion agrège ces risques pour donner une mesure complète du caractère potentiellement frauduleux d'une transaction.

Pour quantifier cette suspicion, nous additionnons les probabilités associées aux n classes de fraude. Étant donné que la somme des probabilités de toutes les $n + 1$ classes est de 1, cette agrégation revient à soustraire la probabilité que la transaction $P_l(t)$ soit légitime de 1. Nous exprimons la probabilité en pourcentage afin de faciliter sa lecture, nous multiplions ce résultat par 100, donnant lieu à la formule de score de suspicion suivante :

$$S(t) = (1 - P_l(t)) * 100 \tag{3.3}$$

Ce score varie de 0 et 100. Une valeur élevée indique un niveau élevé de suspicion. Le défi ultérieur consiste à déterminer un seuil spécifique au-delà duquel une transaction est jugée suspecte.

3.5.3 Évaluation et minimisation du coût

Une fois le score calculé, nous définissons un score de suspicion pour chaque transaction, nous allons définir un seuil à partir duquel une transaction sera classifiée comme suspecte. Ce seuil doit garantir à la fois l'efficacité du modèle en termes d'évaluation et la minimisation de son coût.

Pour évaluer notre modèle, nous utilisons les mesures présentées dans le chapitre 2 :

— la précision
— le rappel
— le f1-score

De plus, pour quantifier le coût financier de notre modèle, nous avons adapté la matrice de coût de Bahnsen et al. [BSAO13], que nous présentons dans le tableau 3.6. Pour une transaction (t_i), si le modèle la prédit frauduleuse ($y_i = 1$), alors elle entrainera un coût administratif Ca

représentant le coût d'un expert pour contrôler une transaction détectée frauduleuse par un modèle. Si t_i est frauduleuse et que le modèle la prédit comme légitime ($y_i = 0$), alors son coût sera égal à son montant (Amt_i). En collaboration avec des experts, nous avons adapté cette méthode de quantification du coût en multipliant par deux le coût financier d'une transaction légitime détectée frauduleuse par le modèle. En effet, nous ne pouvons pas associer le même coût à une prédiction correcte et à une prédiction incorrecte. Ainsi, en augmentant le coût administratif d'une transaction légitime incorrectement prédite, nous prenons en compte cette erreur du modèle. Enfin, pour obtenir le coût du modèle, nous effectuons une somme de trois termes :

— Le coût des transactions frauduleuses prédites frauduleuses représenté par le nombre de vrais positifs (vp) multiplié par le coût administratif Ca :
$Ca \times vp$

— Le coût des transactions légitimes prédites frauduleuses, calculé avec le nombre de faux positifs (fp) multiplié par deux fois le coût administratif Ca :
$2 \times Ca \times fp$

— Le montant des transactions frauduleuses prédites légitimes :
$\sum_{i=1}^{n}(y_i = 0 \wedge y_{true,i} = 1) \times Amt_i$

Ainsi, pour évaluer les performances de notre modèle, nous utilisons la formule coût suivante :

$$Cout = (Ca \times vp) + \left((2 \times Ca \times fp) + \left(\sum_{i=1}^{n}(y_i = 0 \wedge y_{true,i} = 1 \right) \times Amt_i \right) \quad (3.4)$$

TABLE 3.6 – Matrice de risque du coût de Bahnsen [BSAO13] adapté.

		Réalité (t_i)	
		Fraude	Légitime
Prédiction (t_i)	Fraude	Ca	**2** $\times Ca$
	Légitime	Amt_i	0

La détermination d'un seuil de classification est essentielle pour garantir que le modèle fonctionne de manière optimale dans un contexte réel. Dans ce cadre, nous proposons l'algorithme 1 de minimisation de coût financier avec deux phases distinctes sont utilisées pour trouver le seuil optimal : la maximisation du f1-score et la minimisation du coût.

Maximisation du f1-score La première phase vise à déterminer le seuil qui maximise le f1-score, une mesure qui équilibre la précision et le rappel. Nous commençons avec un seuil de 0, et l'augmentons progressivement par un petit incrément, $\Delta\theta$, à chaque itération, jusqu'à atteindre la valeur de 100. À chaque étape, le f1-score est calculé pour le seuil actuel. L'objectif est d'identifier le seuil, noté θ_{best}, pour lequel le modèle affiche le f1-score le plus élevé, noté $f1_{max}$.

Minimisation du coût Cependant, maximiser le f1-score ne garantit pas nécessairement que le modèle sera économiquement viable dans un scénario réel. C'est pourquoi, dans la seconde phase, l'objectif est de trouver un seuil qui minimise le coût financier tout en maintenant le

f1-score proche de sa valeur maximale. Ce coût tient compte du coût administratif associé à la prédiction d'une transaction ainsi que d'autres pénalités liées aux erreurs de classification. Pour ce faire, nous réexaminons tous les seuils possibles. À chaque étape, nous calculons le coût associé à ce seuil et nous le comparons à un coût minimal prédéfini, $cost_{min}$. Si le coût est inférieur à $cost_{min}$ et que le f1-score se situe à une distance acceptable (définie par ϵ) de $f1_{max}$, alors ce seuil est considéré comme optimal.

Il est important de noter que la formule de calcul des coûts dans l'algorithme prend également en compte le montant des transactions frauduleuses non détectées, ce qui souligne l'importance de minimiser les faux négatifs pour les transactions aux montants les plus élevés. En fin de compte, cet algorithme permet d'obtenir un seuil optimal θ_{opt} qui assure le meilleur équilibre possible entre la performance du modèle et le coût financier pour les institutions financières.

Une fois le seuil fixé, toutes les transactions dont le score de suspicion est supérieur au seuil seront considérées suspectes. Cette étape à fait l'œuvre de publications dans des conférences nationales [CACC23a]. Après avoir entraîné le modèle, nous utilisons les techniques d'interprétabilité des modèles pour identifier les schémas de fraude dans la section suivante.

3.5.4 Identification des schémas de fraude

Une transaction frauduleuse suit un schéma spécifique, qui représente un mode opératoire utilisé pour commettre une fraude. En étudiant la littérature, nous nous sommes intéressés aux techniques d'interprétabilité des modèles. Notre modèle multi-classes est capable de prédire des transactions dans $n + 1$ classes, dont n classes sont associées à des clusters.

Nous souhaitons utiliser les techniques d'interprétabilité pour comprendre les prédictions du modèle dans chacune des n classes associées aux clusters. Pour cela, nous avons conservé les transactions suspectes correctement prédites dans les n classes.

Avec cette étape, notre objectif est d'identifier si les clusters peuvent être associés à des schémas de fraude. Pour cela, nous utilisons les valeurs Shapley pour évaluer l'impact des caractéristiques sur les prédictions du modèle dans chacun des clusters. Dans la suite, nous faisons un rappel des valeurs shapley et présentons le framework SHAP.

3.5.4.1 Les valeurs Shapley

Dans le chapitre 2, nous avons identifié plusieurs méthodes pour interpréter les prédictions d'un modèle. Nous avons décidé d'utiliser les méthodes locales agnostiques pour analyser les prédictions des clusters, en particulier les valeurs Shapley [MT20]. Ces dernières nous permettront de comprendre l'importance de chaque caractéristique dans la prédiction de la fraude pour chaque transaction, et ainsi d'identifier si les clusters sont associés à des schémas de fraude.

Les valeurs Shapley sont une méthode d'attribution des prédictions à des caractéristiques individuelles. Elles sont basées sur la théorie des jeux coopératifs et utilisent un modèle de contribution équilibré pour attribuer la contribution de chaque caractéristique à la prédiction.

Algorithm 1 Minimisation du coût financier

Require:

Ca : Coût administratif d'une transaction prédite
p : Score de suspicion
ϵ : Valeur d'écart entre le f1-score maximum et celui du coût optimal
y_{true} : Vérité de terrain des transactions

Ensure:

Seuil de classification optimal θ_{opt}
Coût minimum $cost_{min}$

function MINIMIZECOST($Ca, p, \epsilon, y_{true}$)
    $f1_{max} \leftarrow 0$
    $\theta_{best} \leftarrow 0$
    $\Delta\theta \leftarrow 0.01$
    for $\theta \leftarrow 0$ **to** 1 **step** $\Delta\theta$ **do // Première phase**
        $y \leftarrow$ classer les transactions en fonction de θ et p // Les transactions au-dessus du seuil θ sont suspectes
        $precision \leftarrow$ calcul de la précision avec y et y_{true}
        $rappel \leftarrow$ calcul du rappel avec y et y_{true}
        $f1 \leftarrow 2 \times \frac{precision \times rappel}{precision + rappel}$
        if $f1 > f1_{max}$ **then**
            $f1_{max} \leftarrow f1$
            $\theta_{best} \leftarrow \theta$
        end if
    end for
    $cost_{min} \leftarrow +\infty$
    for $\theta \leftarrow 0$ **to** 1 **step** $\Delta\theta$ **do // Deuxième phase**
        $y \leftarrow$ classer les transactions en fonction de θ et p
        $rappel, precision, f1 \leftarrow$ calcul des mesures
        $cost \leftarrow Ca \times vp + 2 \times Ca \times fp + \sum_{i=1}^{n} (y_i = 0 \wedge y_{true,i} = 1) \times Amt_i$
        if $cost < cost_{min}$ & $f1_{max} <= f1 + \epsilon$ **then**
            $cost_{min} \leftarrow cost$
            $\theta_{opt} \leftarrow \theta$
        end if
    end for
    return θ_{opt}, $cost_{min}$
end function

Cette méthode est particulièrement utile pour l'interprétation des modèles d'apprentissage automatique, car elle fournit une explication de la prédiction.

Dans le domaine de la détection de fraude financière, les valeurs Shapley sont particulièrement importantes. Elles permettent d'attribuer une contribution à chaque caractéristique de la transaction, ce qui permet de comprendre comment les modèles prennent leurs décisions et de détecter les caractéristiques qui sont les plus importantes pour la détection des fraudes.

De plus, les valeurs Shapley présentent l'avantage d'être calculées rapidement et efficacement.

Elles sont également applicables à tous les types de modèles de machine learning, ce qui les rend très utiles pour les approches qui nécessitent une interprétation plus précise et détaillée du modèle.

En comparaisons avec d'autres méthodes, les valeurs Shapley sont également moins sensibles aux interactions entre les caractéristiques, ce qui peut rendre l'interprétation plus difficile avec des méthodes telles que les graphes de dépendance partielle.

Pour une caractéristique i, la valeur Shapley ϕ_i est la contribution marginale moyenne de cette caractéristique i à la prédiction finale y, en prenant en compte toutes les combinaisons possibles de caractéristiques et pour toutes les observations de notre jeu de données. La formule mathématique pour calculer la valeur Shapley ϕ_i d'une caractéristique « i » est la suivante [MT20] :

$$\phi_i(v) = \frac{1}{M} \sum_{S \subseteq \mathcal{M} \setminus \{i\}} \binom{M-1}{|S|}^{-1} (v(S \cup \{i\}) - v(S))$$

— $\phi_i(v)$: La valeur de Shapley de la caractéristique i.
— $\mathcal{M}$: L'ensemble de toutes les caractéristiques du modèle.
— M : Le nombre de caractéristiques.
— S : Une coalition de caractéristiques, un sous-ensemble de $\mathcal{M}$.
— i : Une caractéristique spécifique.
— v : représente la sortie du modèle lors de l'utilisation de l'ensemble S.
— $\binom{M-1}{|S|}$: Une combinaison qui représente le nombre de façons de choisir $|S|$ caractéristiques parmi $M-1$ caractéristiques.

Pour les modèles d'apprentissage automatique, il n'est pas possible d'exclure simplement une caractéristique lors de la détermination d'une prédiction, surtout en présence de valeurs manquantes. La formulation des valeurs Shapley dans le contexte de l'apprentissage automatique simule l'effet d'exclusion d'une caractéristique en calculant la moyenne sur plusieurs échantillons de la distribution empirique des valeurs de la caractéristique.

En général, le processus de calcul des valeurs Shapley pour les modèles d'apprentissage automatique implique les étapes suivantes :

1. Pour chaque observation dans l'ensemble de données, simuler un ensemble de données "baseline" (ou "de base") en définissant toutes les caractéristiques à leur valeur moyenne. Cet ensemble sert de point de comparaison pour évaluer l'impact des caractéristiques individuelles.
2. Sélectionner une caractéristique spécifique et échantillonner aléatoirement ses valeurs à partir de la distribution empirique de cette caractéristique dans l'ensemble de données.
3. Pour chaque observation dans l'ensemble de données, perturber la caractéristique en remplaçant sa valeur par la valeur échantillonnée.
4. Utiliser l'ensemble de données perturbé pour générer des prédictions avec le modèle d'apprentissage automatique, et enregistrer la différence de prédiction par rapport à l'ensemble de données de référence.
5. Répéter les étapes 2 à 4 pour plusieurs échantillons des valeurs de la caractéristique, et moyenner les différences de prédiction sur tous les échantillons.

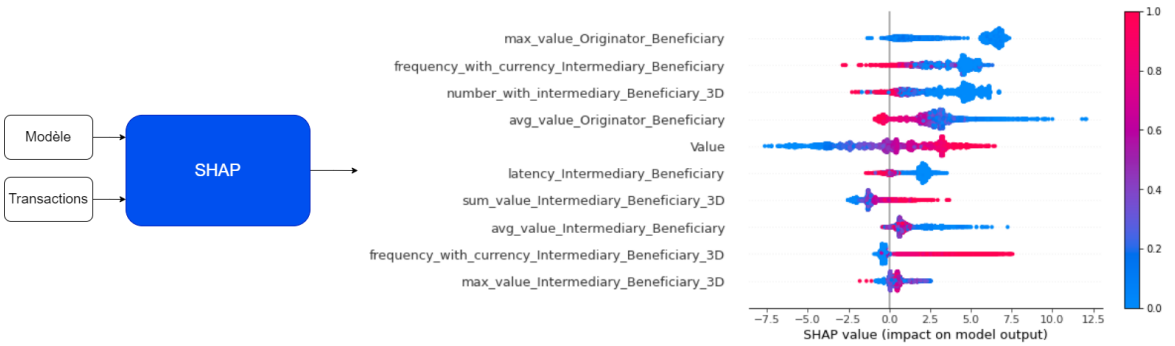

FIGURE 3.6 – Exemple d'utilisation de SHAP

6. Calculer la valeur Shapley comme la différence entre la prédiction moyenne sur toutes les observations et la prédiction moyenne sur toutes les observations avec la caractéristique exclue.

Ce processus peut être coûteux en termes de calcul, en particulier pour les ensembles de données de grande dimension avec un grand nombre de caractéristiques. Par conséquent, de nombreuses méthodes réelles pour calculer les valeurs Shapley pour les modèles d'apprentissage automatique utilisent différentes techniques pour réduire le coût de calcul et produire des estimations de la véritable valeur Shapley.

SHAP est un framework permettant d'utiliser les valeurs Shapley sur les modèles entraînés avec des arbres de décision. Dans la prochaine section, nous présentons ce framework et nous expliquons comment nous l'avons utilisé pour identifier les schémas de fraude.

3.5.4.2 Identification des clusters : SHAP et ses outils de visualisation

Pour associer les clusters à des schémas de fraude en utilisant les valeurs Shapley, nous utilisons l'outil de visualisation SHAP [LL17]. Cet outil permet de visualiser les valeurs Shapley d'un jeu de données à l'aide d'un modèle de classification et il est également compatible avec les algorithmes d'apprentissage ensemblistes basés sur les arbres de décision. Nous présentons un exemple dans la figure 3.6, le framework SHAP prend en entrée le modèle et les transactions avec leurs caractéristiques. SHAP donne une méthode de visualisation où nous pouvons observer les transactions qui sont représentées par des points. Nous pouvons observer :

— La couleur des points, qui représente la valeur des caractéristiques de chaque transaction. La couleur bleue représente une valeur faible et une couleur rouge, une valeur élevée.
— Leur position sur l'axe des abscisses qui représente la valeur Shapley de la transaction pour chaque caractéristique. Si la valeur de Shapley est positive, cela signifie que la caractéristique augmente la probabilité de la prédiction. Si elle est négative, cela indique que la caractéristique réduit la probabilité de la prédiction. Une valeur proche de zéro signifie que la caractéristique n'a pas eu un impact significatif sur la prédiction.

Les transactions de chaque cluster sont fournies au framework d'analyse et chaque cluster est analysé de manière indépendante. Grâce à l'utilisation d'outils de visualisation, nous observerons quelles caractéristiques ont eu le plus d'impact sur la prédiction de chaque transaction. Une caractéristique avec une valeur Shapley très élevée ou très faible indique qu'elle a joué un

rôle déterminant dans la prédiction du modèle pour cette transaction spécifique. En utilisant les valeurs Shapley ainsi que les caractéristiques associées, nous serons en mesure d'identifier les schémas de fraude pour chaque cluster. Les caractéristiques peuvent être liées à des acteurs, des pays ou des devises, en fonction de leur temporalité. Leur valeur et leur impact nous fourniront des informations essentielles pour l'identification et la compréhension des clusters. Ce travail a été publié dans le journal international IJACSA [CACC23b].

À la fin de cette étape, nous avons pour chaque transaction suspecte trois informations :
- Son score de suspicion
- Le schéma de fraude associé
- Les caractéristiques sélectionnées par le modèle pour réaliser la prédiction

Cependant, les experts chargés de mener les enquêtes disposent d'informations complémentaires, notamment sur les acteurs impliqués dans des transactions suspectes. Dans la prochaine section, nous présentons notre ontologie OntoSWIFT développée dans ce travail. Son objectif principal est de formaliser et structurer les connaissances relatives aux différents éléments et acteurs du domaine de la fraude financière, pour faciliter le travail de contrôle des experts.

3.6 Ontologie : aide à la décision sur les transactions suspectes

Nous présentons la dernière partie de notre approche qui vise à guider la décision de l'expert concernant les transactions classées comme suspectes par le modèle. Pour ce faire, nous utilisons des informations issues du modèle ainsi que des informations sémantiques sur les acteurs impliqués. Il est important de souligner que chaque transaction détectée par le système de détection de fraude doit être contrôlée par un expert afin de décider de sa validation ou de son blocage. Pour faciliter la prise de décision de l'expert, nous avons proposé une ontologie spécifiquement axée sur le domaine de la détection des transactions suspectes. Dans la suite de cette section, nous définissons les concepts, les propriétés, les axiomes et les règles de notre ontologie.

3.6.1 Les concepts : Acteur et Transaction

Une transaction SWIFT comporte plusieurs acteurs, ainsi notre ontologie a deux concepts principaux **Transaction** et **Acteur** qui sont illustrés dans la figure 3.7.

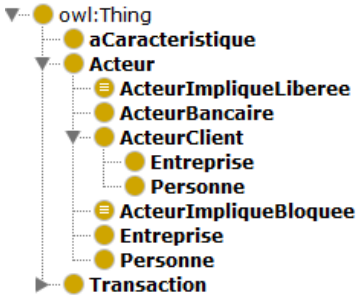

FIGURE 3.7 – Hiérarchie des concepts

Le concept **Transaction** représente les transactions suspectes détectées par notre modèle de classification. L'objectif de l'ontologie est de faciliter le contrôle de ces transactions, et en fonction de la décision de l'expert, elles seront soit bloquées, soit validées. Ainsi, le concept « Transaction » comporte deux sous-concepts : **TransactionLiberee** ou **TransactionBloquee**.

Le concept **Acteur** comporte quatre sous-concepts, le concept **ActeurClient** correspond à l'acteur qui va être ordonnateur ou bénéficiaire de la transaction. Ce concept comporte deux sous-concepts, un client peut être une personne (**ActeurPersonne**) ou une entreprise (**ActeurEntreprise**).

Le concept **ActeurBancaire** représente les banques des acteurs clients qui vont soit être la banque ordonnatrice, intermédiaire ou bénéficiaire d'une transaction.

Enfin, un acteur impliqué dans une transaction bloquée est catégorisé en tant que **ActeurImpliqueBloquee** tandis qu'un acteur impliqué dans plusieurs transactions libérées sans jamais avoir été impliqué dans une transaction est catégorisé en tant qu'**ActeurImpliqueLiberee**.

3.6.2 Les propriétés

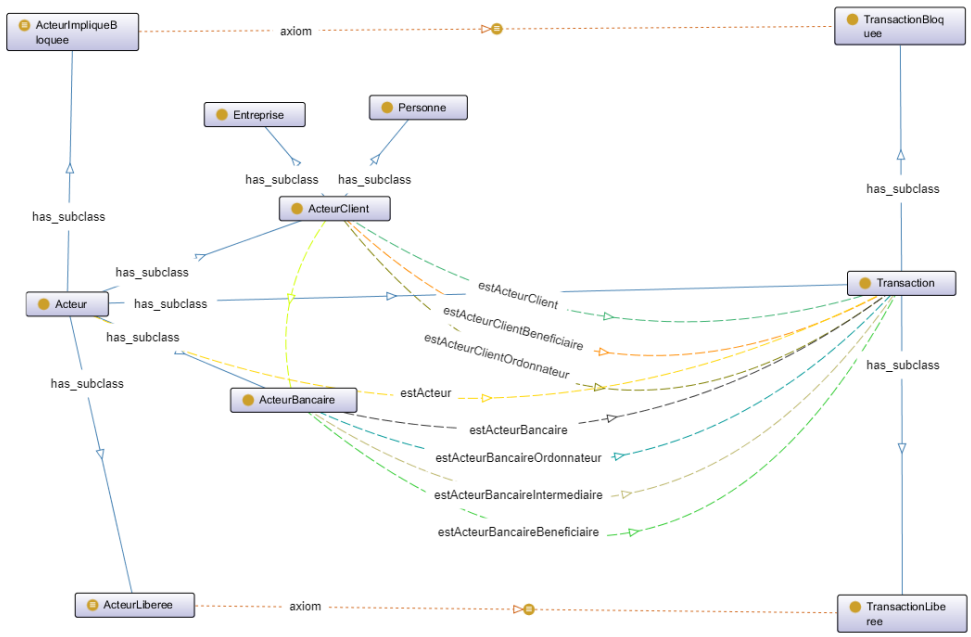

FIGURE 3.8 – Extrait de notre ontologie

Il existe deux types de propriétés : les propriétés d'objets représentant les relations entre les concepts, et les propriétés de données représentant les attributs des concepts. Dans la suite nous présentons les propriétés d'objets de notre ontologie dont un extrait est illustré dans a figure 3.8.

3.6.2.1 Propriété d'objets

Nous avons hiérarchisé les propriétés d'objets de notre ontologie comme illustré dans la figure 3.9. Nous rappelons qu'une transaction comprend deux acteurs clients et au moins deux acteurs

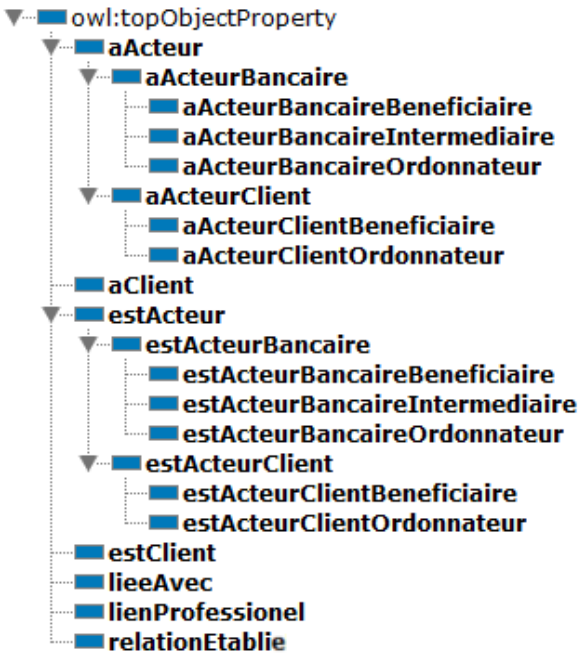

FIGURE 3.9 – Hiérarchie des propriétés d'objets

bancaires. La propriété **estActeur** correspond à la relation orientée entre le concept **Acteur** et **Transaction**.

Plus précisément, nous définissons sa sous-propriété **estActeurClient** qui correspond à la relation entre la transaction et ses acteurs clients. Cette sous-propriété comporte deux autres sous-propriétés : **estClientOrdonnateur** et **estClientBeneficiaire**, qui nous permettent de savoir qui sont les acteurs clients ordonnateurs et bénéficiaires de la transaction. De la même manière, la propriété **estActeurBancaire** est une sous-propriété de **estActeur**, et comporte trois sous-propriétés **estActeurBancaireOrdonnateur**, **estActeurBancaireIntermediaire** et **estActeurBancaireBeneficiaire** qui nous permettent de savoir qui sont les banques ordinatrices, intermédiaires et bénéficiaires d'une transaction.

Un client possède un compte bancaire dans une banque, cette relation est représentée par la propriété **estClient** ainsi que sa propriété inverse **aClient** qui correspondent à la relation entre un client **ActeurClient** et sa banque **ActeurBancaire**.

La propriété **lieeAvec** correspond une relation entre deux transactions (**Transaction**). La propriété **relationEtablie** est une relation entre deux banques lorsqu'elles possèdent une relation d'échange établie. Cette relation permet de savoir si une transaction entre ces deux banques nécessite un intermédiaire.

Dans le cadre de notre recherche, nous avons collaboré avec Jabbari et al. [JSZC20] pour approfondir leur travail, qui met en lumière l'importance de l'extraction d'informations sur les clients afin de répondre aux exigences du processus de *Know Your Customer* (KYC). Le KYC est une démarche réglementaire adoptée par les institutions financières pour prévenir les activités illicites. Il impose à ces institutions de collecter, vérifier et conserver des informations

précises sur l'identité et les activités de leurs clients, cela nous a conduits à la propriété objet **lienProfessionnel** entre les acteurs clients. L'ensemble du processus d'extraction d'informations et la base de connaissances KYC que nous avons développés dans le cadre du projet KBP sont présentés en détail dans l'annexe A.

3.6.2.2 Propriétés de données

Les propriétés des données définissent des attributs spécifiques et leurs valeurs pour les concepts représentés. Dans notre contexte, ces propriétés représentent des informations essentielles sur les transactions et les acteurs. Ainsi, nous exposons ici la hiérarchie structurée des propriétés en fonction des concepts auxquels elles sont associées. La figure 3.10 illustre les propriétés de données. Nous avons hiérarchisé les propriétés de données de notre ontologie comme illustré dans la figure 3.10 avec les caractéristiques retenue par notre modèle. Ces caractéristiques ajoutent des informations complémentaires aux transactions, pouvant ainsi aider les experts dans leur analyse.

FIGURE 3.10 – Hiérarchie des propriétés de données

ActeurPersonne : Une personne possède plusieurs informations telles que son nom, son prénom, son adresse ou son rôle dans son entreprise. Nous avons regroupé ces informations dans la propriété de données **aInformationPersonne** représentant les informations personnelles des personnes **ActeurPersonne**, avec les sous-propriétés **aAdresse** (adresse de la personne), **aNom** (nom de la personne), **aPrenom** (prénom de la personne) et **aProfession** (profession de la personne dans son entreprise).

ActeurEntreprise : Une entreprise possède des informations telles son secteur d'activité, son nom et son pays. Ces informations sont représentées dans la propriété de données **aInfor-**

mationEntreprise représentant les informations sur les entreprises **ActeurEntreprise**, avec les sous-propriétés **aNomEntreprise** (nom de l'entreprise), **aSecteurEntreprise** (secteur d'activité de l'entreprise) et **aPaysEntreprise** (pays de l'entreprise).

ActeurBancaire : Un acteur bancaire est une banque pouvant échanger des fonds avec d'autres banques, elle est représentée par un code BIC. Les informations des banques sont représentées par la propriété de données **aInformationBancaire** qui comporte deux sous-propriétés **aBIC** représentant l'identifiant d'une banque (**ActeurBancaire**) et **aPays** correspondant au pays où la banque est située.

ActeurClient : Un acteur client est un client qui fait appel à sa banque pour réaliser une transaction, il est représenté dans sa banque par un numéro de compte. Les informations des acteurs clients sont représentées par la propriété de données **aInformationClient**. Cette propriété possède deux sous-propriétés **aNumeroCompte** qui représente le numéro de compte d'un client (**ActeurClient**) dans une banque, et **aDateCompte** représentant la date d'ouverture du compte du client dans sa banque.

Transaction : Une transaction contient plusieurs informations de base telles que la devise et le montant, ainsi que des informations sur sa suspicion d'être frauduleuse par le modèle et des informations sur ses caractéristiques. Ces informations sont représentées par la propriété de données **aChamp** qui comporte les sous-propriétés suivantes : **aBase** (informations de base de la transaction), **aCaracteristique** (caractéristiques retenues par le modèle) et **aFraude** (informations sur les transactions suspectes).
aBase représente les informations de base de la transaction (**Transaction**). Elle comprend les sous-propriétés suivantes : **aDate** (date de la transaction), **aDevise** (devise de la transaction) et **aMontant** (montant de la transaction).
La sous-propriété **aFraude** représente les informations relatives aux transactions suspectes. Elle inclut les sous-propriétés suivantes : **aScoreSuspicion** (score de suspicion de la transaction) et **aSchema** (schéma de fraude de la transaction).
Il est important de souligner qu'initialement la propriété **aCaracteristique** n'inclut aucune sous-propriété et qu'après l'entraînement du modèle, nous enrichissons **aCaracteristique** en y intégrant les caractéristiques sélectionnées par le modèle, dans l'étape de réduction des caractéristiques, comme des sous-propriétés de données.

3.6.3 Les axiomes : le statut des acteurs

Pour représenter les concepts des statuts des acteurs impliqués dans des transactions bloquées ou libérées, nous avons défini des axiomes exprimés en syntaxe formelle basée sur la logique de description. **ActeurImpliqueBloquee** ou **ActeurImpliqueLiberee**.

Un acteur de type **ActeurImpliqueBloquee** est un acteur qui a été impliqué dans au moins une transaction qui a été bloquée par un expert, qui est défini comme suit :

$\textbf{ActeurImpliqueeBloquee} \equiv \textbf{Acteur} \cap (\geq 1\ \textbf{estActeur} \cdot \textbf{TransactionBloquee})$
$\textbf{ActeurImpliqueeBloquee} \equiv \textbf{Acteur} \cap \geq 1\ \textbf{estActeur}.\textbf{TransactionBloquee}$

Un acteur de type **ActeurImpliqueLiberee** est un acteur qui a été impliqué dans au moins trois transactions libérées et dans aucune transaction bloquée qui est définie comme suit :

ActeurImpliqueLiberee $\equiv$ **Acteur** ≥ 3 **estActeur**.**TransactionLiberee**$\cap$
$\neg$**ActeurImpliqueeBloquee**

3.6.4 Les règles : le statut des transactions

Après avoir défini les concepts, les propriétés d'objets et de données, nous avons collaboré avec des experts du domaine pour formaliser des règles en utilisant le langage SWRL (Semantic Web Rule Language). Ce choix s'est imposé pour son adéquation avec notre ontologie, facilitant de cette manière l'automatisation du processus de détection et de validation des transactions suspectes. Dans la suite, nous présenterons des exemples de règles pour illustrer la manière dont elles guident le processus de contrôle.

Règle 1 : *Si une transaction suspecte possède un score de suspicion supérieur à 90, alors cette transaction est bloquée*
Règle SWRL : **Transaction**$(?t) \wedge$**aScoreSuspicion**$(?t, ?p) \wedge$**greaterThanOrEqual**$(?p, 90) \rightarrow$ **TransactionBloquee**$(?t)$

Règle 2 : *Si une transaction suspecte a un acteur qui a au moins une transaction bloquée, alors la transaction est bloquée*
Règle SWRL : **Transaction**$(?t) \wedge$ **aActeur**$(?t, ?a) \wedge$ **estActeur**$(?a, ?t2) \wedge$ **TransactionBloquee**$(?t2) \rightarrow$ **TransactionBloquee**$(?t)$

Règle 3 : *Si une transaction suspecte est connectée à une transaction bloquée, alors la transaction est bloquée*
*Règle SWRL :***Transaction**$(?t) \wedge$ **lieeAvec**$(?t, ?t2) \wedge$ **TransactionBloquee**$(?t2) \rightarrow$ **TransactionBloquee**$(?t)$

Règle 4 : *Si une transaction suspecte a un acteur possédant un lien professionnel avec un acteur ayant au moins une transaction bloquée, alors la transaction est bloquée*
Règle SWRL : **Transaction**$(?t) \wedge$ **aActeur**$(?t, ?acteur) \wedge$ **lienProfessionnel**$(?acteur, ?acteur2) \wedge$ **estActeur**$(?acteur2, ?t2) \wedge$ **TransactionBloquee**$(?t2) \rightarrow$ **TransactionBloquee**$(?t)$

En résumé, notre ontologie se positionne au cœur de la stratégie de détection et du contrôle des transactions suspectes. Son objectif principal est de fournir une structure sémantique claire et robuste pour appuyer le processus de décision. Grâce à elle, non seulement nous pouvons représenter efficacement les informations relatives aux transactions et aux acteurs, mais aussi établir des relations et des règles qui reflètent les pratiques et les connaissances des experts du domaine. En intégrant cette ontologie, notre approche favorise une prise de décision plus rapide et automatisée des transactions suspectes. Ce travail à fait l'œuvre de publications de conférence nationales [ACC+22, ACA+23] et d'une internationale [ACA].

3.7 Conclusion

Dans ce chapitre, nous avons présenté notre approche hybride pour la détection et la gestion des transactions suspectes. Notre démarche repose sur une combinaison de techniques d'apprentissage automatique et sur une ontologie spécialement élaborée pour le domaine d'étude.

Cette approche est décomposée en plusieurs étapes distinctes. Tout d'abord, nous procédons à une analyse des distributions, des fréquences et des montants des transactions légitimes et frauduleuses afin d'identifier les schémas potentiels de fraude. Ensuite, nous calculons des caractéristiques spécifiques aux transactions SWIFT, telles que les pays, les devises et les intermédiaires, en tenant compte de différentes échelles temporelles. Nous utilisons ensuite les arbres de décision pour évaluer l'importance de chaque caractéristique et sélectionner les plus pertinentes.

Une fois que les caractéristiques sont calculées et sélectionnées, nous appliquons un algorithme de clustering hiérarchique pour regrouper les transactions frauduleuses en fonction des schémas de fraude identifiés. Cette étape nous permet de mieux comprendre les comportements frauduleux et facilite l'identification des schémas de fraude. Ensuite, nous procédons à la classification des transactions à l'aide d'un modèle entraîné par un algorithme d'apprentissage ensembliste basé sur les arbres de décision, tout en réduisant la dimension des données. Ce modèle de classification multi-classes distingue les transactions légitimes des transactions associées aux schémas de fraude. En utilisant les prédictions de probabilité d'appartenance aux classes frauduleuses, nous attribuons un score de suspicion pour quantifier la probabilité de fraude de chaque transaction. Nous évaluons également les coûts financiers liés aux prédictions et minimisons le coût de notre modèle en sélectionnant le seuil de suspicion optimal à partir duquel une transaction est considérée comme suspecte.

Enfin, nous avons développé une ontologie pour décrire le domaine en définissant les concepts et les propriétés pertinents pour notre étude. L'objectif principal de l'ontologie est de contrôler toutes les transactions suspectes détectées par le modèle. À cet effet, nous utilisons des règles d'inférence pour libérer ou bloquer des transactions suspectes en exploitant les connaissances sur les acteurs provenant de sources externes.

À travers notre approche, nous proposons une solution novatrice en combinant les techniques d'apprentissage automatique avec une ontologie, afin de contribuer à la lutte contre la fraude financière. Notre démarche se concentre spécifiquement sur les transactions effectuées via le réseau SWIFT, et nous accordons une attention particulière à la manière dont nous pouvons enrichir les informations disponibles pour les transactions identifiées comme frauduleuses par notre modèle de classification. Nous avons pris en considération les besoins des experts chargés de la vérification de ces transactions et nous avons développé une ontologie qui représente de manière exhaustive le domaine de la détection de la fraude. Cette ontologie intègre des informations interprétables par les experts, telles que le score de suspicion, les schémas de fraude identifiés, ainsi que des données provenant de sources externes, offrant ainsi une perspective plus complète pour la prise de décisions.

Chapitre 4

Expérimentations et évaluations

Sommaire

4.1 Introduction

Nous décrivons dans ce chapitre les expérimentations que nous avons menée au cours de cette thèse. L'objectif est l'évaluation de notre approche présentée dans le chapitre 3. Nous débutons par une analyse approfondie du jeu de données, en examinant les caractéristiques des transactions et des acteurs. Ensuite, nous abordons le calcul et la sélection des caractéristiques. Nous procédons ensuite au clustering des transactions frauduleuses et à l'identification des schémas de fraude. Les résultats de la classification sont présentés dans la section 4.5. Nous détaillons les résultats des différents modèles pour ensuite analyser les schémas de fraude des transactions classées suspectes. La section 4.6 expose l'utilisation de l'ontologie que nous avons élaborée pour l'aide à la décision. Enfin, dans la section 4.7, nous présentons ST-Fraud, l'outil que nous avons développé pour la détection et l'interprétation des transactions frauduleuses.

4.2 Analyse exploratoire du jeu de données

L'analyse exploratoire représente le point de départ essentiel dans notre approche, visant à obtenir une compréhension approfondie du jeu de données tout en mettant en évidence d'éventuels schémas de fraude potentiels. Cette phase initiale facilite la mise en œuvre ultérieure des techniques que nous allons utiliser en nous fournissant des informations pour orienter notre démarche. Pour mener à bien nos expérimentations, nous nous appuyons sur deux jeux de données clés : $SWIFT_{base}$ et $SWIFT_{base+syn}$. Le premier jeu de données, obtenu en collaboration avec l'entreprise SKaizen Group, est composé de $3\ 500\ 000$ transactions, parmi lesquelles $12\ 756$ sont classées comme frauduleuses, représentant ainsi $0,0034\%$ du total des transactions. Concernant le deuxième jeu de données, nommé $SWIFT_{base+syn}$, il est élaboré en enrichissant $SWIFT_{base}$ avec des transactions frauduleuses synthétiques. Ces transactions synthétiques sont générées selon un schéma de fraude précis, défini en collaboration avec des experts du domaine. Ce schéma se décrit comme suit : *deux acteurs effectuent plusieurs transactions légitimes avec un même intermédiaire. Ensuite, une transaction frauduleuse est générée, se caractérisant par un changement d'intermédiaire*. Un changement soudain d'intermédiaire entre deux acteurs qui ont déjà établi plusieurs transactions entre eux peut être interprété comme un indicateur potentiel de comportement frauduleux. En effet, cette action peut être entreprise dans le but de dissimuler une transaction en optant pour une banque dotée de systèmes de contrôle de fraude moins performants.

Dans le tableau 4.1, nous illustrons quelques transactions à titre d'exemples, conformément au format de notre jeu de données. Il est essentiel de souligner que les données concernant les acteurs, les pays et les devises ont été anonymisées, et que les informations spécifiques sur la nature des transactions frauduleuses ou leur schéma ne sont pas disponibles. Par ailleurs, les montants des transactions ont été uniformisés en étant tous convertis dans une devise unique, dont la nature exacte nous est inconnue.
Il convient de rappeler que les institutions financières sont identifiées par des codes BIC, dont les 4ème et 5ème caractères représentent le code pays. Ainsi, nous disposons de trois caractéristiques catégorielles qui sont les pays des banques ordonnatrices, intermédiaires et bénéficiaires.

Dans la suite de notre étude, nous nous focalisons sur l'analyse des transactions frauduleuses du jeu de données $SWIFT_{base}$. Le jeu de données $SWIFT_{base+syn}$ comporte des transactions

TABLE 4.1 – Exemple de transactions avec les données anonymisées. *Bq : banque, L : légitime, F : frauduleuse*

Bq ordonnatrice	Bq intermédiaire	Bq bénéficiaire	Date	Devise	Montant	Classe
BIC0KZ01 (KZ)	BIC0ET01 (ET)	BIC0LP02 (LP)	210625	ADF	15006	L
BICPOS03 (PO)	-	BIC0ST01 (ST)	210625	POD	33065	L
BIC0KZ04 (KZ)	BIC0WO06 (WO)	BIC0PO05 (PO)	210626	SDI	100325	F

frauduleuses synthétiques dont nous connaissons déjà la nature de leur fraude, ainsi il n'est pas nécessaire d'analyser ce jeu de données.

4.2.1 Analyse des caractéristiques catégorielles

Dans cette section, nous analysons les caractéristiques catégorielles de notre jeu de données. Ces caractéristiques, incluant les acteurs impliqués tels que les banques ordonnatrices, les intermédiaires et les bénéficiaires, ainsi que les pays associés à ces institutions et les devises utilisées dans les transactions, sont essentielles. Elles peuvent en effet nous assister dans la détection de schémas de fraude au sein de notre ensemble de données. Le tableau 4.2 ainsi que la figure 4.1 proposent une analyse approfondie des caractéristiques catégorielles de notre ensemble de données.

TABLE 4.2 – Descriptif des caractéristiques

Caractéristiques	**Nombre total**	**Nombre d'éléments impliqués dans une fraude**	**Nombre d'éléments légitimes**
Acteur banque ordonnatrice	30	30	0
Acteur banque intermédiaire	10 410	5 021	5 389
Acteur banque bénéficiaire	13 236	7 010	6 226
Pays banque ordonnatrice	26	26	0
Pays Banque intermédiaire	248	248	0
Pays Banque bénéficiaire	248	248	0
Devise	151	151	0

Le tableau 4.2 énumère, pour chaque caractéristique, le nombre total d'éléments uniques et distingue le nombre d'éléments (acteurs, pays, devises) impliqués dans des transactions frauduleuses de ceux impliqués dans des transactions légitimes. Quant à la figure 4.1, elle présente deux graphiques distincts. Sur l'axe des ordonnées, ces graphiques représentent le nombre d'éléments, en différenciant les éléments impliqués dans des transactions frauduleuses (affichés en orange) de ceux associés exclusivement à des transactions légitimes (affichés en violet). En ce qui concerne l'axe des abscisses, nous retrouvons les éléments étudiés : dans le graphique de gauche, il s'agit des acteurs bancaires (émetteur, bénéficiaire et intermédiaire), tandis que dans le graphique de droite, nous observons les pays et les devises.
Notons plusieurs observations importantes :

— Nous comptons seulement 30 banques ordonnatrices présentes dans notre ensemble de données, et aucune d'entre elles n'est exclusivement associée à des transactions

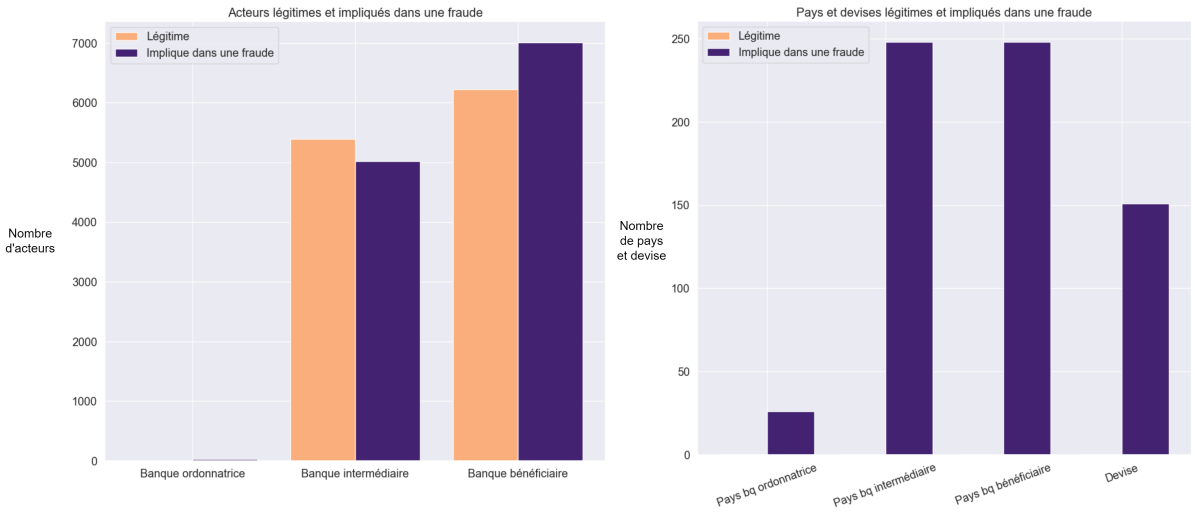

FIGURE 4.1 – Répartition des acteurs, pays et devises entre les transactions légitimes et frauduleuses

légitimes.

— Parmi les 10 410 acteurs banques intermédiaires, 5 021 sont impliqués dans des transactions frauduleuses et 5 389 sont liés à des transactions légitimes.
— Sur les 13 236 acteurs des banques bénéficiaires, 7 010 sont impliqués dans des transactions frauduleuses et 6 226 dans des transactions légitimes.
— Les 26 pays associés aux banques ordonnatrices ainsi que les 248 pays associés aux banques intermédiaires et bénéficiaires sont tous impliqués dans des transactions frauduleuses, d'après nos données.
— Les 151 devises répertoriées dans notre ensemble de données sont toutes utilisées dans des transactions frauduleuses.

Suite à cette analyse, il est important de noter que la distribution des transactions au sein des banques intermédiaires et bénéficiaires est relativement équilibrée entre les transactions frauduleuses et légitimes. Cet équilibre offre une opportunité d'analyse intéressante. Ainsi, dans la section suivante, nous orientons notre étude vers l'analyse des transactions des banques intermédiaires et bénéficiaires.

4.2.2 Analyse du comportement transactionnel des banques intermédiaires et bénéficiaires

Dans cette section, nous portons notre attention sur les banques intermédiaires et bénéficiaires, en mettant l'accent sur la fréquence de leurs transactions. L'objectif est d'étudier le profil transactionnel de ces banques et de déterminer si la fréquence des transactions est un indicateur pertinent pour détecter des comportements frauduleux.
Le tableau 4.3 présente la répartition des banques intermédiaires et bénéficiaires en fonction du nombre de transactions qu'elles traitent. Avec l'aide de nos experts, nous avons séparé nos acteurs en trois groupes d'intervalle : les acteurs réalisant exactement une transaction, ceux réalisant entre 2 et 10 et ceux réalisant plus de 10. Pour chaque intervalle de fréquence de transactions, le tableau indique le nombre total de banques, le nombre de ces banques

impliquées dans des transactions frauduleuses et le nombre de banques qui semblent n'opérer que des transactions légitimes.
La figure 4.2, quant à elle, illustre graphiquement ces données. Les barres en orange représentent le nombre de banques impliquées dans des transactions frauduleuses, et les barres en violet représentent le nombre de banques effectuant des transactions légitimes, pour chaque intervalle de fréquence de transactions.

TABLE 4.3 – Nombre de banques intermédiaires et bénéficiaires légitimes et impliqués dans une fraude par fréquence

Fréquence	**Nombre total**	**Nombre de banques impliquées dans une fraude**	**Nombre de banques légitimes**
Exactement une transaction	3 700	2 800	900
Entre 2 et 10 transactions	604	600	4
Plus de 10 transactions	5 900	3 700	2 200

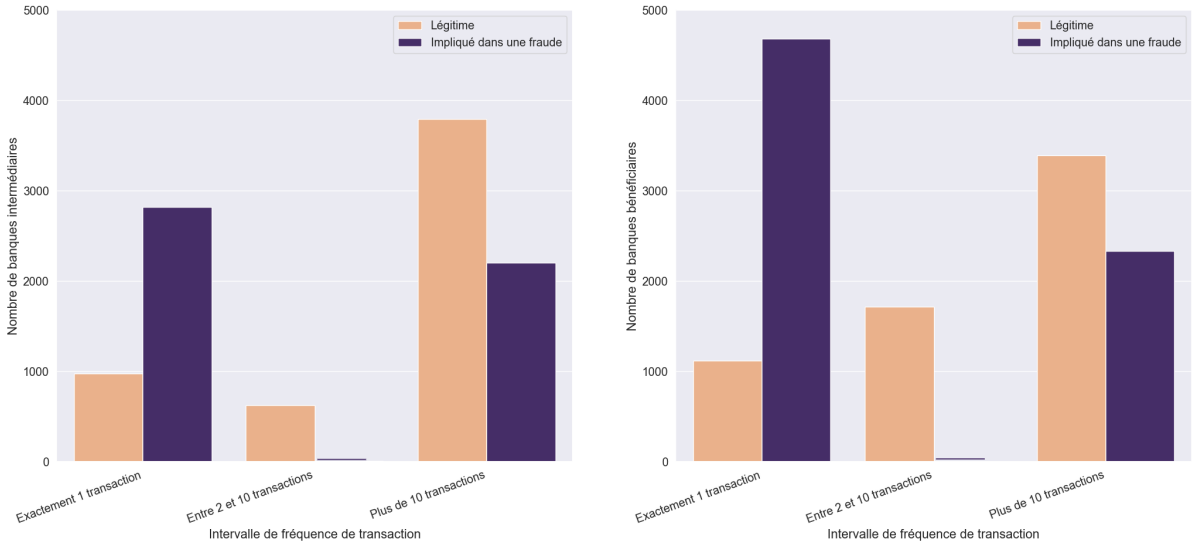

FIGURE 4.2 – Nombre d'intermédiaires et bénéficiaires légitimes et impliqués dans une fraude par fréquence

Il est à noter que parmi les banques n'ayant effectué qu'une seule transaction, un grand nombre d'entre elles (2 800 sur 3 700) sont impliquées dans une transaction frauduleuse. Cette observation suggère que les acteurs frauduleux pourraient utiliser des comptes de banques intermédiaires et bénéficiaires pour effectuer une unique transaction frauduleuse, puis abandonner ces comptes. En revanche, pour les banques ayant effectué plus de 10 transactions, la distribution entre banques impliquées dans des fraudes et banques légitimes est plus équilibrée, indiquant que ces banques sont probablement des entités opérationnelles régulières.

Ces observations mettent en lumière deux groupes d'acteurs distincts : le premier groupe est constitué d'acteurs qui réalisent une seule transaction, parmi lesquels un nombre significatif

(75%) est impliqué dans une fraude. Le second groupe, quant à lui, englobe les acteurs qui réalisent plus de 10 transactions, avec un taux d'implication dans les fraudes de 36%.

Dans la prochaine section, nous approfondissons notre analyse en nous focalisant sur le premier groupe d'acteurs, c'est-à-dire ceux qui réalisent une seule transaction. Nous cherchons à comprendre les spécificités de ce groupe et à déterminer pourquoi une proportion aussi élevée d'acteurs sont impliqués dans des transactions frauduleuses.

4.2.3 Analyse des acteurs réalisant une seule transaction

L'importance des acteurs qui n'effectuent qu'une seule transaction dans le contexte de la fraude est déjà établie. Cependant, pour approfondir notre compréhension, il est essentiel d'examiner les montants des transactions qu'ils réalisent. La répartition des montants peut révéler des tendances qui ne sont pas immédiatement apparentes lorsque l'on se base uniquement sur le nombre de transactions.

Le tableau 4.4 présente les montants moyens des transactions, à la fois légitimes et frauduleuses, en fonction de la fréquence des transactions des intermédiaires et des bénéficiaires. Les valeurs représentent la moyenne des montants des transactions réalisées par les acteurs. En parallèle, la figure 4.3 nous donne une représentation graphique de ces montants. Chaque diagramme correspond à un groupe (intermédiaires à gauche, bénéficiaires à droite), avec les montants moyens des transactions sur l'axe des ordonnées et les fréquences sur l'axe des abscisses. Les transactions légitimes sont illustrées en orange, tandis que les transactions frauduleuses le sont en violet.

TABLE 4.4 – Moyenne des montants des transactions des intermédiaires et bénéficiaires par intervalles de fréquence

Fréquence	**Montants moyens des transactions légitimes**		**Montants moyens des transactions frauduleuses**	
	Intermédiaire	Bénéficiaire	Intermédiaire	Bénéficiaire
Exactement une transaction	9 050	9 370	258 050	259 440
Entre 2 et 10 transactions	9 960	9 450	-	-
Plus de 10 transactions	277 580	277 340	305 700	305 410

Nous pouvons réaliser les observations suivantes :

- Acteurs réalisant une seule transaction : Pour les intermédiaires ou les bénéficiaires, les montants moyens des transactions frauduleuses (258 050 et 259 440) sont nettement supérieurs à ceux des transactions légitimes (9 050 et 9 370). Cela renforce l'idée que ce groupe, est lié à un schéma de fraude.
- Acteurs réalisant entre 2 et 10 transactions : Une différence de montants est également constatée dans cet intervalle, cependant, les résultats ne sont pas pris en compte en raison du faible nombre d'acteurs impliqués dans une fraude (seulement 4).
- Acteurs réalisant plus de 10 transactions : Les montants moyens sont très similaires entre les transactions frauduleuses (305 700 et 305 410) et légitime (277 580 et 277 340).

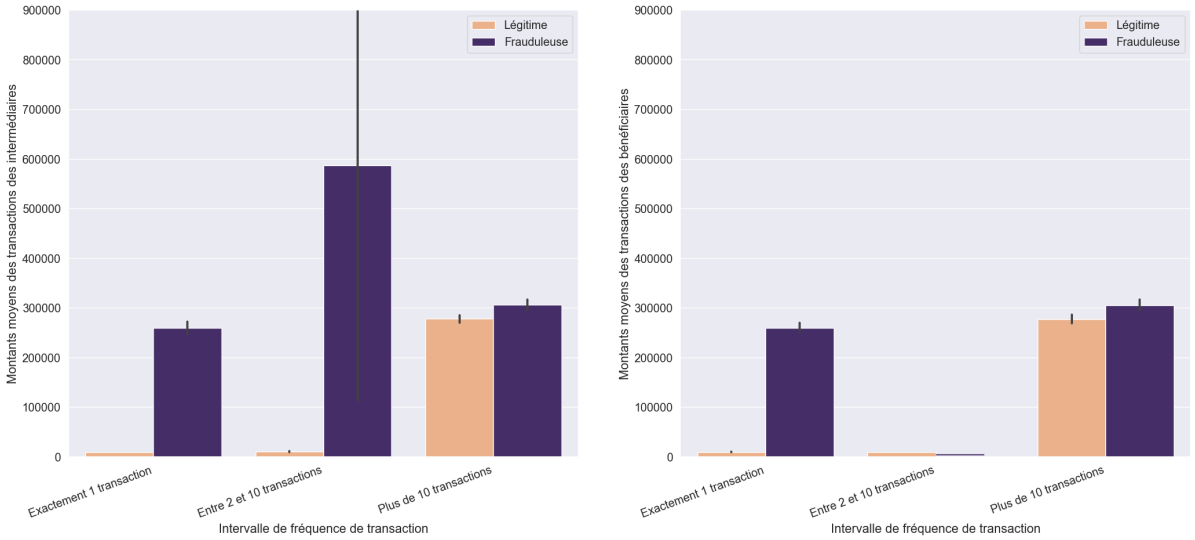

FIGURE 4.3 – Moyenne des montants des transactions des intermédiaires et bénéficiaires par intervalles de fréquence

Ceci suggère que d'autres indicateurs pourraient être plus pertinentes pour détecter les transactions frauduleuses pour ce groupe.

Ainsi, grâce à cette analyse et en collaboration avec des experts, nous avons identifié un schéma de fraude spécifique au groupe d'acteurs effectuant une seule transaction. Ce schéma implique de nouveaux acteurs qui créent des comptes, effectuent une transaction, souvent d'un montant élevé, puis abandonnent ces comptes.

Dans la prochaine section, nous porterons notre attention sur le groupe d'acteurs réalisant plus de 10 transactions, afin d'y déceler d'éventuelles tendances ou schémas de fraude.

4.2.4 Analyse des acteurs réalisant plus de 10 transactions

Dans cette section, nous étudions l'aspect temporel des transactions afin de détecter d'éventuelles tendances parmi les acteurs impliqués dans des fraudes. Nous examinons le temps moyen entre deux transactions pour chaque acteur, en distinguant les transactions légitimes des transactions frauduleuses. L'objectif est de déterminer si les transactions frauduleuses sont réalisées avec un intervalle de temps plus court que les transactions légitimes. Pour cela, nous utilisons l'attribut de la date à laquelle les transactions ont été réalisées.

4.2.4.1 Temps moyen entre deux transactions

Le tableau 4.5 présente le temps moyen entre deux transactions, à la fois pour les intermédiaires et les bénéficiaires, en distinguant entre les acteurs légitimes et ceux impliqués dans une fraude.

D'après le tableau 4.5, nous pouvons observer que les acteurs impliqués dans des fraudes effectuent des transactions plus fréquemment, avec des délais moyens de 710 et 421 minutes, tandis que les acteurs légitimes ont des délais moyens plus élevés de 900 et 2033 minutes. Ainsi, pour approfondir notre analyse, nous allons étudier les transactions réalisées avec un

TABLE 4.5 – Temps moyen entre deux transactions

Acteur / Type	Légitime	Impliqué dans une fraude
Bénéficiaire	900 minutes	710 minutes
Intermédiaire	2033 minutes	421 minutes

intervalle de moins de 60 minutes, car un laps de temps aussi court peut être un indicateur d'une activité potentiellement suspecte.

4.2.4.2 Transactions à moins de 60 minutes d'intervalle

Dans notre étude, nous avons également cherché à comprendre la fréquence à laquelle les acteurs effectuent des transactions sur une courte période. Plus précisément, nous avons examiné le nombre moyen de transactions réalisées par un acteur dans un intervalle de moins de 60 minutes. Le tableau 4.6 résume ces informations, en différenciant les transactions légitimes et frauduleuses, et en mettant en évidence les transactions effectuées avec un écart inférieur à 60 minutes.

TABLE 4.6 – Nombre de transactions moyen d'un acteur total et avec moins de 60 minutes

Acteur / Type	Transactions légitimes		Transactions frauduleuses	
	Nombre	Nombre avec moins de 60mn	Nombre	Nombre avec moins de 60mn
Bénéficiaire légitime	575	220	0	0
Bénéficiaire impliqué dans une fraude	733	290	3	2
Intermédiaire légitime	250	60	0	0
Intermédiaire impliqué dans une fraude	1224	850	4	3

Nous pouvons formuler les observations suivantes :

- Pour les bénéficiaires :
  - Les bénéficiaires légitimes ont en moyenne un nombre de transactions de 575, parmi lesquelles 220 sont effectuées à un intervalle inférieur à 60 minutes.
  - Les bénéficiaires impliqués dans des fraudes présentent un comportement légèrement différent, avec une moyenne totale de 733 transactions. Parmi celles-ci, 290 transactions sont réalisées à un intervalle inférieur à 60 minutes. De plus, ces acteurs affichent une moyenne de 3 transactions frauduleuses, dont 2 sont effectuées dans un court laps de temps.
- Pour les intermédiaires :
  - Les intermédiaires légitimes ont en moyenne 250 transactions, parmi lesquelles seulement 60 sont effectuées en moins de 60 minutes.
  - En revanche, les intermédiaires impliqués dans des fraudes présentent un nombre de transactions nettement plus élevé, avec une moyenne de 1224. Parmi ces transactions, 850 sont effectuées dans un intervalle inférieur à 60 minutes. De plus, ces

intermédiaires ont une moyenne de 4 transactions frauduleuses, dont 3 sont rapprochées en termes de temps.

D'après ces observations, il apparaît que les acteurs impliqués dans des fraudes, qu'ils soient bénéficiaires ou intermédiaires, ont tendance à effectuer des transactions à des intervalles plus rapprochés que leurs homologues légitimes. Cette tendance plus prononcée vers des transactions rapprochées, particulièrement dans le cas des transactions frauduleuses, suggère un comportement d'urgence ou un mécanisme de fraude où l'acteur tente d'effectuer plusieurs transactions avant d'être détecté.

4.2.5 Schémas de fraude identifiés

Au cours de notre analyse exploratoire, nous avons identifié deux schémas principaux de fraude :

1. Acteurs occasionnels avec montants de transactions élevés : Le premier schéma concerne les acteurs qui réalisent une seule transaction avec un montant très élevé. Nos experts ont déjà observé que de nombreux acteurs nouvellement apparus créent des comptes, effectuent des transactions (souvent frauduleuses), puis abandonnent ces comptes. Ce comportement pourrait être le fait d'acteurs malveillants cherchant à exploiter le système en établissant rapidement une fraude sans établir de trace durable ou en évitant la détection grâce à la non-récurrence de leur activité.
2. Transactions rapprochées dans le temps : Le second schéma est caractérisé par des transactions réalisées à des intervalles très rapprochés avec des montants similaires à ceux des transactions légitimes, en particulier chez les acteurs impliqués dans des fraudes. Ces acteurs, qu'ils soient bénéficiaires ou intermédiaires, tendent à effectuer des transactions à des intervalles de moins de 60 minutes plus fréquemment que leurs homologues légitimes. Ce comportement suggère un mécanisme où l'acteur malveillant tente de réaliser un maximum de transactions avant d'éventuellement être repéré et bloqué. Cela pourrait aussi indiquer une tentative de profiter d'une fenêtre d'opportunité limitée ou d'exploiter un maillon faible détecté dans le système pendant une courte période.

Après avoir identifié ces schémas distinctifs de fraude, nous entamons la prochaine phase de l'ingénierie des caractéristiques. Cette étape permet de calculer et sélectionner les caractéristiques pour détecter les transactions frauduleuses appartenant aux schémas de fraude.

4.3 Calcul et sélection des caractéristiques

Nous avons extrait des caractéristiques de nos deux jeux de données en utilisant les informations du tableau 3.4 présenté dans la section 3.2. Ces caractéristiques ont été dérivées à deux échelles temporelles : annuelle et sur une période de 3 jours. L'échelle annuelle vise à modéliser le comportement global des acteurs et des transactions à travers différents pays et devises. En revanche, l'échelle de 3 jours offre une modélisation plus fine de ces comportements. Cette phase d'enrichissement des données a considérablement augmenté la dimensionnalité de nos jeux de données, atteignant désormais 513 dimensions.

Comme expliqué dans la section 3.3.2.1, nous avons adopté une méthode de sélection de caractéristiques intégrée. Cette méthode consiste à entraîner un modèle d'apprentissage automatique

qui attribue un score d'importance à chaque caractéristique. Dans les algorithmes basés sur des arbres de décision, ce score est généralement déterminé par le nombre de fois qu'une caractéristique est utilisée pour diviser les données, ainsi que par l'amélioration apportée au modèle à chaque division. Les caractéristiques ayant les scores les plus élevés sont alors retenues. Dans ce contexte, nous avons opté pour l'algorithme Catboost pour les raisons suivantes :

— Sa performance avérée en classification de transactions frauduleuses, confirmée par plusieurs études dans la littérature [AF22, GOV22].
— Son utilisation des arbres de décision, ce qui élimine le besoin de normalisation ou de mise à l'échelle des données.
— Sa capacité à gérer directement les caractéristiques catégorielles, un aspect crucial pour notre étude.

Pour faciliter la compréhension des caractéristiques, nous présentons dans le tableau 4.7 la liste des caractéristiques avec leur description. Il faut savoir que les caractéristiques sont calculées pour chaque transaction. Par exemple, pour une transaction donnée avec la caractéristique « avg_value_Beneficiary_Currency_1Y » nous rajoutons à la transaction le montant moyen de son bénéficiaire dans la devise de la transaction durant l'année qui précède cette dernière.

Dans la section suivante, nous présentons les 10 caractéristiques sélectionnées avec leurs scores d'importance pour les jeux de données SWIFT$_{base}$ et SWIFT$_{base+syn}$. Nous rappelons que SWIFT$_{base}$ est le jeu de données de base et que SWIFT$_{base+syn}$ est composé du jeu de données de base dans lequel nous avons ajouté des transactions synthétique suivant le schéma de fraude d'un changement soudain d'intermédiaire entre une banque ordonnatrice et bénéficiaire.

4.3.1 SWIFT$_{base}$

Les résultats de la sélection de caractéristiques par ordre décroissant d'importance pour le jeu de données SWIFT$_{base}$ sont présentés dans le tableau 4.8.

TABLE 4.8 – Caractéristiques sélectionnées pour SWIFT$_{base}$ avec leurs scores d'importance

Caractéristique	**Score d'Importance**
value	6.7
avg_value_Beneficiary_Currency_1Y	2.8
sum_value_Beneficiary_IntermediaryCountry_3D	2.1
sum_value_Beneficiary_Currency_1Y	1.8
sum_value_Beneficiary_Currency_3D	1.2
max_value_Beneficiary_Currency_1Y	1.2
sum_value_Intermediary_Currency_3D	1
frequency_Beneficiary_Currency_3D	0.7
min_value_Currency_3D	0.7
sum_value_Intermediary_Beneficiary_3D	0.7

Nous pouvons observer que la caractéristique la plus influente est « value » avec un score d'importance de 6.7, indiquant l'importance du montant de la transaction pour détecter les transactions frauduleuses.

TABLE 4.7 – Description des caractéristiques

Caractéristique	**Description**
Value	Montant de la transaction.
avg_value_Beneficiary_Currency_1Y	Montant moyen des transactions du bénéficiaire dans la devise de la transaction au cours de l'année.
sum_value_Beneficiary_Intermediary Country_3D	Somme des montants des transactions du bénéficiaire vers le pays de l'intermédiaire au cours des trois derniers jours précédant la transaction.
sum_value_Beneficiary_Currency_1Y	Somme des montants des transactions du bénéficiaire dans la devise de la transaction au cours de l'année.
sum_value_Beneficiary_Currency_3D	Somme des montants des transactions du bénéficiaire dans la devise de la transaction au cours des trois derniers jours.
max_value_Beneficiary_Currency_1Y	Montant maximal réalisé par le bénéficiaire dans la devise de la transaction au cours de l'année.
sum_value_Intermediary_Currency_3D	Somme des montants des transactions de l'intermédiaire dans la devise de la transaction au cours des trois derniers jours.
frequency_Beneficiary_Currency_3D	Nombre de transactions réalisées par le bénéficiaire dans la devise de la transaction au cours des trois derniers jours.
min_value_Currency_3D	Montant minimal réalisé dans la devise de la transaction au cours des trois derniers jours.
sum_value_Intermediary_Beneficiary_3D	Somme des montants des transactions entre l'intermédiaire et le bénéficiaire au cours des trois derniers jours.
latency_BeneficiaryCountry_Currency_1Y	Temps en secondes passé dans le pays du bénéficiaire dans la devise de la transaction au cours de l'année.
count_intermediary_OriginatorCountry_ BeneficiaryCountry_1Y	Nombre de transactions avec un intermédiaire entre le pays de l'ordonnateur et le pays du bénéficiaire au cours de l'année.
frequency_Beneficiary_IntermediaryCountry _1Y	Nombre de transactions entre le bénéficiaire et le pays de l'intermédiaire au cours de l'année.
max_value_Intermediary_Beneficiary_3D	Montant maximal réalisé entre l'intermédiaire et le bénéficiaire au cours des trois derniers jours.

Les caractéristiques associées à la devise de la transaction, notamment au bénéficiaire, telles que « avg_value_Beneficiary_Currency_1Y » et « asum_value_Beneficiary_Currency_1Y »,

suggèrent l'importance de suivre le comportement des transactions d'un bénéficiaire au fil du temps.
De plus, les caractéristiques basées sur une période de trois jours, telles que « sum_value_Beneficiary_IntermediaryCountry_3D » et « sum_value_Beneficiary_Currency_3D », indiquent que les transactions récentes sont également déterminantes pour identifier des comportements frauduleux.
Enfin, l'importance relative des caractéristiques associées à l'intermédiaire, telles que « sum_value_Intermediary_Currency_3D », montre que le rôle et le comportement de l'intermédiaire dans le réseau de transactions sont également pertinents pour la détection de fraude.

4.3.2 SWIFT$_{base+syn}$

Les résultats de la sélection de caractéristiques par ordre décroissant d'importance pour le jeu de données SWIFT$_{base+syn}$ sont détaillés dans le tableau 4.9

TABLE 4.9 – Caractéristiques sélectionnées pour SWIFT$_{base+syn}$ avec leurs scores d'importance

Caractéristique	**Score d'Importance**
Value	6.9
avg_value_Beneficiary_Currency_1Y	2.9
sum_value_Beneficiary_Currency_1Y	2.3
latency_BeneficiaryCountry_Currency_1Y	1.9
sum_value_Beneficiary_Currency_3D	1.7
sum_value_Intermediary_Beneficiary_3D	1.4
count_intermediary_OriginatorCountry_BeneficiaryCountry_1Y	1.2
frequency_Beneficiary_IntermediaryCountry_1Y	1.2
max_value_Intermediary_Beneficiary_3D	1.1
max_value_Beneficiary_Currency_1Y	1.0

Nous pouvons encore observer que la caractéristique « Value » reste prédominante avec un score d'importance de 6.9, mettant en évidence le rôle central du montant de la transaction pour identifier des comportements frauduleux.
Les caractéristiques en rapport avec le bénéficiaire, telles que « avg_value_Beneficiary_Currency_1Y (2.9) » et « sum_value_Beneficiary_Currency_1Y (2.3) », confirment l'importance de suivre les transactions associées à un bénéficiaire spécifique sur une période d'un an.
La caractéristique « latency_BeneficiaryCountry_Currency_1Y (1.9) » introduit une dimension temporelle, montrant que la durée d'une transaction peut être un indicateur pertinent de fraude.
Les caractéristiques mettant en avant les interactions avec l'intermédiaire, ou son pays d'origine, telles que « sum_value_Intermediary_Beneficiary_3D (1.4) », « max_value_Intermediary_Beneficiary_3D (1.1) », « count_intermediary_OriginatorCountry_BeneficiaryCountry_1Y (1.2) » et « frequency_Beneficiary_IntermediaryCountry_1Y (1.2) », accentuent l'importance des intermédiaires dans les mécanismes de fraude observés sur ce jeu de données. Ceci s'explique par

le schéma de fraude synthétique ajouté à ce jeu de données, correspondant à un changement soudain d'intermédiaire.

Après avoir identifié et analysé les caractéristiques pertinentes pour distinguer les transactions frauduleuses au sein de notre base de données, nous avons obtenu un aperçu des points clés contribuant à la nature frauduleuse d'une transaction. Dans la suite, nous abordons l'étape du regroupement (clustering) des transactions frauduleuses. Cette démarche vise à regrouper les transactions similaires dans des clusters associés à des schémas de fraude.

4.4 Clustering des transactions frauduleuses

Dans cette section, nous utilisons le regroupement hiérarchique (CAH) pour former des clusters à partir des transactions frauduleuses. Suite à cette étape, ces clusters seront analysés et associés à des schémas spécifiques de fraude. Les résultats de ce clustering sont présentés sous la forme d'un dendrogramme, un outil graphique qui permet de visualiser les regroupements formés à chaque étape de l'agglomération.

Un dendrogramme est composé de branches verticales reliées par des branches horizontales, formant une structure d'arbre. Sur l'axe des abscisses, nous retrouvons les observations. L'axe des ordonnées, quant à lui, représente la distance à laquelle les clusters ont fusionné, reflétant le niveau de similarité entre eux. Dans notre cas, nous avons utilisé la distance euclidienne comme mesure de similarité. Ainsi, plus la fusion entre deux clusters se produit à une hauteur élevée sur le dendrogramme, plus ces clusters sont distincts l'un de l'autre. En observant et en interprétant ce dendrogramme, nous pourrons déterminer le nombre optimal de clusters et obtenir des informations sur les regroupements de transactions.

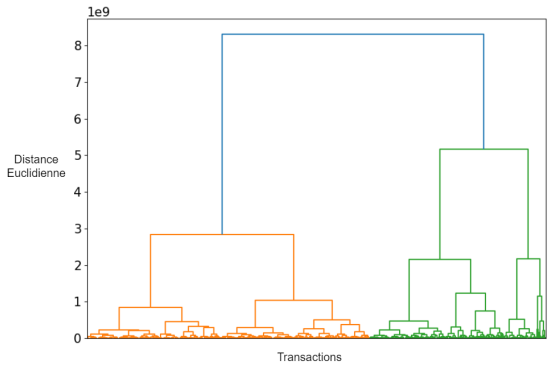

FIGURE 4.4 – Dendrogramme SWIFT$_{base}$

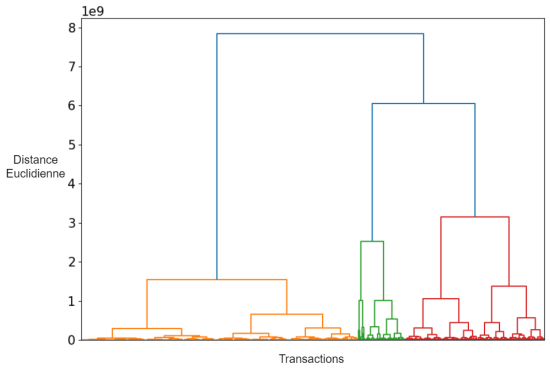

FIGURE 4.5 – Dendrogramme SWIFT$_{base+syn}$

Après avoir effectué le clustering hiérarchique sur nos deux jeux de données, nous nous sommes concentrés sur la représentation graphique de ces clusters à l'aide de dendrogrammes, comme le montrent les figures 4.4 et 4.5.

La figure 4.4 représente le dendrogramme pour le jeu de données SWIFT$_{base}$. Deux clusters distincts y sont clairement visibles, différenciés par les couleurs orange et verte. Ceci suggère

que, dans notre jeu de données initial, deux groupes majeurs de transactions peuvent être identifiés. Nous rappelons que dans notre analyse exploratoire, nous avions identifié deux schémas de fraude potentiels, qui sont les nouveaux acteurs réalisant des montants de transactions élevés et les transactions rapprochées dans le temps.

Cependant, la figure 4.5 montre le dendrogramme pour le jeu de données SWIFT$_{base+syn}$. Nous observons trois clusters clairement définis, symbolisés par les couleurs orange, vert et rouge. Ce troisième cluster pourrait être lié au schéma de fraude synthétique ajouté à ce jeu de données SWIFT$_{base+syn}$.

Jeu de données	Cluster	Nombre de transactions	Montant moyen
$SWIFT_{base}$	cluster 0	1 746	1 998 330
	cluster 1	11 621	466 870
$SWIFT_{base+syn}$	cluster 0	3 412	1 478 490
	cluster 1	11 801	318 990
	cluster 2	154	2 634 140

TABLE 4.10 – Tableau récapitulatif du clustering

Le tableau 4.10 présente un récapitulatif des clusters formés pour nos deux jeux de données, à savoir $SWIFT_{base}$ et $SWIFT_{base+syn}$ avec leur nombre de transactions et le montant moyen de ces transactions. L'examen de ces clusters révèle des nuances quant à la nature des transactions dans nos jeux de données.

Dans le jeu de données $SWIFT_{base}$, deux clusters distincts émergent. Le cluster 0, bien qu'il ne compte que $1\,746$ transactions, présente un montant moyen élevé de $1\,998\,330$. En contraste, le cluster 1, qui est nettement plus volumineux avec $11\,621$ transactions, a un montant moyen plus faible de $466\,870$.

Dans le jeu de données $SWIFT_{base+syn}$, le cluster 0 et le cluster 1 montrent des comportements similaires à ceux observés dans $SWIFT_{base}$, avec des montants moyens de $1\,478\,490$ et $318\,990$ respectivement. Cependant, le cluster 2, malgré son faible volume de 154 transactions, présente un montant moyen exceptionnellement élevé de $2\,634\,140$. Cette valeur est particulièrement marquante et suggère une nature de transaction spécifique, probablement liée au schéma de fraude avec les nouveaux comptes réalisant une seule transaction.

Dans la section suivante, nous nous concentrons sur l'entraînement d'un modèle de classification. Son rôle est essentiellement double. Premièrement, le modèle doit être en mesure de distinguer nettement entre les transactions frauduleuses et les transactions légitimes. Deuxièmement, pour les transactions identifiées comme frauduleuses, le modèle devra également déterminer à quel cluster elles appartiennent. Cela nous permet non seulement d'identifier la fraude, mais également de comprendre sa nature spécifique, en se basant sur les caractéristiques propres à chaque cluster.

4.5 Classification des transactions suspectes

Dans cette section, nous présentons l'entraînement de nos modèles de classification multiclasses. Ces modèles s'appuient sur des algorithmes ensemblistes fondés sur des arbres de

décision. Les jeux de données $SWIFT_{base}$ et $SWIFT_{base+syn}$ ont été utilisés et subdivisés en deux ensembles : un ensemble d'entraînement, représentant 80% des transactions, et un ensemble de test, constituant les 20% restants.

Dans le cadre de nos expérimentations, nous avons procédé à l'entraînement de quatre modèles distincts en utilisant des algorithmes renommés pour leur efficacité et leur performance en classification : Random Forest, CatBoost, XGBoost et LightGBM. Ces algorithmes, fondés sur des techniques d'apprentissage ensemblistes basées sur des arbres de décision, sont reconnus dans la littérature pour leur capacité à traiter les données tabulaires, comme c'est le cas dans la détection de fraude financière. Dans la suite, nous comparons les performances de chaque algorithme.

4.5.1 Comparaison des algorithmes de classification

Les modèles que nous entraînons avec les 4 algorithmes mentionnés ci-dessus, sont entraînés de telle sorte à discriminer entre les classes frauduleuses associées aux clusters identifiés lors de l'étape précédente et la classe représentant les transactions légitimes. Avant de présenter les résultats de comparaison, nous allons rappeler les mesures avec lesquelles ils seront comparés. Tout d'abord, les 4 modèles réalisent des prédictions entre les classes des clusters frauduleux et légitime. Comme décrit dans le chapitre 3, nous utilisons les prédictions pour calculer un score de suspicion en additionnant les probabilités associées aux classes frauduleuses, puis nous multiplions le résultat par 100. Nous souhaitons exprimer la probabilité en pourcentage afin de faciliter sa lecture. Nous utilisons l'algorithme 1 défini dans la section 3.5.3 pour déterminer un seuil optimal, les transactions dont le score de suspicion est supérieur à ce seuil seront frauduleuses et celles dont le score est inférieur seront légitimes. Cet algorithme vise à trouver le seuil pour maximiser f1-score du modèle, tout en minimisant le coût financier associé aux prédictions du modèle. En s'inspirant des travaux de [BSAO13], le coût associé à la détection de transactions frauduleuses est défini à partir d'un coût administratif (Ca) fixé à 5. Il est à noter que nous ne spécifions pas de devise précise, car tous les montants des transactions ont été uniformisés dans une devise qui demeure inconnue à nos travaux. Par ailleurs, dans notre algorithme, un paramètre epsilon est introduit. Il représente l'écart toléré par rapport à la valeur maximale que peut atteindre le f1-score du modèle. Après consultation d'experts, cette tolérance a été fixée à 0.01, permettant ainsi de rechercher le coût minimal dans cet intervalle autour du f1-score optimal. Toutes transactions avec un score de suspicion supérieur à ce seuil sera considérée frauduleuse par le modèle.

Les résultats obtenus à partir des différents algorithmes utilisés pour nos expériences sont détaillés dans les tableaux suivants :

- Tableau 4.11 pour l'algorithme Random Forest
- Tableau 4.12 pour l'algorithme CatBoost
- Tableau 4.13 pour l'algorithme XGBoost
- Tableau 4.14 pour l'algorithme LightGBM

Chaque tableau présente les mesures d'évaluation (précision, rappel et f1-score) obtenues pour les classes « légitime » et « frauduleuse », ainsi que la moyenne de ces métriques pour les deux classes. De plus, le coût associé à la détection de fraude et le seuil optimal pour chaque modèle sont également présentés. Les résultats sont présentés dans les tableaux 4.11, 4.12, 4.13 et 4.14 :

Jeu de données	Classe	Précision	Rappel	f1-score	Coût	Seuil
$SWIFT_{base}$	légitime	0.99	0.99	0.99	852 559 251	25
	frauduleuse	0.80	0.72	0.76		
	moyenne des deux	0.90	0.86	0.88		
$SWIFT_{base+syn}$	légitime	0.99	0.99	0.99	746 300 323	25
	frauduleuse	0.83	0.80	0.81		
	moyenne des deux	0.91	0.90	0.90		

TABLE 4.11 – Résultats pour l'algorithme Random Forest

Jeu de données	Classe	Précision	Rappel	f1-score	Coût	Seuil
$SWIFT_{base}$	légitime	0.99	0.99	0.99	974 459 380	20
	frauduleuse	0.70	0.60	0.65		
	moyenne des deux	0.85	0.80	0.82		
$SWIFT_{base+syn}$	légitime	0.99	0.99	0.99	735 164 827	20
	frauduleuse	0.75	0.76	0.76		
	moyenne des deux	0.88	0.87	0.88		

TABLE 4.12 – Résultats pour l'algorithme CatBoost

Jeu de données	Classe	Précision	Rappel	f1-score	Coût	Seuil
$SWIFT_{base}$	légitime	0.99	0.99	0.99	824 443 056	20
	frauduleuse	0.73	0.68	0.70		
	moyenne des deux	0.86	0.84	0.85		
$SWIFT_{base+syn}$	légitime	0.99	0.99	0.99	717 186 134	20
	frauduleuse	0.79	0.79	0.79		
	moyenne des deux	0.89	0.89	0.89		

TABLE 4.13 – Résultats pour l'algorithme XGBoost

Jeu de données	Classe	Précision	Rappel	f1-score	Coût	Seuil
$SWIFT_{base}$	légitime	0.99	0.99	0.99	1 069 569 596	15
	frauduleuse	0.48	0.54	0.51		
	moyenne des deux	0.73	0.77	0.75		
$SWIFT_{base+syn}$	légitime	0.99	0.99	0.99	1 010 110 831	10
	frauduleuse	0.55	0.69	0.59		
	moyenne des deux	0.77	0.82	0.79		

TABLE 4.14 – Résultats pour l'algorithme LightHBM

En analysant les performances des modèles, l'algorithme Random Forest présente un f1-score élevé de 0.90 avec un seuil optimal de 25 sur le jeu de données $SWIFT_{base+syn}$. XGBoost possède une valeur très proche avec un f1-score de 0.89 et un seuil de 20. CatBoost a également une bonne performance avec un f1-score de 0.88 et un seuil de 20 sur le même jeu de données. En revanche, LightGBM a affiché les performances les plus basses avec un f1-score de seulement 0.79 et un seuil de 10.

Lorsqu'on examine les coûts associés à chaque modèle, XGBoost s'avère être le plus économique avec un coût de 717186134, tandis que LightGBM est le plus coûteux avec 1010110831 sur le jeu de données $SWIFT_{base+syn}$.

En résumé, même si l'algorithme Random Forest offre le meilleur f1-score, rendant son utilisation idéale pour la détection de fraude dans notre contexte, XGBoost semble être le choix le plus judicieux en prenant en compte à la fois la performance et le coût, avec un seuil de 20.

Dans la section suivante, nous explorons l'analyse des valeurs Shapley pour approfondir notre compréhension de l'importance des caractéristiques dans la classification des transactions. Ceci nous permet d'identifier les schémas de fraude associés aux différents clusters.

4.5.2 Identification des schémas de fraude

Après avoir finalisé l'étape de classification, nous analysons les clusters identifiés lors de l'étape de clustering pour les associer à des schémas de fraude. Pour ce faire, nous utilisons des techniques d'interprétabilité. Notre modèle attribue à chaque transaction un label associé à un cluster. Notre prochaine étape consiste à exploiter le modèle de classification avec le framework SHAP (SHapley Additive exPlanations). Pour rappel, SHAP est un outil puissant qui explique la prédiction d'un modèle en attribuant à chaque caractéristique une importance relative à la prédiction en utilisant les valeurs Shapley. Nous nous concentrons sur les prédictions correctes du modèle effectuées sur le jeu de données de test. Nous précisons que nous n'utilisons pas les données d'entraînement pour cette analyse, car le modèle les connaît déjà et pourrait donc introduire un biais dans notre interprétation. Dans le tableau 4.15, nous présentons le nombre de transactions et la moyenne de leur montant pour chaque cluster des deux jeux de données.

Jeu de données	Cluster	Nombre de transactions	Montant moyen
$SWIFT_{base}$	cluster 0	87	1 387 443
	cluster 1	1 208	376 653
$SWIFT_{base+syn}$	cluster 0	298	1 101 618
	cluster 1	1714	278 188
	cluster 2	11	1 837 318

TABLE 4.15 – Transactions des prédictions correctes sur le jeu de données test

Dans ce qui suit, nous allons explorer en détail chaque cluster issu des ensembles de données $SWIFT_{base}$ et $SWIFT_{base+syn}$. Pour cette exploration, nous avons recours à une visualisation spécifique : le Beeswarm, tiré du framework SHAP. Le Beeswarm de SHAP est une représentation graphique qui illustre clairement l'importance des caractéristiques d'un modèle et l'effet de celles-ci sur ses prédictions.

Avant de poursuivre l'analyse des clusters, il est essentiel de rappeler les schémas de fraude spécifiques présents dans nos jeux de données. Ces schémas, identifiés lors de nos analyses exploratoires, guident notre compréhension des tendances et des comportements frauduleux potentiellement observés.

Lors de la phase d'analyse exploratoire du jeu de données $SWIFT_{base}$ dans la section 4.2, nous avons identifié deux schémas de fraude dominants :

— **Acteurs occasionnels et montants de transactions élevés** : Ce schéma est caractérisé par l'émergence de nouveaux acteurs sur le réseau réalisant une seule transaction avec un montant anormalement élevé.

— **Transactions rapprochées dans le temps** : Dans ce schéma nous observons des transactions qui sont exécutées à des intervalles de temps très rapprochés avec des montants normaux, suggérant une tentative potentielle de diviser un grand montant en plusieurs petites transactions afin d'échapper à la détection.

Nous rappelons que nous avons ajouté synthétiquement un schéma de fraude dans notre jeu de données $SWIFT_{base+syn}$:

— **Changement d'intermédiaire** : Ce schéma signale un changement soudain dans l'intermédiaire utilisé pour réaliser une transaction.

Ces schémas de fraude fournissent le contexte nécessaire pour interpréter les tendances et les caractéristiques saillantes que nous observerons dans les clusters.

4.5.2.1 SWIFT base

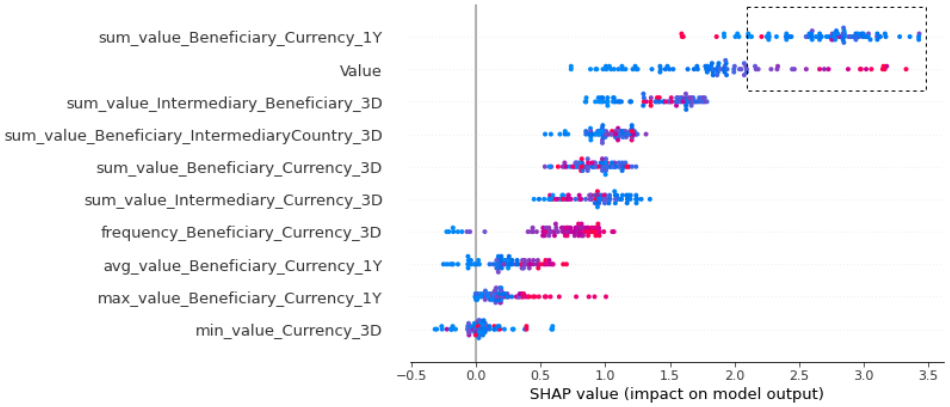

FIGURE 4.6 – Beeswarm du cluster 0 du jeu de données $SWIFT_{base}$

Cluster 0 En examinant le Beeswarm pour le cluster 0 du jeu de données $SWIFT_{base}$ (figure 4.6), deux caractéristiques se démarquent avec une forte importance :

— sum_value_Beneficiary_Currency_1Y possède des valeurs faibles indiquées pas la couleur bleue des correspondants.

— Value possède des valeurs élevées indiquées par la couleur rouge de ses points.

La faible valeur de « sum_value_Beneficiary_Currency_1Y » suggère que le total des transactions effectuées par cet acteur sur une période d'un an est faible. Cela pourrait indiquer que l'acteur a réalisé peu de transactions, voire une seule, durant cette période. D'autre part, les valeurs élevées associées à « Value » soulignent que, même s'il y a eu peu de transactions, leurs montants étaient significatifs.

En mettant ces observations en relation avec les schémas de fraude que nous avons identifiés précédemment, le cluster 0 semble correspondre au schéma **nouveaux acteurs et montants de transactions élevés**. Un nouvel acteur réalisant une ou quelques transactions avec des montants notables s'aligne sur ce schéma.

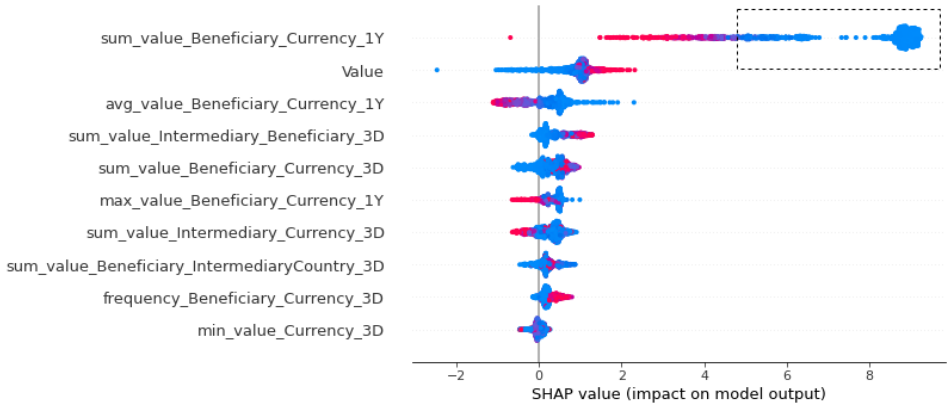

FIGURE 4.7 – Beeswarm du cluster 1 du jeu de données $SWIFT_{base}$

Cluster 1 En étudiant le Beeswarm pour le cluster 1 du jeu de données $SWIFT_{base}$ (figure 4.7), une caractéristique se démarque avec une forte importance :

— « sum_value_Beneficiary_Currency_1Y » qui possède des valeurs faibles.

La faible valeur de « sum_value_Beneficiary_Currency_1Y » suggère que le total des transactions réalisées par cet acteur sur une période d'un an est bas. Cependant, cette seule observation ne fournit pas une image complète des activités ou des comportements de l'acteur.

Ainsi, sur la base de cette information isolée, il est difficile de rattacher clairement ce cluster à l'un des schémas de fraude que nous avons identifiés précédemment. Nous ne disposons pas d'assez d'éléments probants pour faire une telle association.

4.5.2.2 SWIFT base+syn

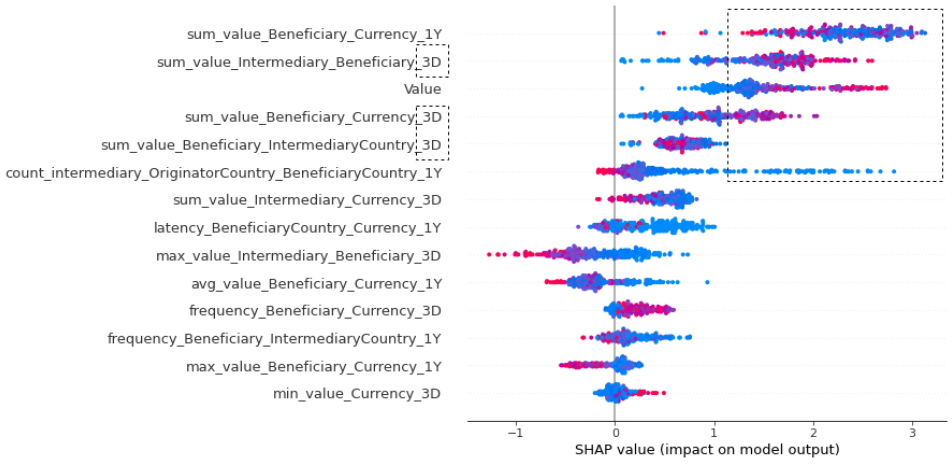

FIGURE 4.8 – Beeswarm du cluster 0 du jeu de données $SWIFT_{base+syn}$

Cluster 0 En étudiant le Beeswarm pour le cluster 0 du jeu de données $SWIFT_{base+syn}$ (figure 4.8), plusieurs caractéristiques ont une forte importance :

— sum_value_Beneficiary_Currency_1Y : Cette caractéristique possède des valeurs faibles. Cela suggère que le bénéficiaire a réalisé un nombre limité de transactions sur une période d'une année ou des transactions de faible valeur.

— sum_value_Intermediary_Beneficiary_3D : Ici, nous observons des valeurs significativement élevées, indiquant que le volume des transactions entre un intermédiaire et un bénéficiaire est considérable sur une courte période de 3 jours.
— sum_value_Beneficiary_Currency_3D : Bien que l'impact de cette caractéristique soit légèrement plus faible que le précédent, ses valeurs élevées sur une période de 3 jours sont notables. Cela nous indique que le montant des transactions pour une devise spécifique a été particulièrement élevé dans ce court laps de temps.
— count_intermediary_OriginatorCountry_BeneficiaryCountry_1Y : Des valeurs faibles pour cette caractéristique suggèrent que le nombre de transactions entre un pays d'origine et un pays bénéficiaire via un intermédiaire spécifique a été limité sur l'année écoulée.

En synthétisant ces observations, le motif apparent du cluster se rapproche du schéma de fraude **transactions rapprochées dans le temps**. La combinaison d'un volume de transactions élevé sur une période très courte (3 jours) et la faible activité sur une période plus longue (1 an) suggère une tentative d'exécuter plusieurs transactions en peu de temps, potentiellement pour échapper à la détection.

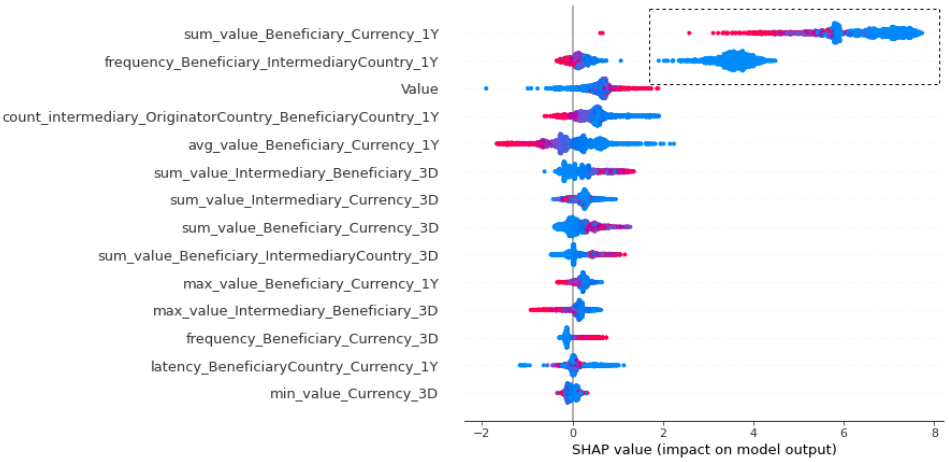

FIGURE 4.9 – Beeswarm du cluster 1 du jeu de données $SWIFT_{base+syn}$

Cluster 1 En étudiant le Beeswarm pour le cluster 1 du jeu de données $SWIFT_{base+syn}$ (figure 4.9), deux caractéristiques ont une forte importance :

— sum_value_Beneficiary_Currency_1Y : Cette caractéristique possède des valeurs faibles, suggérant que le montant total des transactions liées à une monnaie spécifique pour un bénéficiaire donné a été relativement bas sur une période d'un an.
— frequency_Beneficiary_IntermediaryCountry_1Y : Avec des valeurs faibles, cette caractéristique indique que la fréquence des transactions entre un bénéficiaire et les intermédiaires d'un pays particulier a été limitée durant l'année écoulée.

La combinaison de ces deux observations nous donne des indices sur le comportement transactionnel. Un bénéficiaire ayant une faible somme de transactions pour une monnaie donnée couplé à une fréquence faible de transactions avec les intermédiaires d'un pays spécifique peut suggérer que ce bénéficiaire a changé d'intermédiaire ou a modifié ses habitudes transactionnelles.
Ces tendances concordent avec le schéma de fraude **changement d'intermédiaire soudain**.

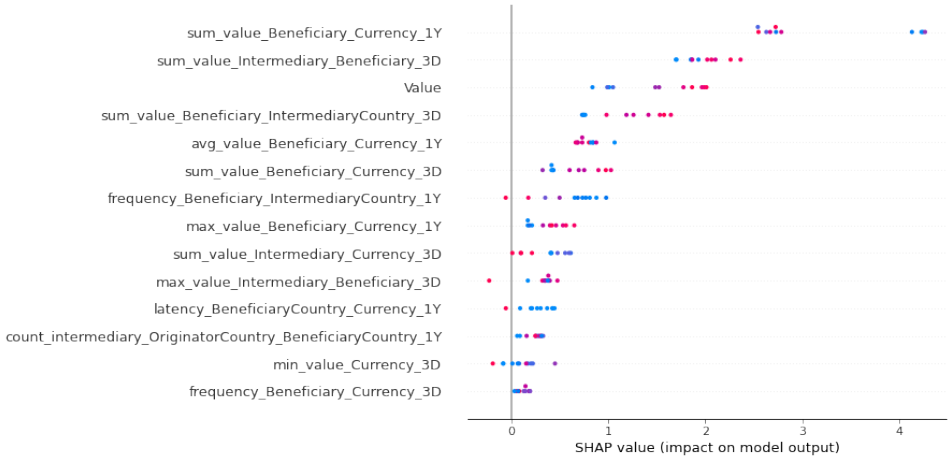

FIGURE 4.10 – Beeswarm du cluster 2 du jeu de données $SWIFT_{base+syn}$

Cluster 2 Le cluster 2 de $SWIFT_{base+syn}$ se caractérise par un faible nombre de transactions, précisément 11 transactions, comme le montrent les résultats illustrés dans la figure 4.10 qui peuvent être difficiles à interpréter de manière détaillée. Cependant, en se référant au tableau 4.15, nous remarquons que ces transactions possèdent une moyenne de montant élevée atteignant à 1 837 318. Néanmoins, un élément clé ressort : les transactions de ce cluster présentent des montants exceptionnellement élevés. Cette observation nous conduit à conclure que ce cluster pourrait être associé au schéma de fraude caractérisé par des **montants élevés**.

TABLE 4.16 – Récapitulatif des schémas de fraude

Clusters	**Caractéristiques importantes**	**Schéma de fraude**
Cluster 0	Bénéficiaire avec faible activité annuelle mais montants de transactions élevés.	Nouveaux acteurs et montants de transactions élevés
Cluster 1	Bénéficiaire avec faible somme de transactions annuelles.	Indéterminé
Cluster 0 (base+syn)	Bénéficiaire actif sur 3 jours avec montants élevés, mais faible activité annuelle.	Transactions rapprochées dans le temps
Cluster 1 (base+syn)	Bénéficiaire avec faible activité annuelle et peu de transactions avec ses intermédiaires.	Changement d'intermédiaire soudain
Cluster 2 (base+syn)	Transactions rares mais avec des montants significativement élevés.	Montants élevés

4.5.3 Synthèse : Classification et identification des schémas de fraude

Au cours de notre étude, nous avons évalué quatre algorithmes majeurs pour la classification des transactions : Random Forest, XGBoost, CatBoost et LightGBM. Suite à une analyse comparative de leurs performances, XGBoost s'est démarqué comme le choix optimal en tenant compte à la fois du f1-score et des coûts associés. Il est à noter que ce modèle étiquette une transaction comme frauduleuse si elle présente un score de suspicion supérieur à 20. Ce seuil a été sélectionné grâce à notre algorithme présenté dans le chapitre 3 permettant de trouver le seuil qui maximise le f1-score tout en minimisant le coût des prédictions de notre modèle.

Après avoir défini notre modèle de référence, nous l'avons utilisé comme outil principal pour identifier les clusters associés aux transactions frauduleuses au sein de nos jeux de données. Nous avons utilisé une technique d'interprétabilité des modèles avec le framework SHAP, qui, à travers l'utilisation des valeurs de Shapley, nous a permis de déterminer les caractéristiques majeures influençant la classification.

Grâce à cette approche, nous avons réussi à identifier et à associer des schémas de fraude spécifiques à la majorité des clusters, un récapitulatif des schémas de fraude identifiés est présentée dans le tableau 4.16. Toutefois, il convient de mentionner que le cluster 1 du jeu de données $SWIFT_{base}$ est resté sans association claire à un schéma de fraude spécifique.

La prochaine étape s'articule autour de l'utilisation de l'ontologie pour une analyse approfondie des transactions présentant un score de suspicion supérieur au seuil sélectionné à 20. Il est essentiel de souligner que ces transactions, bien qu'ayant un score élevé de suspicion, recevront le statut de « suspecte » et non de « frauduleuse ». En effet, une transaction ne peut être définitivement étiquetée comme frauduleuse qu'après une vérification rigoureuse et un contrôle par un expert du domaine. Ainsi, notre objectif est donc de fournir à ces experts des outils et des informations pertinentes pour faciliter leur travail d'inspection sur les transactions suspectes détectées par notre approche.

4.6 L'ontologie pour l'aide à la décision

Nous avons présenté dans le chapitre 3, l'ontologie développée dans le cadre de ce travail. Notre ontologie a pour but de de fournir un cadre structuré et sémantique pour assister l'analyse des transactions et aider les experts à prendre des décisions plus facilement sur la validation ou le blocage des transactions. Cette ontologie se concentre principalement sur les transactions et les différents acteurs y participant.

L'expérimentation de cette partie a été mise en difficulté par le caractère anonyme de notre jeu de données. En effet, sans informations concrètes sur les banques ou les clients, il est complexe d'exploiter pleinement l'ontologie et de déduire des informations pertinentes ou des relations sémantiques significatives. Cette absence d'informations spécifiques limite notre capacité à contextualiser et à comprendre les transactions suspectes de nos jeux de données.

La figure 4.11 présente de manière concise le processus intégral de notre approche pour le contrôle des transactions suspectes. Initialement, nous réalisons un calcul et une sélection des caractéristiques de la transaction. Sur cette base, notre modèle d'apprentissage automatique

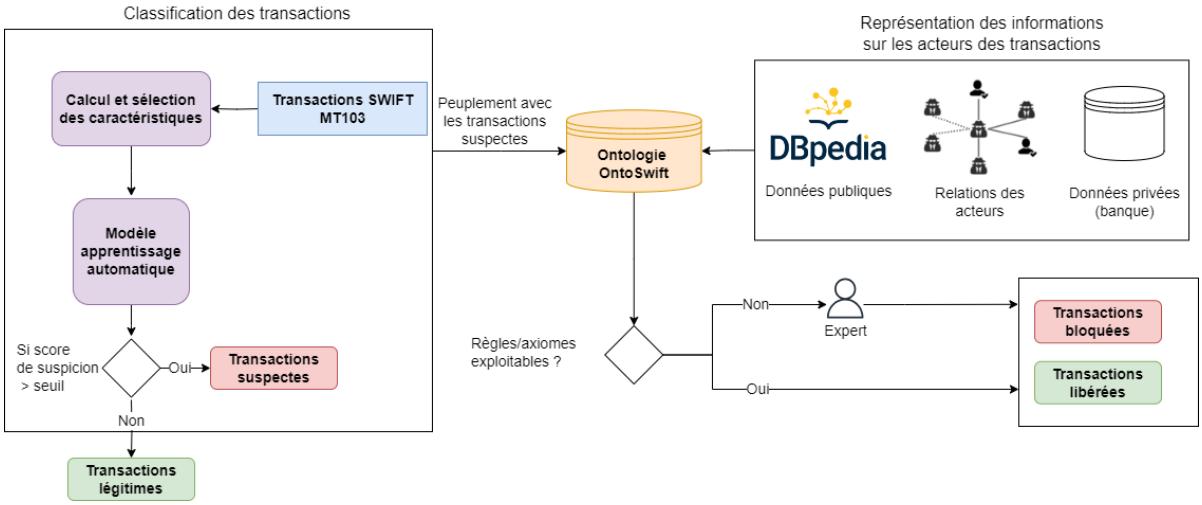

FIGURE 4.11 – Processus de contrôle des transactions

est utilisé pour déduire un score de suspicion. Une transaction est étiquetée comme suspecte si son score dépasse un seuil préétabli, autrement, elle est considérée comme légitime. L'ontologie est peuplée avec les transactions, leurs attributs, leurs caractéristiques et leur score de suspicion associés. Ensuite, nous exploitons les connaissances dans notre ontologie pour utiliser les règles et les axiomes définis. L'objectif principal consiste à prendre une décision concernant ces transactions suspectes : à savoir, les libérer ou les bloquer. Néanmoins, il existe des situations où, face à l'absence ou l'insuffisance de données, les règles et axiomes ne sont pas exploitables pour parvenir à une décision automatique. Dans ce cas, l'expert doit décider de lui-même avec les informations contenues dans l'ontologie pour libérer ou bloquer la transaction.

Dans la section suivante, nos expérimentations se concentrent sur un échantillon de 20 transactions suspectes impliquant 32 acteurs bancaires. Ces données peuvent être interrogées grâce au langage SPARQL. La figure 4.12 présente les règles définies dans notre ontologie en SWRL.

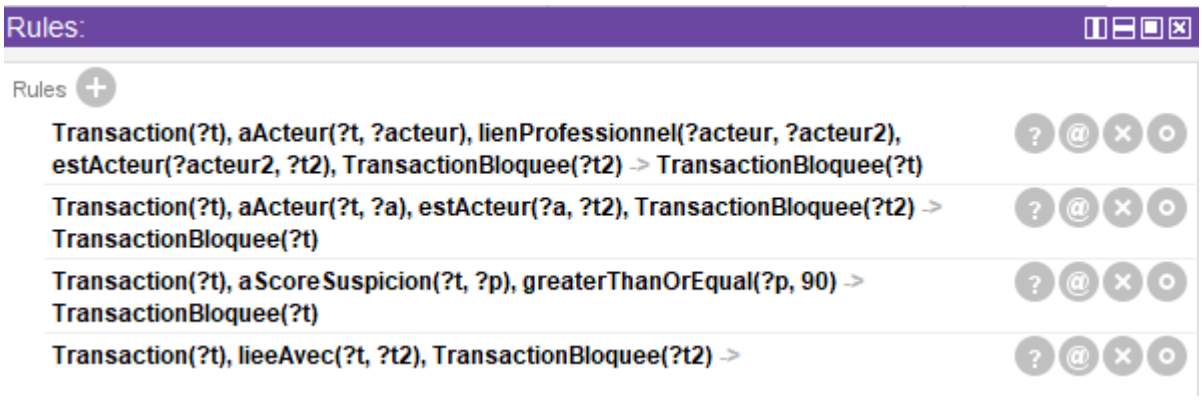

FIGURE 4.12 – Règles définies dans l'ontologie

Nous avons exécuté ces règles, ce qui a entraîné le blocage et la libération de certaines transactions. Par exemple, la première règle bloque les transactions avec un score de suspicion supérieur à 90, ce qui a entraîné le blocage de 6 transactions affichées dans la figure 4.13.

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX onto: <http://test.org/ontoSwift.owl#>

SELECT
( STR(?montant) as ?Montant )
( STR(?devise) as ?Devise )
( STR(?date) as ?Date )
( STR(?score) as ?ScoreSuspicion )

WHERE {
?transaction rdf:type onto:TransactionBloquee .
?transaction onto:aDevise ?devise .
?transaction onto:aDate ?date .
?transaction onto:aValue ?montant .
?transaction onto:aScoreSuspicion ?score .
}
```

Montant	Devise	Date	ScoreSuspicion
"248264"	"3F5"	"01/01/2019, 09:40:00"	"100"
"134348"	"A42"	"01/01/2019, 11:20:00"	"100"
"878634"	"A4B"	"01/01/2019, 09:46:00"	"100"
"381717"	"F30"	"01/01/2019, 08:34:00"	"100"
"249509"	"75D"	"01/01/2019, 10:25:00"	"100"
"76331"	"1E6"	"01/01/2019, 10:13:00"	"100"

FIGURE 4.13 – Transactions bloquées à la suite de la première règle SWRL

Dans la figure 4.14, nous rappelons les deux axiomes de notre ontologie **ActeurImpliqueBloquee** et **ActeurLiberee**. Suite à l'application de ces axiomes, le raisonneur a inféré les instances de ces nouveaux concepts pour les acteurs.

FIGURE 4.14 – Axiomes définis dans l'ontologie

Dans la figure 4.15, nous affichons à l'aide d'une requête, les acteurs impliqués dans une transaction bloquée avec leur nombre de transactions bloquées.
Notre ontologie offre la possibilité d'intégrer un large éventail de règles en exploitant des données publiques provenant de sources telles que dbpedia [1] ou des réseaux sociaux. L'objectif est d'enrichir les règles avec des informations plus spécifiques sur les acteurs, afin de faciliter la prise de décision, que ce soit pour bloquer ou libérer les transactions ou pour faciliter l'analyse de l'expert.

1. https://www.dbpedia.org/

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX onto: <http://test.org/ontoSwift.owl#>

SELECT
 (STR(?bic) as ?BIC)
 (STR("Impliquee Bloquee") as ?Statut_Acteur)
 (STR(COUNT(?transaction)) as ?Nombre_Transactions_Bloquees)

WHERE {
   ?acteur rdf:type onto:ActeurImpliqueBloquee .
   ?acteur onto:aBIC ?bic .
   ?transaction onto:aActeur ?acteur .
   ?transaction rdf:type onto:TransactionBloquee .
}
GROUP BY ?bic
```

BIC	Statut_Acteur	Nombre_Transactions_Bloquees
"RJSGLKXX"	"Impliquee Bloquee"	"1"
"JBDQDEXX"	"Impliquee Bloquee"	"1"
"RMZCMYXX"	"Impliquee Bloquee"	"1"
"JOBXNGXX"	"Impliquee Bloquee"	"1"
"XOAWPYXX"	"Impliquee Bloquee"	"1"
"CBVPKWXX"	"Impliquee Bloquee"	"1"
"AHBKNZXX"	"Impliquee Bloquee"	"1"

FIGURE 4.15 – Acteurs impliqués dans des transactions bloquées

Les décisions de blocage prises par notre ontologie reposent sur des règles explicites. Cela s'avère particulièrement utile pour les experts, rendant le contrôle des transactions suspectes plus intuitif et efficace. Cette ontologie se positionne comme un complément solide à notre modèle initialement construit sur des techniques d'apprentissage automatique. Dans l'optique d'optimiser davantage la tâche des experts, nous avons développé un outil reprenant les travaux de notre méthode. Son objectif est d'assister les experts en proposant 3 modules. Le premier permet l'entraînement d'un modèle de classification à partir d'un jeu de données, le deuxième sert à réaliser un clustering et à identifier les schémas de fraude. Et le dernier se base sur l'ontologie pour analyser les transactions suspectes.

4.7 L'outil ST-Fraud

Dans cette section, nous présentons l'outil ST-Fraud développé dans le cadre de cette thèse pour détecter les transactions suspectes et faciliter leur examen par les experts en charge de la validation des transactions bloquées par les systèmes anti-fraudes. ST-Fraud est conçu pour la détection et l'interprétation de transactions suspectes, et se compose de trois modules principaux :

- Module 1 (entraînement du modèle) : Import des données, calcul des caractéristiques et entraînement des modèles.
- Module 2 (identification des schémas de fraude) : Clustering des transactions frauduleuses, interprétation et association des clusters avec des schémas de fraude.
- Module 3 (analyse des transactions suspectes) : Présentation des connaissances sur la transaction et les acteurs pour assister l'expert dans son analyse.

Nous allons maintenant présenter chacun de ces modules.

4.7.1 Module d'entraînement du modèle

Le premier module de ST-Fraud permet l'entraînement d'un modèle de détection de transactions suspectes. L'interface de ce module, illustrée dans la figure 4.16, guide l'utilisateur à travers les différentes étapes, de l'importation des données brutes à la finalisation de l'entraînement du modèle.

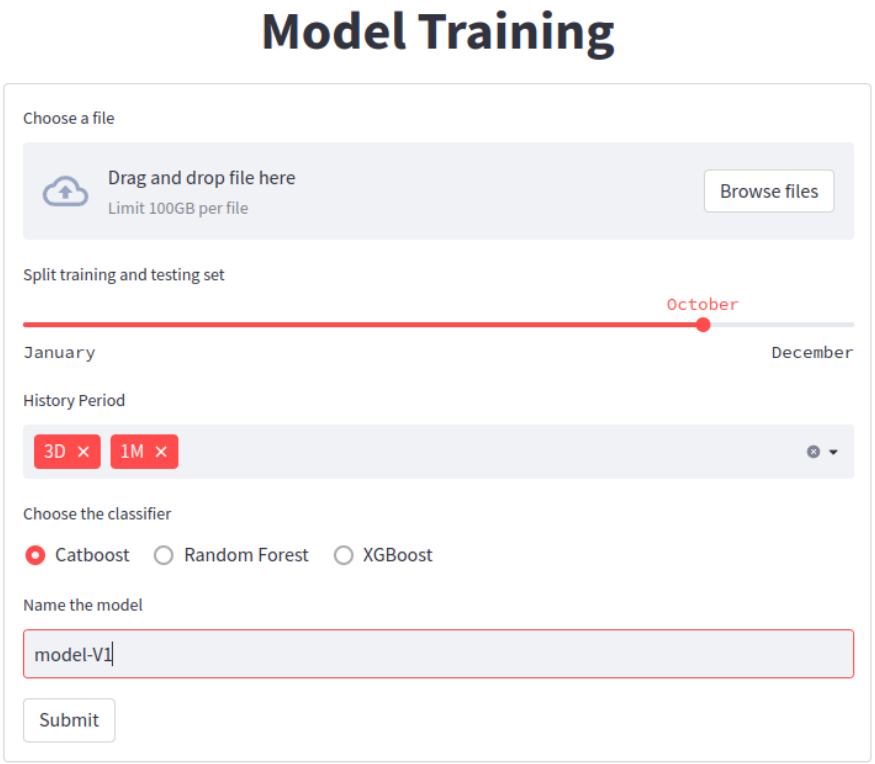

FIGURE 4.16 – ST-Fraud : module d'entraînement d'un modèle

Les principales fonctionnalités de ce module comprennent :

- **Importation des données (*Choose a file*) :** Permet d'intégrer un fichier CSV contenant les transactions. L'outil a été conçu pour importer des données CSV, mais il pourrait aussi intégrer des données de base de données si les banques le souhaitent.
- **Séparation des données (*Split training and testing set*) :** Répartition automatique des données en ensembles d'entraînement et de test pour garantir une évaluation rigoureuse du modèle.
- **Sélection des périodes temporelles (*History Period*) :** Offre la flexibilité de définir des intervalles temporels spécifiques pour le calcul des caractéristiques, offrant ainsi une perspective multi-échelle sur les potentiels schémas de fraude.
- **Choix de l'algorithme de classification (*Choose the classifier*) :** L'utilisateur a la possibilité de sélectionner les algorithmes de classification CatBoost, Random Forest ou XGBoost, en fonction de son volume de données, de la nature de ses attributs et du nombre de dimensions.
- **Lancement de l'entraînement :** Avec tous les paramètres en place, l'utilisateur est prêt à initier l'entraînement. Le module gère alors l'intégralité du processus, en s'appuyant sur les données et caractéristiques fournies.

Ce module offre une solution complète pour la préparation, la configuration et l'entraînement d'un modèle de détection de transactions suspectes.

4.7.2 Identification des schémas de fraude

Le deuxième module de ST-Fraud est dédié à l'exploration et l'identification des schémas de fraude. Il offre aux experts des outils conçus pour analyser ces schémas. Ce module comporte deux principales fonctionnalités : la première permet de sélectionner le nombre de caractéristiques pour réaliser un clustering des transactions frauduleuses. À l'aide d'un dendrogramme, l'expert peut choisir le nombre optimal de clusters pour réaliser ce clustering. La deuxième fonctionnalité permet d'associer des schémas de fraude aux clusters, en utilisant des outils visuels et des informations sur les clusters.

Paramétrage pour l'identification Cette première fonctionnalité (illustrée à la figure 4.17) permet à l'expert à configurer les paramètres essentiels pour l'identification des schémas de fraude. Au départ, l'expert sélectionne le modèle (*Select model*) à utiliser, qui associe les données, l'algorithme et les caractéristiques prédéfinies, optimisé pour une réduction dimensionnelle.

L'expert peut ensuite définir le nombre de caractéristiques (*Features number*) sur lesquelles il souhaite réaliser son analyse. Un nombre trop élevé peut améliorer les résultats, mais cela peut rendre complexe l'analyse des relations entre ces caractéristiques. En revanche, un nombre trop faible peut compromettre l'efficacité du modèle, car les schémas de fraude analysés ne seront pas représentatifs de ceux présents dans le jeu de données.

Le nombre de schémas de fraude (*Clusters number*) à identifier est également paramétrable, guidé par un dendrogramme. Le dendrogramme est un outil visuel utilisé en clustering pour représenter la structure hiérarchique des données. À travers ce diagramme, il est possible d'observer et d'interpréter les regroupements de transactions et leurs niveaux de similarité.

FIGURE 4.17 – ST-Fraud : processus de réduction dimensionnelle et de clustering

Analyse détaillée des clusters La seconde fonctionnalité dont l'interface est décrite dans la figure 4.18, met l'accent sur l'étude des clusters formés. Après avoir défini les paramètres, l'expert peut se focaliser sur l'analyse des regroupements, dans le but d'associer un schéma de fraude à chaque cluster.

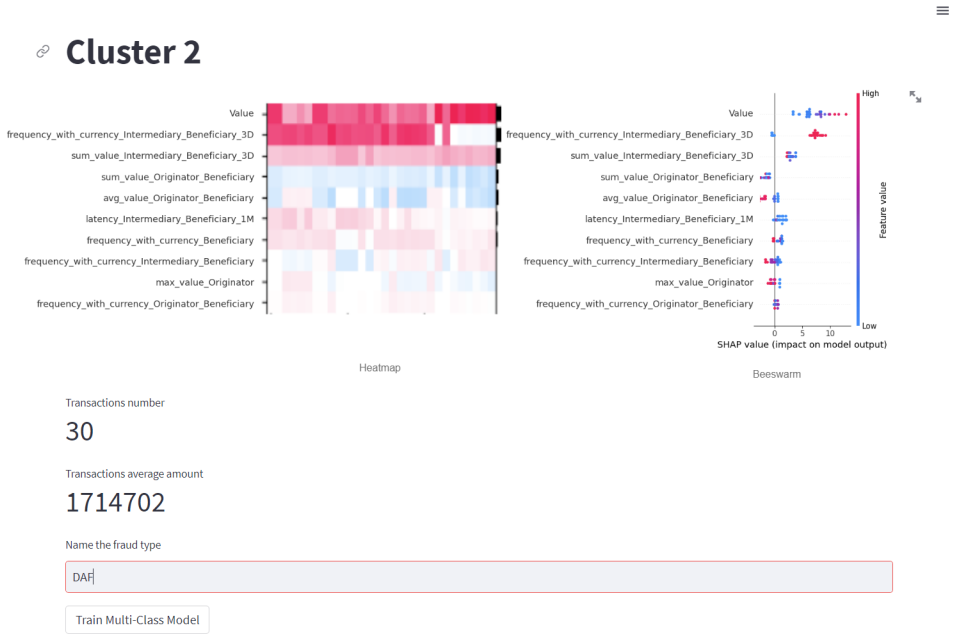

FIGURE 4.18 – ST-Fraud : Démarche d'analyse et d'identification

Nous utilisons deux outils de visualisation principaux du framework SHAP : "beeswarm" et "heatmap". Ces outils offrent à l'expert une vision de synthèse sur l'influence des différentes

caractéristiques sur la classification des transactions. Le "beeswarm" présente visuellement comment chaque caractéristique contribue à l'évaluation de chaque transaction, permettant ainsi de déceler les motifs qui poussent le modèle à considérer une transaction comme suspecte. Quant à la "heatmap", elle met en évidence les relations entre différentes caractéristiques et comment elles interagissent pour influencer la classification. Chaque colonne de la heatmap représente une transaction, ce qui permet de visualiser l'importance des caractéristiques dans la prédiction de la transaction dans ce cluster. Plus la couleur est foncée, plus la caractéristique a un impact sur la prédiction. Cela offre une perspective plus globale des tendances et motifs au sein des données.
De plus, le volume de transactions (*Transactions number*) ou leur montant moyen (*Transactions average amount*) affichés permettent d'affiner la compréhension des schémas de fraude. Si l'expert arrive à identifier un schéma de fraude, il peut lui attribuer un nom (*Name the fraud type*).

En conclusion, ces fonctionnalités permettent la création d'un modèle multi-classes, capable d'attribuer une classe associée à un schéma de fraude à chaque transaction frauduleuse du jeu de données.

4.7.3 L'étude des transactions suspectes basée sur notre ontologie

Le troisième module de notre outil vise à étudier individuellement chaque transaction suspecte. Afin de faciliter l'analyse de ces transactions par les experts, nous organisons les informations en deux catégories distinctes. La première catégorie comprend les informations sur la transaction elle-même et les acteurs avec des connaissances provenant des historiques de transactions des acteurs ou de sources externes (figure 4.19). La deuxième catégorie présente des informations à partir de représentations graphiques en s'appuyant sur le modèle et les transactions liées à la transaction étudiée (figure 4.20).

Informations sur la transaction et ses acteurs La figure 4.19 présente les informations concernant la transaction et les acteurs impliqués. Par ailleurs, nous peuplons également notre ontologie avec des informations externes quand les acteurs sont dans des bases de données publiques telles que dbpedia. Cependant, pour nos expérimentations, nous avons du faire face à l'absence d'informations sur les clients. En conséquence, nous avons été contraints d'associer aléatoirement des personnalités connues aux transactions. Cette démarche a pour objectif de tester le fonctionnement de notre outil.

Transactions liées et choix Pour obtenir une compréhension visuelle plus complète, l'expert dispose d'un graphique des transactions illustré dans la figure 4.20. Les transactions sont liées si elles partagent au moins un acteur commun. La transaction étudiée est en couleur orange. La couleur des autres transactions dépend de leur score de suspicion, si elles sont suspectes, alors elles ont la couleur rouge, autrement si elles sont légitimes, elles ont la couleur verte. Cette visualisation du graphe de transactions liées permet à l'expert de repérer les connexions potentielles entre les transactions. Le tableau à droite du graphe affiche toutes les transactions liées dans le graphique, avec les attributs en commun avec la transaction suspecte mise en évidence. Les attributs en commun avec la transaction étudiée sont en surbrillance rouge.

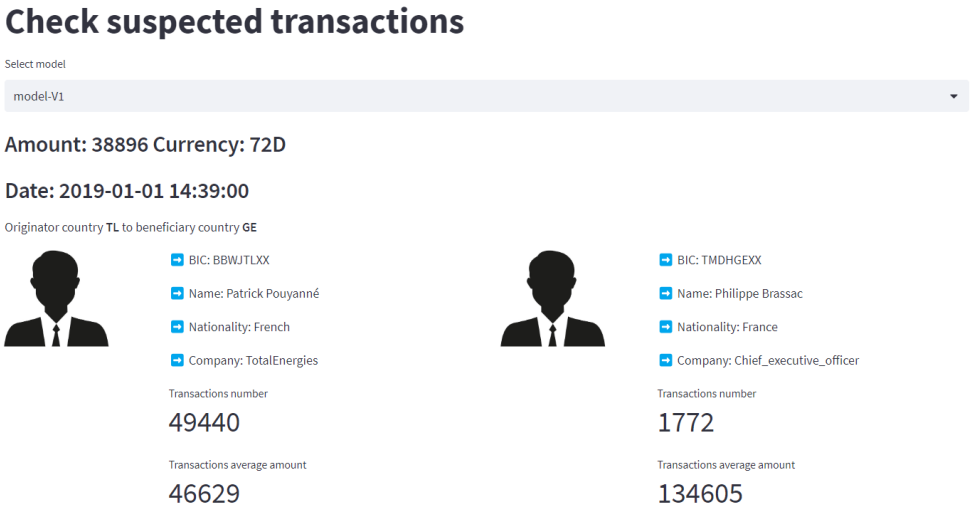

FIGURE 4.19 – ST-Fraud : Détails sur la transaction et ses acteurs

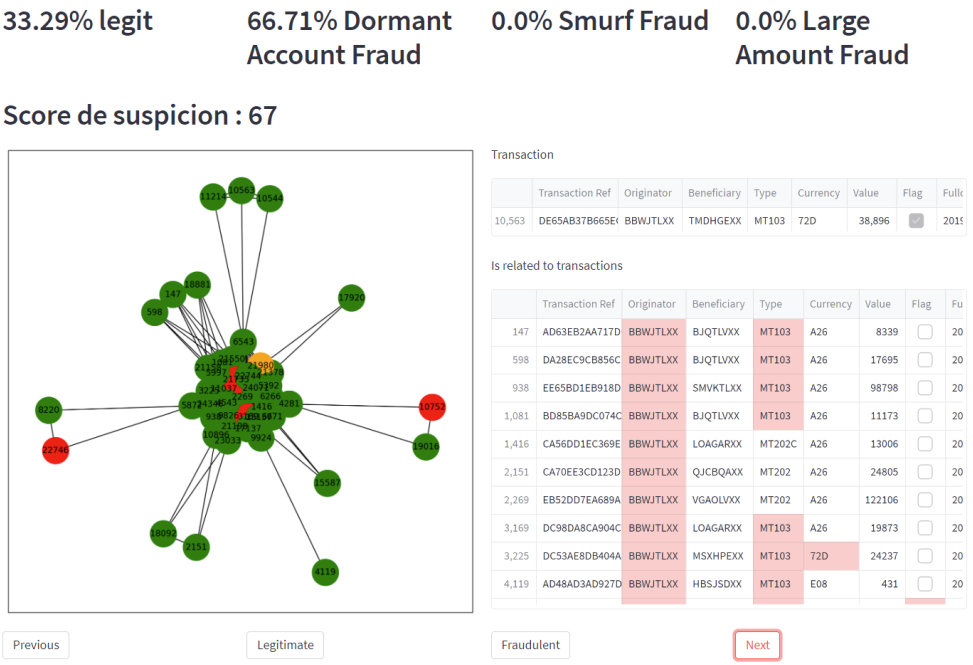

FIGURE 4.20 – ST-Fraud : Visualisation et contrôle

Nous affichons avec le modèle de classification la probabilité d'appartenance de la transaction étudiée aux différentes classes, à savoir les schémas de fraude identifiés. Cette information guide la décision finale de l'expert en fournissant une estimation de la probabilité que la transaction soit frauduleuse, accompagné du score de suspicion de la transaction.

Enfin, renseigné par toutes ces informations, l'expert peut prendre une décision pour bloquer ou libérer la transaction suspecte. Cette décision peut être éclairée par les informations présentées dans le module.

En conclusion, ST-Fraud est un outil qui vise à contribuer à la lutte contre la fraude financière.

En combinant des techniques d'apprentissage automatique avec les éléments de notre ontologie, il vise à apporter une aide supplémentaire aux experts. Cet outil à fait l'œuvre d'une publication dans une conférence nationale [CACC23c]. L'outil ST-Fraud pourrait potentiellement offrir une aide précieuse aux institutions financières en leur fournissant une ressource complémentaire dans leur quête constante d'amélioration des méthodes de détection et de prévention de la fraude.

4.8 Conclusion

Nous avons présenté dans ce chapitre nos expérimentations, à travers l'utilisation de deux ensembles de données : $SWIFT_{base}$, fruit d'une collaboration avec le groupe SKAIZen, et $SWIFT_{base+syn}$, une extension du premier ensemble enrichi d'un schéma de fraude additionnel. Près de 500 caractéristiques, englobant divers éléments des transactions SWIFT comme les pays, devises, acteurs ou encore intermédiaires, ont été calculées. Pour optimiser la pertinence de notre ensemble de données, nous avons eu recours à des techniques basées sur les arbres de décision, nous permettant de réduire la dimensionnalité à 10 caractéristiques pour $SWIFT_{base}$ et à 14 pour $SWIFT_{base+syn}$.

En exploitant le clustering hiérarchique sur les transactions identifiées comme frauduleuses, nous avons dégagé 2 clusters distincts pour $SWIFT_{base}$ et 3 pour $SWIFT_{base+syn}$. Ce regroupement nous a permis de réaliser une tâche de classification plus précise pour classifier les transactions dans cluster associé ou dans la classe légitime. Cette démarche a été suivie d'un entraînement de 4 modèles avec 4 algorithmes distincts sur l'ensemble des transactions, l'objectif était d'identifier le modèle le plus performant en termes de f1-score et de coût. Le modèle basé sur Random Forest s'est distingué par son f1-score, tandis que le modèle XGBoost s'est avéré être le plus économique avec un f1-score très proche de celui du Random Forest. En outre, nous avons calculé un score de suspicion basé sur les prédictions des modèles, offrant ainsi une information supplémentaire de la probabilité de fraude pour chaque transaction.

Nous avons retenu le modèle entraîné via l'algorithme XGBoost combinant ainsi un coût minimal et une performance appréciable. Dans le but d'analyser les schémas de fraude potentiels au sein de nos ensembles de données, nous avons utilisé les valeurs Shapley à travers le framework SHAP. Cela nous a permis d'associer les schémas de fraude identifiés, lors de la phase exploratoire, aux clusters.

Par la suite, l'intégration de nos transactions suspectes identifiées dans notre ontologie a ouvert la voie à l'application de règles, d'axiomes, à la vérification de cohérence, ainsi qu'à la réalisation de requêtes, même si la limitation des données disponibles a quelque peu restreint l'étendue de nos expérimentations.

Enfin, nous avons présenté l'outil ST-Fraud permettant l'utilisation de notre méthode avec une interface. Structuré en trois modules distincts, il offre aux utilisateurs une solution pour importer leurs données, entraîner un modèle de classification, l'identification de schéma de fraude et enfin le contrôle des transactions suspectes avec des visualisations.

Chapitre 5

Conclusion et perspectives

Sommaire

5.1 Synthèse

La détection des transactions frauduleuses revêt une importance cruciale pour les institutions financières. À mesure que les échanges de transactions internationales deviennent de plus en plus complexes et rapides, elles exposent davantage les institutions financières à des fraudes sophistiquées. Ces fraudes représentent une menace pour la stabilité économique, tant au niveau des institutions financières que des particuliers. Dans ce contexte, les méthodes traditionnelles fondées sur des systèmes de règles atteignent leurs limites. C'est précisément dans ce contexte que les techniques fondées sur l'apprentissage automatique peuvent apporter une contribution significative à l'amélioration des systèmes de détection de fraude.
Les travaux présentés dans cette thèse se situent dans le contexte de la détection de fraude financière dans le réseau SWIFT, en se focalisant sur les techniques d'apprentissage automatique. Les principales contributions de cette thèse sont des approches de détection et d'analyse de schémas de fraude en utilisant des approches d'apprentissage automatique et sémantique. En outre, nous avons introduit l'outil ST-Fraud, conçu pour assister les experts du domaine dans leur processus de prise de décision concernant le blocage et la libération de transactions.

Dans cette thèse, nous avons abordé principalement le problème de détection de transactions frauduleuses, l'analyse de schémas de fraude, ainsi l'interprétabilité des modèles de classification de transaction. Afin de situer nos travaux, nous avons présenté dans le chapitre 2 les techniques basées sur l'apprentissage automatique dans le domaine de la détection des transactions frauduleuses. Nous avons également présenté un résumé des approches sémantiques, explorant l'utilisation des ontologies du domaine pour représenter les connaissances contextuelles et faciliter l'interprétation des résultats de détection.
Cet état de l'art nous a conduits à conclure que malgré les avancées significatives dans ces deux domaines, peu d'études ont exploré l'intégration de l'apprentissage automatique et des ontologies dans le contexte spécifique des transactions SWIFT.

Le chapitre 3 a exposé la méthode que nous avons adoptée pour résoudre les limitations des travaux de l'état de l'art et répondre à nos questions de recherche. La méthode présentée est une approche hybride qui combine l'apprentissage automatique pour la détection de fraude, le clustering hiérarchique pour l'identification de schémas de fraude, et l'ontologie pour le contrôle des transactions.

Le chapitre 4 sur l'expérimentation a constitué une partie importante de notre travail. Nous avons présenté les résultats de nos expérimentations et les avons analysés en profondeur. Ces résultats ont servi à valider la méthode que nous avons élaborée.
Enfin, nous avons développé l'outil ST-Fraud utilisant la méthode proposée. Cet outil, composé de trois modules, permet aux utilisateurs d'appliquer la méthode pour détecter et identifier des transactions frauduleuses, offrant ainsi une solution pratique et efficace pour lutter contre la fraude dans le domaine des transactions SWIFT.

5.2 Contributions

Les travaux de cette thèse combinent des approches d'apprentissage automatique et des bases de connaissances sémantiques pour résoudre le problème de la détection de transactions frau-

duleuses dans le réseau SWIFT. Notre méthodologie consiste à classifier les transactions en catégories frauduleuses et légitimes, puis à analyser en profondeur les schémas de fraude afin d'assister les experts dans leur processus de prise de décision.
Nos contributions se composent de deux parties complémentaires, ainsi que d'un outil pour la **détection** et l'**analyse** des fraudes financières.

Détection de transactions frauduleuses Nous avons présenté notre approche hybride pour la détection et le contrôle des transactions suspectes dans le réseau SWIFT. Cette approche combine des techniques d'apprentissage automatique avec une ontologie créée pour le domaine de recherche. Notre modèle de classification multi-classe permet de séparer les transactions légitimes de celles associées à des schémas de fraude. En parallèle, nous calculons un score de suspicion pour évaluer de manière quantitative la probabilité de fraude pour chaque transaction. Enfin, pour établir le seuil de suspicion permettant de qualifier une transaction de suspecte, nous avons proposé une évaluation quantitative des coûts financiers liés aux prédictions de notre modèle.

Analyse de transactions frauduleuses Dans cette thèse, nous avons défini les caractéristiques pertinentes liées aux transactions SWIFT, en considérant des facteurs tels que les pays et les devises, puis en sélectionnant les caractéristiques les plus significatives. Ensuite, une analyse collaborative avec les experts a permis de définir les schémas de fraude dans le jeu de données et les transactions frauduleuses ont été regroupées en clusters grâce à des méthodes d'apprentissage non supervisé. Pour finir, des techniques d'interprétation et de visualisation ont été utilisées pour associer chaque cluster à un schéma de fraude spécifique.
Nous avons créé une ontologie du domaine qui a permis une analyse approfondie des transactions suspectes détectées par le modèle, contribuant ainsi à la validation par les experts du domaine.

Un outil a été développé pour valider les propositions avancées dans le domaine de la détection de fraude dans le réseau SWIFT. Cet outil offre une approche complète, combinant des calculs de caractéristiques pertinentes, des outils de visualisation, ainsi que le modèle de classification pour distinguer les transactions légitimes des transactions frauduleuses.

Cette thèse a apporté des contributions significatives au domaine de détection de fraude dans le réseau international et interbancaire SWIFT. Elle ouvre également la voie à de nouvelles problématiques de recherche dans ce domaine.

5.3 Perspectives

Ces travaux ont posé les fondations pour de nouvelles avancées dans le domaine de la détection de fraude. Ils ouvrent de nombreuses perspectives dans ce domaine. Parmi ces perspectives, nous pouvons citer quatre futurs travaux de recherche :

Analyse du comportement et surveillance basée sur les graphes : L'analyse du comportement des fraudeurs est importante pour les institutions financières. Nous avons étudié dans cette thèse le comportement du client à travers son historique de transactions. Notre analyse s'est trouvée limitée, car nous avons seulement considéré une partie des transactions, ne prenant pas en compte l'intégralité des échanges transactionnels.

Nous envisageons d'approfondir cette analyse en temps réel avec une prise de décision automatique en prenant en compte l'intégralité des transactions et des connexions entre les acteurs. Les techniques d'apprentissage automatique sont utilisées au sein de la communauté scientifique pour la détection de fraude. Elles permettent de classifier des volumes de données conséquents. Plus récemment, des techniques basées sur des graphes sont étudiés pour extraire des relations entre les clients, ainsi que pour identifier des sous-graphes présentant des potentiels comportements frauduleux ou schémas de fraude. Le *graph embedding* est une approche efficace pour la détection de fraude et de schémas complexes dans un réseau de transaction financières. Les fraudes qui représentent des anomalies dans les graphes peuvent être détectées par des connexions suspectes ou des comportements anormaux. La représentation des nœuds du graphe permet d'obtenir des caractéristiques représentant des informations contextuelles sur les relations entre les acteurs ou les transactions, ces caractéristiques peuvent ainsi contribuer à l'amélioration des systèmes de détection de fraude.

Extraction de connaissances à partir de données non structurées : L'utilisation de techniques de traitement automatique du langage naturel (NLP) dans la détection de fraude peut améliorer la capacité des institutions financières à détecter les activités frauduleuses. Les messages SWIFT, bien que semi-structurés, contiennent des informations textuelles relatives aux transactions, ainsi que d'autres données pertinentes. Les techniques NLP peuvent être utilisées pour analyser ces messages pour la détection d'anomalies dans le texte, et des incohérences dans les données.

Les méthodes de traitement du langage naturel peuvent être appliqués à des sources externes telles que les forums en ligne ou les réseaux sociaux pour identifier des discussions liées à une fraude potentielle. L'extraction d'entités nommées et de relations permet d'identifier des concepts tels que les noms de personnes ou d'entreprises, ainsi les relations existantes qui pourraient être associées à des activités frauduleuses. De plus, ces méthodes peuvent également être utilisées dans le cadre de la conformité KYC (Know Your Customer) afin d'obtenir des informations sur les acteurs impliqués dans des transactions financières.

Génération de transactions synthétiques pour améliorer la détection : Le manque de données constitue un obstacle majeur au développement de nouvelles approches dans le domaine financier. Bien que des travaux aient été entrepris pour générer synthétiquement des

transactions financières, ils ne prennent pas en compte les spécificités interbancaires et internationales des transactions SWIFT. La proposition de travaux dédiés à la génération synthétique de transactions SWIFT serait bénéfique pour les institutions financières lorsqu'elles mettent à jour leurs systèmes de traitement des flux financiers.
Actuellement, ces institutions font face à des défis considérables lorsqu'il s'agit de tester ces mises à jour. En générant un grand volume de transactions, ces institutions seraient en mesure de réaliser des tests de bien meilleure qualité. Cette approche permettrait de repérer les potentiels bugs dans leurs plateformes.
De plus, les techniques émergentes de génération de données, notamment l'utilisation des réseaux antagonistes génératifs (GAN), offrent la possibilité de créer des données synthétiques. Les GAN sont des algorithmes qui permettent de générer des données synthétiques en s'appuyant sur des modèles d'apprentissage profond.

Alignement des ontologies pour une vision unifiée : Les institutions financières gèrent des informations de sources multiples et hétérogènes, telles que les transactions, les informations client (KYC), les réglementations de banques étrangères, etc. Ces connaissances peuvent être représentées et structurées à l'aide d'ontologies.
L'alignement d'ontologies permet de mettre en évidence les liens de correspondance entre deux ontologies. Dans le domaine de détection de fraude, il peut être utilisé pour identifier les correspondances anormales entre les différents acteurs impliqués dans les transactions en alignant les ontologies des différentes entités acteurs : client, banque, entreprises, etc.

Annexe A

Projet KBP : Knowledge Base Population

A.1 Présentation du projet

Initiée au sein du pôle R&D de SKAIZen Group, la démarche KBP est consacrée à travers deux publications [JSC19, JSZC20].

Le secteur financier est constamment soumis à d'importantes réglementations, élaborées pour réduire les risques et prévenir les opérations illicites telles que le blanchiment d'argent et le financement du terrorisme. Au cœur de ces réglementations se trouve le processus KYC (Know Your Customer), qui exige des institutions financières qu'elles maintiennent une connaissance actualisée et approfondie de leurs clients afin de sécuriser chaque transaction financière. Cependant le mise œuvre et la mise à jour régulière du KYC peuvent s'avérer être des processus manuels et fastidieux. C'est dans cette optique que le projet KBP a été développé, avec pour objectif principal d'automatiser la collecte et l'analyse des informations nécessaires au KYC, rendant ainsi le processus plus efficace tout en garantissant une totale conformité. Pour ce faire, le projet KBP a constitué un corpus annoté basé sur des articles financiers en langue française, permettant ainsi de décrypter et d'analyser les tendances financières.

A.2 KYC : Know Your Customer

Know Your Customer est le processus par lequel les institutions financières valident et vérifient l'identité de leurs clients. Ce processus joue un rôle déterminant dans la lutte contre les crimes financiers, tels que la fraude et le blanchiment d'argent, garantissant ainsi la sécurité des transactions financières et protégeant à la fois l'institution et ses clients des menaces potentielles. Institué par des recommandations du Groupe d'action financière (GAFI) et adopté par plus de 190 pays, le KYC se décompose en trois piliers principaux.

Le premier pilier repose sur le CIP (Customer Identification Program), un programme d'identification du client qui implique une vérification rigoureuse des informations fournies par le client, qu'il s'agisse d'une personne physique ou d'une entreprise, afin de garantir leur authenticité. Cela englobe la collecte et la validation de données telles que le nom, l'adresse, la date de naissance, ainsi que d'autres identifiants officiels.

Le deuxième pilier, est le CDD (Customer Due Diligence), qui consiste en une évaluation du niveau de risque associé à chaque client. Cette évaluation nécessite une diligence accrue pour les clients présentant un risque élevé, tandis que les clients à risque plus faible peuvent

bénéficier d'une diligence allégée.

Le troisième pilier concerne la surveillance continue. Les institutions financières sont tenues de surveiller activement les activités de leurs clients afin de détecter tout changement potentiel pouvant indiquer un risque accru pour le client ou la nécessité d'obtenir des informations complémentaires.

Bibliographie

[AAM+21] Mansoor AHMED, Kainat ANSAR, Cal B MUCKLEY, Abid KHAN, Adeel ANJUM et Muhammad TALHA : A semantic rule based digital fraud detection. *PeerJ Computer Science*, 7:e649, 2021.

[ACA] Lylia ABROUK, Hamza CHERGUI et Hamid AHAGGACH : Ontofic : an ontology for financial fraud detection and customer behavior modeling. *In International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2023*.

[ACA+23] Lylia ABROUK, Hamza CHERGUI, Hamid AHAGGACH, Benjamin AUGER et Dominique CHERON : Ontofic : une ontologie pour la détectionde fraude financière et la modélisation descomportements des clients. *In INFormatique des Organisations et Systèmes d'Information et de Décision, INFORSID 2023, La Rochelle, France*, pages 97–102, 2023.

[ACC+22] Benjamin AUGER, Hamza CHERGUI, Yara CHEHADE, Jana El KADRI, Lylia ABROUK et Nicolas CABIOCH : Construction d'une ontologie dans le domaine financier pour la détection de fraudes. *In INFORSID'22*, pages 157–162, 2022.

[AF22] Noor Saleh ALFAIZ et Suliman Mohamed FATI : Enhanced credit card fraud detection model using machine learning. *Electronics*, 11(4):662, 2022.

[AJA20] Doaa ALMHAITHAWI, Assef JAFAR et Mohamad ALJNIDI : Example-dependent cost-sensitive credit cards fraud detection using SMOTE and Bayes minimum risk. *SN Applied Sciences*, 2(9):1–12, 2020.

[AMPK18] Girija ATTIGERI, Manohara Pai MM, Radhika M PAI et Rahul KULKARNI : Knowledge base ontology building for fraud detection using topic modeling. *Procedia Computer Science*, 135:369–376, 2018.

[BCMM21] Candice BENTÉJAC, Anna CSÖRGŐ et Gonzalo MARTÍNEZ-MUÑOZ : A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54:1937–1967, 2021.

[Bel61] RE BELLMAN : Adaptive control processes, 255 princeton university press. *Princeton, New Jersey*, 1961.

[BH02] Richard J BOLTON et David J HAND : Statistical fraud detection : A review. *Statistical science*, 17(3):235–255, 2002.

[BJTW11] Siddhartha BHATTACHARYYA, Sanjeev JHA, Kurian THARAKUNNEL et J Christopher WESTLAND : Data mining for credit card fraud : A comparative study. *Decision support systems*, 50(3):602–613, 2011.

[BKG20] Dor BANK, Noam KOENIGSTEIN et Raja GIRYES : Autoencoders. *arXiv preprint arXiv :2003.05991*, 2020.

[BKN+08] Robert C BLATTBERG, Byung-Do KIM, Scott A NESLIN, Robert C BLATTBERG, Byung-Do KIM et Scott A NESLIN : *Why database marketing ?* Springer, 2008.

[BLS+22] Vadim BORISOV, Tobias LEEMANN, Kathrin SESSLER, Johannes HAUG, Martin PAWELCZYK et Gjergji KASNECI : Deep neural networks and tabular data : A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[BMT06] Brain Leke BETECHUOH, Tshilidzi MARWALA et Thando TETTEY : Autoencoder networks for HIV classification. *Current Science (00113891)*, 91(11), 2006.

[Bon17] R. BONNIN : *Machine Learning for Developers*. Packt Publishing, 2017.

[Bre96] Leo BREIMAN : Bagging predictors. *Machine learning*, 24:123–140, 1996.

[Bre01] Leo BREIMAN : Random forests. *Machine learning*, 45(1):5–32, 2001.

[BSAO13] Alejandro Correa BAHNSEN, Aleksandar STOJANOVIC, Djamila AOUADA et Björn OTTERSTEN : Cost sensitive credit card fraud detection using Bayes minimum risk. *In 2013 12th international conference on machine learning and applications*, volume 1, pages 333–338. IEEE, 2013.

[CAA+22] Hamza CHERGUI, Romain A. ALFRED, Lylia ABROUK, Ali JABBARI et Nadine CULLOT : Détection d'anomalies : une méthode appliquée aux transactions interbancaires. *In* Sihem AMER-YAHIA et Arnaud SOULET, éditeurs : *Extraction et Gestion des Connaissances, EGC 2022, Blois, France, 24 au 28 janvier 2022*, volume E-38 de *RNTI*, pages 479–480. Editions RNTI, 2022.

[Cab22] Guillaume CABANAC : Decontamination of the scientific literature. *arXiv preprint arXiv :2210.15912*, 2022.

[CACC22] Hamza CHERGUI, Lylia ABROUK, Nadine CULLOT et Nicolas CABIOCH : Détection de fraude financière dans unsystème de transactions interbancaires. *In INFormatique des Organisations et Systèmes d'Information et de Décision, INFORSID 2022, Dijon, France*, pages 141–156, 2022.

[CACC23a] Hamza CHERGUI, Lylia ABROUK, Nadine CULLOT et Nicolas CABIOCH : Réduction du risque du coût d'un modèle dans la détection de fraude financière. *In* Catherine FARON et Sabine LOUDCHER, éditeurs : *Extraction et Gestion des Connaissances, EGC 2023, Lyon, France, 16 - 20 janvier 2023*, volume E-39 de *RNTI*, pages 641–642. Editions RNTI, 2023.

[CACC23b] Hamza CHERGUI, Lylia ABROUK, Nadine CULLOT et Nicolas CABIOCH : Semi-supervised method to detect fraudulent transactions and identify fraud types while minimizing mounting costs. *International Journal of Advanced Computer Science and Applications*, 14(2), 2023.

[CACC23c] Hamza CHERGUI, Lylia ABROUK, Nadine CULLOT et Nicolas CABIOCH : St-fraud : un outil de détection de transactions frauduleuses. *In INFormatique des Organisations et Systèmes d'Information et de Décision, INFORSID 2023, La Rochelle, France*, pages 105–109, 2023.

[CAO15] Alejandro CORREA BAHNSEN, Djamila AOUADA et Björn OTTERSTEN : Example-dependent cost-sensitive decision trees. *Expert Systems with Applications*, 42(19):6609–6619, 2015.

[CBHK02] Nitesh V. CHAWLA, Kevin W. BOWYER, Lawrence O. HALL et W. Philip KEGELMEYER : SMOTE : synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[CFRS19] Salvatore CARTA, Gianni FENU, Diego Reforgiato RECUPERO et Roberto SAIA : Fraud detection for e-commerce transactions by employing a prudential multiple consensus model. *Journal of Information Security and Applications*, 46:13–22, 2019.

[CHB+15] Tianqi CHEN, Tong HE, Michael BENESTY, Vadim KHOTILOVICH, Yuan TANG, Hyunsu CHO, Kailong CHEN, Rory MITCHELL, Ignacio CANO, Tianyi ZHOU *et al.* : XGBoost : extreme gradient boosting. *R package version 0.4-2*, 1(4):1–4, 2015.

[CHC+09] Israel COHEN, Yiteng HUANG, Jingdong CHEN, Jacob BENESTY, Jacob BENESTY, Jingdong CHEN, Yiteng HUANG et Israel COHEN : Pearson correlation coefficient. *Noise reduction in speech processing*, pages 1–4, 2009.

[CLBC+21] Fabrizio CARCILLO, Yann-Aël LE BORGNE, Olivier CAELEN, Yacine KESSACI, Frédéric OBLÉ et Gianluca BONTEMPI : Combining unsupervised and supervised learning in credit card fraud detection. *Information sciences*, 557:317–331, 2021.

[CPC19] Diogo V CARVALHO, Eduardo M PEREIRA et Jaime S CARDOSO : Machine learning interpretability : A survey on methods and metrics. *Electronics*, 8(8): 832, 2019.

[CS08] Kim CHURCH et Rod SMITH : Rea ontology-based simulation models for enterprise strategic planning. *Journal of information systems*, 22(2):301–329, 2008.

[CTN+18] Zhiyuan CHEN, Ee Na TEOH, Amril NAZIR, Ettikan Kandasamy KARUPPIAH, Kim Sim LAM *et al.* : Machine learning techniques for anti-money laundering (aml) solutions in suspicious transaction detection : a review. *Knowledge and Information Systems*, 57(2):245–285, 2018.

[CXD20] Liming CHEN, Baoxin XIU et Zhaoyun DING : Finding misstatement accounts in financial statements through ontology reasoning. *IEEE Access*, 2020.

[DBM21] Roxane DESROUSSEAUX, Gilles BERNARD et Jean-Jacques MARIAGE : Profiling money laundering with neural networks : a case study on environmental crime detection. *In 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 364–369. IEEE, 2021.

[DJ21] Mwamba Kasongo DAHOUDA et Inwhee JOE : A deep-learned embedding technique for categorical features encoding. *IEEE Access*, 9:114381–114391, 2021.

[DZ04] Saso DZEROSKI et Bernard ZENKO : Is combining classifiers with stacking better than selecting the best one ? *Machine learning*, 54:255–273, 2004.

[Elk01] Charles ELKAN : The foundations of cost-sensitive learning. *In International joint conference on artificial intelligence*, volume 17, pages 973–978. Lawrence Erlbaum Associates Ltd, 2001.

[EOB19] Ahmed EL ORCHE et Mohamed BAHAJ : Approach to use ontology based on electronic payment system and machine learning to prevent fraud. *In Proceedings of the 2nd International Conference on Networking, Information Systems & Security*, pages 1–6, 2019.

[EOBA18] Ahmed EL ORCHE, Mohamed BAHAJ et Soumya Ain ALHAYAT : Ontology based on electronic payment fraud prevention. pages 143–148, 2018.

[Fla12] Peter FLACH : *Machine learning : the art and science of algorithms that make sense of data*. Cambridge university press, 2012.

[FLW+19] Yixuan FENG, Chao LI, Yun WANG, Jian WANG, Guigang ZHANG, Chunxiao XING, Zhenxing LI et Zengshen LIAN : Anti-money laundering (AML) research : a system for identification and multi-classification. *In Web Information Systems and Applications : 16th International Conference, WISA 2019, Qingdao, China, September 20-22, 2019, Proceedings 16*, pages 169–175. Springer, 2019.

[Fos19] Leopold Ghemmogne FOSSI : *Gestion des règles basée sur l'indice de puissance pour la détection de fraude : Approches supervisées et semi-supervisées*. Thèse de doctorat, Université de Lyon ; Università degli studi (Milan, Italie), 2019.

[Fre95] Yoav FREUND : Boosting a weak learning algorithm by majority. *Information and computation*, 121(2):256–285, 1995.

[GE03] Isabelle GUYON et André ELISSEEFF : An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[GM02] Guido L GEERTS et William E MCCARTHY : An ontological analysis of the economic primitives of the extended-rea enterprise information architecture. *International Journal of Accounting Information Systems*, 3(1):1–16, 2002.

[GOV22] Léo GRINSZTAJN, Edouard OYALLON et Gaël VAROQUAUX : Why do tree-based models still outperform deep learning on tabular data ? *arXiv preprint arXiv :2207.08815*, 2022.

[GWBV02] Isabelle GUYON, Jason WESTON, Stephen BARNHILL et Vladimir VAPNIK : Gene selection for cancer classification using support vector machines. *Machine learning*, 46:389–422, 2002.

[HDO+98] Marti A. HEARST, Susan T DUMAIS, Edgar OSUNA, John PLATT et Bernhard SCHOLKOPF : Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[HGL22] Adamu HUSSAINI, Zahia GUESSOUM et Eunika Mercier LAURENT : Elaboration of financial fraud ontology. *Annals of Computer Science and Information Systems*, 32:277–285, 2022.

[HM01] Volker HAARSLEV et Ralf MÖLLER : Racer system description. *In Automated Reasoning : First International Joint Conference, IJCAR 2001 Siena, Italy, June 18–22, 2001 Proceedings 1*, pages 701–705. Springer, 2001.

[HTF01] Trevor HASTIE, Robert TIBSHIRANI et Jerome FRIEDMAN : The elements of statistical learning. Springer series in statistics. *New York, NY, USA*, 2001.

[JBB15] Alan JOVIĆ, Karla BRKIĆ et Nikola BOGUNOVIĆ : A review of feature selection methods with applications. *In 2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)*, pages 1200–1205. IEEE, 2015.

[Jen97] David JENSEN : Prospective assessment of ai technologies for fraud detection : A case study. *In AAAI Workshop on AI Approaches to Fraud Detection and Risk Management*, pages 34–38, 1997.

[JSC19] Ali JABBARI, Olivier SAUVAGE et Nicolas CABIOCH : Towards a knowledge base of financial relations : Overview and project description. *In 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pages 313–316. IEEE, 2019.

[JSZC20] Ali JABBARI, Olivier SAUVAGE, Hamada ZEINE et Hamza CHERGUI : A French corpus and annotation schema for named entity recognition and relation extraction of financial news. *In Proceedings of the 12th Language Resources and Evaluation Conference*, pages 2293–2299, 2020.

[KC08] Riydah KENAYA et Ka C CHEOK : Euclidean art neural networks. *In Proceedings of the world congress on engineering and computer science*, numéro 1, 2008.

[KDT20] Ashwini KUMAR, Sanjoy DAS et Vishu TYAGI : Anti money laundering detection using Naïve Bayes classifier. *In 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, pages 568–572. IEEE, 2020.

[KFMBEN21] Shima KHALILINEZHAD, Hamed FAZLOLLAHTABAR, Behrouz MINAEI-BIDGOLI et Hamid ESLAMI NOSRATABADI : Detecting valuable customers using the trade patterns of financial transactions applying integrated RFM and OLAP. *Int. J. Ind. Eng. Prod. Res*, 32:1–15, 2021.

[KMF+17] Guolin KE, Qi MENG, Thomas FINLEY, Taifeng WANG, Wei CHEN, Weidong MA, Qiwei YE et Tie-Yan LIU : Lightgbm : A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.

[Kno20] Andres KNOBEL : Swift data can be a global vantage point for tackling global money laundering. *In Tax Justice Network*, Sep 2020.

[KSV04] John KINGSTON, Burkhard SCHAFER et Wim VANDENBERGHE : Towards a financial fraud ontology : A legal modelling approach. *AI & L.*, 12:419, 2004.

[KT11] Liu KEYAN et Yu TINGTING : An improved support-vector network model for anti-money laundering. *In 2011 Fifth International Conference on Management of e-Commerce and e-Government*, pages 193–196. IEEE, 2011.

[KZ17] Zahra KAZEMI et Houman ZARRABI : Using deep networks for fraud detection in the credit card transactions. *In 2017 IEEE 4th International conference on knowledge-based engineering and innovation (KBEI)*, pages 0630–0633. IEEE, 2017.

[LH11] Asma S LARIK et Sajjad HAIDER : Clustering based anomalous transaction reporting. *Procedia Computer Science*, 3:606–610, 2011.

[LJZ08] Lin-Tao LV, Na JI et Jiu-Long ZHANG : A RBF neural network model for anti-money laundering. *In 2008 International conference on wavelet analysis and pattern recognition*, volume 1, pages 209–215. IEEE, 2008.

[LKK10] Nhien An LE KHAC et M-Tahar KECHADI : Application of data mining for anti-money laundering detection : A case study. *In 2010 IEEE International Conference on Data Mining Workshops*, pages 577–584. IEEE, 2010.

[LL17] Scott M LUNDBERG et Su-In LEE : A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

[LRA12] Edgar Alonso LOPEZ-ROJAS et Stefan AXELSSON : Money laundering detection using synthetic data. *In Annual workshop of the Swedish Artificial Intelligence Society (SAIS)*. Linköping University Electronic Press, Linköpings universitet, 2012.

[LSA+20] Joana LORENZ, Maria Inês SILVA, David APARÍCIO, João Tiago ASCENSÃO et Pedro BIZARRO : Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity. *In Proceedings of the First ACM International Conference on AI in Finance*, pages 1–8, 2020.

[M+07] Tom Michael MITCHELL *et al.* : *Machine learning*, volume 1. McGraw-Hill New York, 2007.

[Man16] Louise MANNING : Food fraud : policy and food chain. *Current Opinion in Food Science*, 10:16–21, 2016. Innovation in food science • Foodomics technologies.

[MBB+18] Guillaume METZLER, Xavier BADICHE, Brahim BELKASMI, Elisa FROMONT, Amaury HABRARD et Marc SEBBAN : Tree-based cost sensitive methods for fraud detection in imbalanced data. *In International Symposium on Intelligent Data Analysis*, pages 213–224. Springer, 2018.

[MC12] Fionn MURTAGH et Pedro CONTRERAS : Algorithms for hierarchical clustering : an overview. *Wiley Interdisciplinary Reviews : Data Mining and Knowledge Discovery*, 2(1):86–97, 2012.

[MK11] Krzysztof MICHALAK et Jerzy KORCZAK : Graph mining approach to suspicious transaction detection. *In 2011 Federated conference on computer science and information systems (FedCSIS)*, pages 69–75. IEEE, 2011.

[Mol19] Christoph MOLNAR : *Interpretable Machine Learning*. 2019. https://christophm.github.io/interpretable-ml-book/.

[MR93] Andrzej MAĆKIEWICZ et Waldemar RATAJCZAK : Principal components analysis (PCA). *Computers & Geosciences*, 19(3):303–342, 1993.

[MT20] Luke MERRICK et Ankur TALY : The explanation game : Explaining machine learning models using Shapley values. *In Machine Learning and Knowledge Extraction : 4th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2020, Dublin, Ireland, August 25–28, 2020, Proceedings 4*, pages 17–38. Springer, 2020.

[MTGS20] Sumit MISRA, Soumyadeep THAKUR, Manosij GHOSH et Sanjoy Kumar SAHA : An autoencoder based model for detecting fraudulent credit card transaction. *Procedia Computer Science*, 167:254–262, 2020.

[NP01] Ian NILES et Adam PEASE : Towards a standard upper ontology. *In Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001*, pages 2–9, 2001.

[oEB23] The Editors of ENCYCLOPAEDIA BRITANNICA : Fraud, 2023.

[PGV+18] Liudmila PROKHORENKOVA, Gleb GUSEV, Aleksandr VOROBEV, Anna Veronika DOROGUSH et Andrey GULIN : CatBoost : unbiased boosting with categorical features. *Advances in neural information processing systems*, 31, 2018.

[PL18] Apapan PUMSIRIRAT et Yan LIU : Credit card fraud detection using deep learning based on auto-encoder and restricted Boltzmann machine. *International Journal of advanced computer science and applications*, 9(1), 2018.

[PLCM16] Ebberth L PAULA, Marcelo LADEIRA, Rommel N CARVALHO et Thiago MARZAGAO : Deep learning anomaly detection as support fraud investigation in Brazilian exports and anti-money laundering. *In 2016 15th IEEE international conference on machine learning and applications (ICMLA)*, pages 954–960. IEEE, 2016.

[PM18] Utkarsh PORWAL et Smruthi MUKUND : Credit card fraud detection in e-commerce : An outlier detection approach. *arXiv preprint arXiv :1811.02196*, 2018.

[POKB20] Tahereh POURHABIBI, Kok-Leong ONG, Booi H KAM et Yee Ling BOO : Fraud detection : A systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems*, 133:113303, 2020.

[PR19] Debachudamani PRUSTI et Santanu Kumar RATH : Fraudulent transaction detection in credit card by applying ensemble machine learning techniques. *In 2019 10th international conference on computing, communication and networking technologies (ICCCNT)*, pages 1–6. IEEE, 2019.

[Qui86] J. Ross QUINLAN : Induction of decision trees. *Machine learning*, 1:81–106, 1986.

[RE20] Naoufal RTAYLI et Nourddine ENNEYA : Selection features and support vector machine for credit card risk identification. *Procedia Manufacturing*, 46:941–948, 2020.

[RH11] Saleha RAZA et Sajjad HAIDER : Suspicious activity reporting using dynamic bayesian networks. *Procedia Computer Science*, 3:987–991, 2011.

[RKLH14] Quratulain RAJPUT, Nida Sadaf KHAN, Asma LARIK et Sajjad HAIDER : Ontology based expert-system for suspicious transactions detection. *Computer and Information Science*, 7(1):103, 2014.

[RM05] Lior ROKACH et Oded MAIMON : Clustering methods, 2005.

[RSM+18] Abhimanyu ROY, Jingyi SUN, Robert MAHONEY, Loreto ALONZI, Stephen ADAMS et Peter BELING : Deep learning detecting fraud in credit card transactions. *In 2018 Systems and Information Engineering Design Symposium (SIEDS)*, pages 129–134. IEEE, 2018.

[Rum11] D.J. RUMSEY : *Statistics For Dummies*. –For dummies. Wiley, 2011.

[SBHS+21] Branka STOJANOVIĆ, Josip BOŽIĆ, Katharina HOFER-SCHMITZ, Kai NAHRGANG, Andreas WEBER, Atta BADII, Maheshkumar SUNDARAM, Elliot JORDAN et Joel RUNEVIC : Follow the trail : Machine learning for fraud detection in fintech applications. *Sensors*, 21(5):1594, 2021.

[SDPH+21] Bram STEENWINCKEL, Dieter DE PAEPE, Sander Vanden HAUTTE, Pieter HEYVAERT, Mohamed BENTEFRIT, Pieter MOENS, Anastasia DIMOU, Bruno VAN DEN BOSSCHE, Filip DE TURCK, Sofie VAN HOECKE *et al.* : Flags : A methodology for adaptive anomaly detection and root cause analysis on sensor data streams by fusing expert knowledge with machine learning. *Future Generation Computer Systems*, 116:30–48, 2021.

[SHK18] Tom SWEERS, Tom HESKES et Jesse KRIJTHE : Autoencoding credit card fraud. *Bachelor Thesis*, 2018.

[SM22] G. SAPORTA et S. MARANEY : *Practical Fraud Prevention : Fraud and AML Analytics for Fintech and ECommerce, Using SQL and Python*. O'Reilly Media, Incorporated, 2022.

[SMH08] Robert DC SHEARER, Boris MOTIK et Ian HORROCKS : Hermit : A highly-efficient OWL reasoner. *In Owled*, volume 432, page 91, 2008.

[SPG+07] Evren SIRIN, Bijan PARSIA, Bernardo Cuenca GRAU, Aditya KALYANPUR et Yarden KATZ : Pellet : A practical OWL-DL reasoner. *Journal of Web Semantics*, 5(2):51–53, 2007.

[SSS18] Zulazeze SAHRI, Shuhaida Mohammed SHUHIDAN et Zuraidah Mohd SANUSI : An ontology-based representation of financial criminology domain using text analytics processing. *International Journal of Computer Science and Network Security*, 18(2):56–62, 2018.

[SZA22] Ravid SHWARTZ-ZIV et Amitai ARMON : Tabular data : Deep learning is not all you need. *Information Fusion*, 81:84–90, 2022.

[TGIH17] Alaa THARWAT, Tarek GABER, Abdelhameed IBRAHIM et Aboul Ella HASSANIEN : Linear discriminant analysis : A detailed tutorial. *AI communications*, 30(2):169–190, 2017.

[Tho84] Bruce THOMPSON : *Canonical correlation analysis : Uses and interpretation*. Numéro 47. Sage, 1984.

[Tib97] Robert TIBSHIRANI : The lasso method for variable selection in the Cox model. *Statistics in medicine*, 16(4):385–395, 1997.

[TL20] Rodrigo Araujo Lima TORRES et Marcelo LADEIRA : A proposal for online analysis and identification of fraudulent financial transactions. *In 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 240–245. IEEE, 2020.

[TLYW18] Xiao-Bo TANG, Guang-Chao LIU, Jing YANG et Wei WEI : Knowledge-based financial statement fraud detection system : based on an ontology and a decision tree. *Knowledge organization*, 45(3):205–219, 2018.

[TY05] Jun TANG et Jian YIN : Developing an intelligent data discriminating system of anti-money laundering based on SVM. *In 2005 International conference on machine learning and cybernetics*, volume 6, pages 3453–3457. IEEE, 2005.

[Vel20] Alfredo VELLIDO : The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural computing and applications*, 32(24):18069–18083, 2020.

[VKS+19] Dejan VARMEDJA, Mirjana KARANOVIC, Srdjan SLADOJEVIC, Marko ARSENOVIC et Andras ANDERLA : Credit card fraud detection-machine learning methods. *In 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, pages 1–5. IEEE, 2019.

[VVBC+15] Véronique VAN VLASSELAER, Cristián BRAVO, Olivier CAELEN, Tina ELIASSI-RAD, Leman AKOGLU, Monique SNOECK et Bart BAESENS : Apate : A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems*, 75:38–48, 2015.

[Wal74] Immanuel WALLERSTEIN : The rise and future demise of the capitalist world system. *Comparative Studies in Society and History*, 16(4):387–415, 1974.

[WD09] Xingqi WANG et Guang DONG : Research on money laundering detection based on improved minimum spanning tree clustering and its application. *In 2009 Second international symposium on knowledge acquisition and modeling*, volume 2, pages 62–64. IEEE, 2009.

[WHJ+09] Christopher WHITROW, David J HAND, Piotr JUSZCZAK, David WESTON et Niall M ADAMS : Transaction aggregation as a strategy for credit card fraud detection. *Data mining and knowledge discovery*, 18:30–55, 2009.

[WM97] D Randall WILSON et Tony R MARTINEZ : Improved heterogeneous distance functions. *Journal of artificial intelligence research*, 6:1–34, 1997.

[WP03] Xuechuan WANG et Kuldip K PALIWAL : Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern recognition*, 36(10):2429–2439, 2003.

[Zhe15] A. ZHENG : *Evaluating Machine Learning Models : A Beginner's Guide to Key Concepts and Pitfalls*. O'Reilly Media, 2015.

[ZS+15] Masoumeh ZAREAPOOR, Pourya SHAMSOLMOALI *et al.* : Application of credit card fraud detection : Based on bagging ensemble classifier. *Procedia computer science*, 48(2015):679–685, 2015.

[ZZJ19] Junyi ZOU, Jinliang ZHANG et Ping JIANG : Credit card fraud detection using autoencoder neural network. *arXiv preprint arXiv :1908.11553*, 2019.