REPUBLIQUE DU SENEGAL

Un Peuple - Un But - Une Foi

Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation

Direction de l'Enseignement Supérieur Privé

Institut Supérieur d'Informatique

ISI

Mémoire de fin de Cycle pour l'obtention du master professionnel en Réseaux et systèmes informatiques

Etude et mise en place d'un système basé sur le machine learning pour la détection de fraudes monétiques

Présenté et soutenu par :

M. William N. TOLOFON

Sous la direction de

M. Lamine DIENG

Spécialité : Ingénieur en virtualisation et cloud

Année Académique : 2018 -2019

Dédicace

Je dédie ce mémoire, fruit de mon travail, à mes parents pour les sacrifices et soutiens qu'ils ont consentis à mon égard pour me garantir un avenir radieux.

Remerciements

L'écriture d'une telle page n'est pas aisée. Bien-sûr, des noms viennent en tête immédiatement à l'esprit ; mais combien d'anonymes ou d'oubliées ont directement ou indirectement contribué à ce mémoire ? Difficile de le dire, d'autant plus qu'en la matière, la subjectivité est reine. Que ceux et celles que nous n'aurons pas cités reçoivent l'expression de nos pensées.

Nous remercions, tout d'abord Dieu pour toutes les grâces dont Il ne cesse de nous combler.

Nous tenons à remercier ensuite tous nos enseignants de l'Institut Supérieur d'Informatique pour la qualité du savoir qu'ils nous ont transmis.

Nous exprimons notre profonde gratitude et nos sincères remerciements à toutes les personnes qui ont contribué à rendre notre travail meilleur notamment :

- ✓ Notre Maître de mémoire
- ✓ Mme Cissé pour ses précieux conseils
- ✓ Aux membres du jury qui nous font l'honneur d'apprécier notre travail
- ✓ A M. Nassirou YOUNGA, pour son ouverture d'esprit, et sa disponibilité inconditionnelle
- ✓ A M. Rody KINGNIDE, M. Eden AGBOGBA, M. Noé KODJO pour leur aide, soutien et disponibilité
- ✓ A M. Julien Ethè BENISSAN-MESSAN, Mme. Anna TINE, M. Beau-Gars AWI pour leur soutien moral et psychologique
- ✓ A mes camarades de classe, en particulier El-Hassane TCHA-COROUDOU, Sall NDIAYE pour leur aide et leur soutien
- ✓ Toutes les personnes qui ont contribué de près ou de loin à la rédaction de ce mémoire

Merci à tous et à toutes

Avant-propos

L'Institut Supérieur d'informatique (ISI) est un institut d'enseignement supérieur, créé en 1991, spécialisé dans la formation continue de type académique et particulièrement dans des domaines tels que l'informatique, la gestion, la comptabilité et l'organisation des entreprises. Il s'active dans la formation académique afin de participer à la politique mise en place par le gouvernement pour le développement dans le secteur de l'éducation et de la formation. Les programmes académiques offerts par ISI sont complets et intégrés au système LMD.

Ce travail a été réalisée dans le but d'adhérer à la règle selon laquelle tout étudiant en fin de cycle doit rédiger un mémoire de fin d'étude et ce conformément aux normes académiques en vigueur en République Sénégalaise en général, et à ISI en particulier. Non seulement, il se conforme aux règles, mais il répond à un important problème de notre société. Cette œuvre est intitulée **Etude et mise en place d'un système basé sur le machine learning pour la détection de fraudes monétiques**. Elle met l'accent sur le sérieux problème de l'accroissement des fraudes bancaires.

On assiste à une utilisation accrue des cartes de crédit, et à une recrudescence des paiements électroniques avec le développement d'internet ; et ceci se fait remarquer autant du côté des entreprises que des particuliers. Cette recrudescence observée présente beaucoup d'avantages ; mais elle présente aussi des inconvénients. En réalité, on constate les fraudes monétiques qui ne cessent de croitre malgré les moyens mis pour faire face à la complexité et de la diversité des attaques. En effet, il existe des approches de détections de fraudes mais qui semblent être classique à nos jours. Ce type d'approche est principalement fondé sur l'application de règles préétablies, simples ou avancées basé sur des fraudes connues. Ces approches de solutions ont pu prouvé, et prouvent encore leur efficacité dans la détection de fraudes usuelles. Mais de nos jours, la diversité et la complexité des attaques s'amplifient. Et l'effet prévisible de ces solutions font qu'elles peuvent être contournées.

Nous sollicitons de la part des membres du jury leur indulgence quant à l'évaluation.

Sommaire

Glossaire

ACS : Serveur gérant la sécurité des paiements 3D Secure

ATM : Automated Teller Machine

BCEAO : Banque Centrale des Etats de l'Afrique de l'Ouest

BIN : Bank Identification Number

CEDEAO : Communauté des Etats de l'Afrique de l'Ouest

CIP-UEMOA : Centrale des Incidents de Paiement de l'UEMOA

DAB : Distributeur Automatique de Billets

GAB : Guichet Automatique de Banque

IA : Intelligence artificielle

IBAN : International Bank Account Number

ISI : Institut Supérieur d'Informatique

MPI : Message Passing Interface

SVM : Support Vector Machine

TIC : Technologie de l'information et de la communication

TPE : Terminal de paiement électronique

UEMOA : Union Economique et Monétaire Ouest Africaine

Liste des figures

Liste des tableaux

Introduction

Les progrès continus des Nouvelles Technologies de l'Information et de la Communication définissent, à nos jours, un nouveau style de vie. Les nouvelles technologies sont devenues incontournables de par leur intervention dans tous les secteurs d'activités que ce soit la science, le commerce, ou encore la monétique. Ce nouveau style de vie qui commence à s'ancrer dans notre quotidien nous pousse à prendre de nouvelles habitudes. Le secteur de la monétique, qui est caractérisé par l'utilisation des cartes bancaires, est encore loin d'être à son apogée. Ceci est dû au développement des TICS et à l'explosion d'internet. Son impact est sans appel sur notre quotidien, elle a modifié le comportement des usagers face aux transactions. Elle a permis une certaine facilité dans les transactions, et s'est révélé être un super gagne-temps pour les utilisateurs.

On assiste désormais à une utilisation accrue des cartes de crédit, et à une recrudescence des paiements électroniques avec le développement d'internet ; et ceci se fait remarquer autant du côté des entreprises que des particuliers. Cette recrudescence observée présente beaucoup d'avantages et montre combien le développement est effectif, et encore à son aurore ; mais elle présente aussi des inconvénients. L'un de ces inconvénients est l'augmentation des fraudes monétiques. Et malgré, les moyens mises en place qui semblent être usitées à nos jours, elle ne cesse d'augmenter, du fait que les fraudeurs utilisent de nouvelles méthodes chaque jour. Comment anticiper sur ces nouvelles attaques ? Quelle nouvelle approche de détection doit-on mettre en place pour une détection proactive ?

C'est dans ce cadre que s'inscrit notre thème, intitulé « étude et mise en place d'un système basé sur le machine learning pour la détection de fraudes monétiques » ; ce qui dévoilera le rôle important de l'intelligence artificielle dans la gestion des fraudes. En effet, elle permettra, par exemple, aux banques de se rapprocher du parfait équilibre entre protection et satisfaction des clients, aussi elle pourrait révolutionner le processus du « Know Your Custumer » pour les marchands.

En réalité, on constate de plus en plus de fraudes monétiques. Elles ne cesseront de croître du fait de l'utilisation par les fraudeurs des techniques complexes et de plus en plus sophistiquées. Il va sans dire qu'il existe des approches de détections de fraudes mais qui semblent parfois dépassées par rapport aux attaques auxquelles on fait face de nos jours. Ce

type d'approche est principalement fondé sur l'application de règles préétablies, simples ou avancées basé sur des fraudes connues. Ces approches de solutions ont su prouvé, et continuent de prouver leur efficacité dans la détection de fraudes usuelles. Mais de nos jours, la diversité et la complexité des attaques augmentent. L'effet prévisible de ces solutions fait qu'elles peuvent être contournées.

Pour remédier à cet effet prévisible des solutions, il faut adopter un nouveau regard sur les bases et stratégies de détection. Aujourd'hui, plutôt que d'utiliser des règles préétablies, on veut pouvoir détecter des comportements de fraudeurs en analysant les transactions monétiques, et ce à posteriori ou en temps réel. Ceci pourrait être possible grâce aux approches fondées sur des algorithmes et technologies d'analyse de grands volumes de données et traités à grande vitesse. Ainsi, on pourrait concevoir un profil de porteur en se basant sur les données qu'il produit.

Pour y arriver, nous nous sommes intéressées à la monétique plus précisément aux techniques de fraudes monétiques, et aux algorithmes du machine learning, afin de trouver l'algorithme qui pourrait nous aider au mieux dans la lutte contre ces fraudes.

Pour arriver à mieux cerner la problématique, nous avons subdivisé notre travail en trois grandes parties : dans la première partie, nous avons présenté le sujet et abordé les différents concepts généraux de notre travail, puis dans la deuxième partie, nous avons défini les fraudes, présenté les modèles mathématiques de machine learning, nous avons ensuite retenu un modèle selon un type de fraude donné, pour enfin dans la troisième partie, modéliser, réaliser le système et présenter des résultats de notre système.

I. Cadre méthodologique et théorique

I. Cadre Méthodologique et théorique

Cette partie met en exergue la problématique qui pèse sur la monétique, surtout sur les transactions monétiques, et elle présente les différents concepts pour mieux appréhender les différentes notions qui seront abordées tout au long du document.

1.1. Présentation du sujet

1.1.1. Contexte

Aujourd'hui, les Nouvelles Technologies ont joué un rôle important dans la révolution monétique. De ce fait, on constate une ouverture des économies, et une multiplication de services financiers qui fait intervenir beaucoup plus l'utilisation des cartes. En effet, de plus en plus, de nouveaux instruments de paiements sont offerts par les banques. Néanmoins, sans un certain sentiment de sécurité, qu'il soit physique ou virtuel, il ne peut réellement régner un climat de confiance dans la bonne tenue des affaires et transactions, c'est pour cela que ces nouveaux instruments doivent présenter un niveau de sécurité acceptable afin de garantir la protection des clients. C'est dans cette optique, que plusieurs mécanismes sont mises en place dans le but d'offrir ce niveau de sécurité recherché.

1.1.2. Problématique

La lutte contre la fraude représente un fléau croissant pour un bon nombre d'entreprises. Aujourd'hui, la vente sur internet, par exemple, qui connait une croissance forte est confrontée à la problématique de sécurisation des paiements ; aussi, en exemple, constate-t-on une augmentation de transactions par paiement électronique. Ce type de transactions repose sur des algorithmes complexes qui peuvent être embarqués sur des terminaux physiques, être liés à des applications. L'ensemble de ces transactions représente un marché qui s'appuie fortement sur la confiance que l'ensemble des acteurs (banques, marchands, clients) a dans la manière d'opérer ces systèmes.

Parallèlement à cette hausse du volume de transactions, on assiste à un accroissement des cas de fraudes. Et pourtant, il existe un bon nombre de solutions de sécurisation. Le principal hic reste la complexité de mise en œuvre qui constitue un obstacle. Les commerçants préfèrent à ce stade, des moyens de paiements simples et moins sécurisés, au risque de supporter la fraude

plutôt que des solutions de paiements plus robustes (authentification tripartite (3DS))[1] qui leur font perdre du chiffre d'affaire car demandant un nombre de clics aux utilisateurs, incompatible avec l'achat pulsionnel. Dans un autre registre, afin d'accroitre la concurrence et donc de baisser les prix, il a été récemment décidé de simplifier les procédures permettant la réalisation des virements entre comptes bancaires de particuliers, au risque de faciliter la tâche du fraudeur[2].

De manière générale, plus la solution de paiement est généralisée et donc utilisée par un grand nombre d'individus, plus elle est exposée. Par exemple, Windows est plus vulnérable que MacOS car plus attaqué. Les schémas de fraude classiques simple à mettre en œuvre, sont facilement contournables, il suffit d'en identifier les seuils ou paliers paramétrés. L'effet prévisible à la mise en place de ces règles est une complexification du comportement du fraudeur qui tend à ressembler à « Mr tout le monde ». En effet, si vous bloquez, par exemple, les transactions d'un montant élevé, vous aurez un nombre plus important de transactions avec un montant plus faible. La question revient alors à savoir, comment détecter, reconnaitre et arrêter une fraude, ou comment évaluer un risque de fraude possible, et ce, à postériori ou en temps réel (qui est préférable).

C'est dans ce cadre que l'analyse statistique prend tout son sens. L'objectif est d'identifier des comportements atypiques par la combinaison optimisée d'un ensemble de données et non plus d'une seule tel que le montant d'une transaction dans l'exemple précédent.

1.1.3. Objectifs

Objectif général

Le principal but de notre travail est de montrer les possibilités et apports que nous offre le machine Learning, dans la monétique, pour aider dans la détection de fraudes monétiques.

[1] *La technologie 3-D Secure permet de transmettre au serveur d'authentification de la banque du porteur, une demande d'authentification du porteur de carte. Il s'agit de s'assurer lors de chaque paiement que la carte est bien utilisée par son titulaire. L'authentification du porteur se fait par la saisie d'un identifiant personnel en plus de l'authentification habituelle*

[2] *Les statistiques au service de la lutte contre la fraude, Yves Péchiné*

Pour ce faire, nous mettrons en place un système basé sur le machine learning pour améliorer le système de fraudes existant

Objectifs spécifiques

Sur la base de l'objectif général défini ci-dessus, nous avons pu avoir quelques objectifs spécifiques de notre projet que voici :

- Réduire le temps d'analyse des relevés de transactions en optimisant l'analyse
- Mettre en place un système de détection moderne, flexible et modifiable
- Modéliser un profil d'utilisateur en se basant sur ces habitudes

1.2. Hypothèses et Approches de solutions

1.2.1. Hypothèses

L'hypothèse principale derrière notre approche est qu'en utilisant des exemples de fraude fournis, une méthode supervisée pourra généraliser à de nouvelles données et aussi, en se basant sur les comportements des porteurs de carte, on pourra plus facilement détecter des comportements anormaux.

Une autre hypothèse est, qu'avec notre système, on pourrait optimiser l'analyse des transactions, ce qui réduirait le temps d'analyse

1.2.2. Etat de l'art

Aujourd'hui, les Nouvelles Technologies ont un rôle important dans la révolution monétique. De ce fait, on constate une ouverture des économies, et une multiplication de services financiers qui font intervenir beaucoup plus l'utilisation des cartes. De plus en plus, de nouveaux instruments de paiements sont offerts par les banques, et parmi celles-ci, il y en a, qui permettent d'effectuer des transactions sur Internet. Néanmoins, sans un certain sentiment de sécurité, qu'il soit physique ou virtuel, il ne peut réellement régner un climat de confiance dans la bonne tenue des affaires et transactions, c'est pour cela que ces nouveaux instruments doivent présenter un niveau de sécurité acceptable afin de garantir la protection des clients.

De nombreux outils et méthodes permettent de garantir ce niveau de sécurité pour les nouveaux systèmes de paiements. La base commune à ces méthodes est la mise en place d'un système de sécurité basé sur un processus cyclique d'identifications, d'évaluations et de traitements de risques[3], l'objectif étant d'identifier les fraudes, et de mener des actions pour les contrer.

[3] *Le risque ici est considéré comme la fraude. Selon la thèse de doctorat intitulé « Sécurisation d'un système de transactions sur terminaux mobiles », la fraude fait partie de la catégorie du risque opérationnel qui correspond*

L'approche traditionnelle pour la protection des clients repose sur des solutions de protection multiples telles que l'authentification (simple ou à double facteur) qui constitue une première couche de sécurité essentielle. L'authentification à double facteur[4] offre des gages de sécurité bien plus importants que les autres types d'authentification, son utilisation se démocratise largement, notamment pour les paiements et les opérations sensibles. Pour autant, l'authentification étant bien souvent le premier niveau de sécurité rencontré par un client, il est aussi le premier à être attaqué. Des moyens de protection additionnels existent (limitation des options proposées, temporisation des opérations…), mais ils dégraderaient l'expérience client, ce qui va à l'encontre de la simplification et de la fluidification des parcours client recherchées par les entreprises.

Dans le souci de lutter efficacement contre les fraudes et de renforcer la confiance des clients dans les instruments de paiement scripturaux-chèques, cartes bancaires, lettres de change- dans l'espace UEMOA, la BCEAO a mis en place la Centrale des incidents de paiement de l'UEMOA (CIP-UEMOA). Une démarche qui est aujourd'hui appuyée par des institutions comme GIM-UEMOA. Parmi les instruments régionaux dédiés à cette lutte figurent la directive sur la lutte contre la cybercriminalité dans l'espace de la CEDEAO et la convention de l'Union africaine sur la cyber sécurité et la protection des données à caractère personnel, ainsi que des approches de détection de fraudes.

L'approche classique de détection de fraude, largement répandue aujourd'hui, consiste à détecter des schémas de fraude déjà rencontrés par le passé (Les schémas de fraude classiques font l'objets de règles simples du type « si . . . alors . . . »). Cette approche est principalement fondée sur l'application de règles préétablies, simples ou avancées, sur le flux de transactions :

- La détection unitaire, qui consiste à définir une règle métier où le non-respect d'un critère peut générer une alerte (virement vers un compte/IBAN sous surveillance…)

au risque de subir des pertes suite à un dysfonctionnement dans des processus internes, dans le système, les composants techniques ou dans des facteurs externes

[4] *L'authentification à double facteur est un processus de sécurité par lequel l'utilisateur fournit deux modes d'identification à partir de catégories de données distinctes : l'une se présente généralement sous la forme d'un jeton physique, comme une carte, et l'autre sous forme d'informations mémorisées, par exemple un code de sécurité*

- La corrélation d'événements, qui consiste à mettre en œuvre des règles métiers plus avancées corrélant plusieurs types de données (réalisation d'une opération depuis un pays sous surveillance et dépassement d'un seuil cumulé sur 24h…)

Dans certains contextes où la menace n'est pas prépondérante et où le risque de subir une fraude sophistiquée est faible, ces solutions peuvent être suffisantes et ont prouvé leur efficacité dans les situations de fraudes usuelles. Dans le cas contraire, ces solutions peuvent être très vite dépassées…

Une nouvelle approche de détection existe néanmoins, elle fait intervenir des méthodes statistiques, l'objectif étant de détecter à postériori ou en temps réel des comportements de fraudeurs en analysant les statistiques des transactions. Ces nouvelles solutions de fraude par l'analyse statistique offrent des avantages non négligeables dans la détection de fraudes. C'est dans cette optique que s'inscrira notre recherche pour trouver et adapter une solution basée sur des analyses statistiques afin de renforcer le domaine de la détection de fraudes dans la zone UEMOA.

1.2.3. Approches de solution

L'objectif est de pouvoir détecter à posteriori ou « en temps réel » des comportements de fraudeurs. Une démarche, relativement nouvelle, est possible grâce notamment à l'augmentation des capacités de calcul des ordinateurs, au développement de solutions logicielles adaptées et à l'utilisation des algorithmes statistiques de modélisation à ce jour très efficaces. Ces nouvelles solutions de détection de fraude par l'analyse statistique des transactions offrent des avantages non négligeables. Il s'agit tout d'abord de solutions peu couteuses à mettre en place. La mise en place d'outils décisionnels est infiniment moins couteuse que mettre à jour ou changer les algorithmes de sécurité d'un parc de millions de cartes bancaires. La réactivité est également un des avantages des méthodes statistiques. Le recours à de nouvelles règles, modèles statistiques ou stratégies de détection des transactions de fraude est beaucoup plus rapide, quelques jours voire quelques heures, que la mise au point d'une nouvelle génération de puces et son déploiement sur l'ensemble d'un parc de cartes et terminaux de paiement, de l'ordre d'une dizaine d'années.

Il existe à nos jours plusieurs moyens qui nous permettront de mettre en place ce type de solutions. Plusieurs choix s'offrent à nous dont nous pouvons retenir :

Tableau 1 : Comparaison entre les solutions

Solution	Avantages	Inconvénients	Décisions
Solutions Propriétaires	Complètes et fiables Sécurisés Facile d'utilisation	- Non modifiables - Payantes - Code source non disponible - Modules complémentaires et mise à jour payants - Dépendance de l'entreprise	Non adapté
Solution de conception native	Interface unique Maitrise absolue du code source Possibilités de modifier le code source à volonté Outil unique selon les besoins de l' entreprise Gestion de la sécurité	Nécessité de mise à jour régulières jusqu'à stabilisation	Adapté

La solution de conception native a été celle retenue à cause des avantages qu'elle nous offre.

1.3. Généralités sur la monétique et le machine learning

1.3.1. Généralités sur le machine learning

1.3.1.1. Définition

La monétique désigne l'ensemble des traitements électroniques, informatiques et télématiques nécessaires à la gestion de cartes bancaires ainsi que des transactions associées. Elle peut être aussi définie comme la gestion automatique (électronique) de la monnaie.

Elle a pour rôle :

- De faire des cartes bancaires un moyen de paiement ;
- De fiabiliser les services des distributeurs de billets.

Elle permet tout aussi de :

- Mettre un client en relation avec sa banque peu importe sa localisation ;
- Réduire les risques liés à la manipulation de la monnaie physique (perte d'argent, vol, etc. …).

1.3.1.2. Composants de la monétique

Il existe principalement deux composants dans la monétique :

- Le support

Le support est tout moyen de paiement ou d'encaissement présenté sous forme de carte plastique, équipée d'une bande magnétique et éventuellement d'une puce électronique. Il existe plusieurs sortes de carte, en fonction de leur vocation.

La carte bancaire possède un numéro qui est variable, mais en standard est de seize (16) caractères.

- Les six (06) premiers chiffres représentent le BIN : Bank Identification Number. Il identifie la banque qui a émis la carte. Une banque peut avoir plusieurs BIN. Une carte peut être adossée à zéro (0), un (01) ou plusieurs comptes bancaires ;
- Les chiffres suivants (neuf (09) à douze (12) chiffres) constituent l'identifiant de la carte chez la banque émettrice et sont attribués par la banque même ;
- Le dernier chiffre est une clé de contrôle permettant de vérifier que le numéro de la carte est conforme à la norme. Cette clé de contrôle est calculé selon une formule appelé formule de Luhn.

La puce électronique d'une carte est un élément clé de la sécurité. L'augmentation des capacités de calcul et de stockage de la puce a permis d'y stocker davantage d'informations et de programmes. C'est un circuit intégré qui se compose d':

- Un microprocesseur. Le microprocesseur permet d'effectuer des chiffrements lorsque c'est nécessaire ;
- Une mémoire. La mémoire de la carte est divisée en deux parties : une partie en lecture publique, les données classiques (prénom, nom, numéro de carte, etc.) sont présentes dans la partie publique sous forme claire et sous forme chiffrée ; et une partie en lecture cachée, la partie illisible contient la clé privée de la carte à laquelle il est théoriquement impossible d'accéder. Seule la carte peut la lire.

La piste magnétique permet de stocker des informations sur la carte et l'identité du titulaire visible au recto et le code confidentiel sous forme encodée ou cryptée. La sécurité de ces informations est assez limitée, puisqu'elles ne sont pas protégées, ni en écriture ni en lecture.

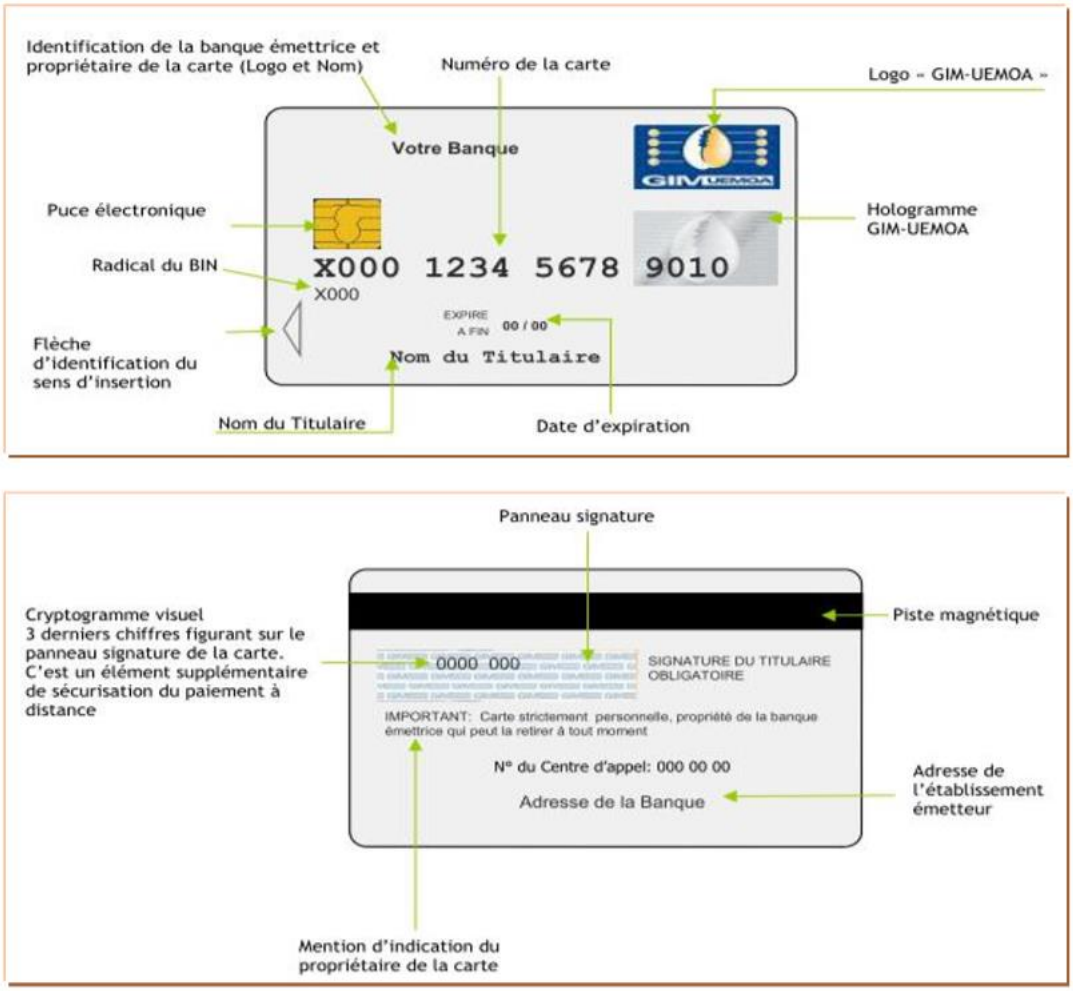

Figure 1 :la carte bancaire

- Le système de traitement

Ce sont des appareils électroniques permettant de lire les informations contenues dans les différents supports de la monétique. Ils sont généralement connectés à un centre de gestion des comptes des utilisateurs. Il en existe plusieurs dont :

- Les automates bancaires (DAB/GAB) : ce sont des appareils électromécaniques et électronique situés soit à l'intérieur ou à l'extérieur d'une banque ou même dans des lieux publics. La fonction principale d'un automate bancaire est de permettre de réaliser des opérations bancaires de base telles que les retraits d'espèces, les dépôts, des transferts de fonds, l'impression de relevés de compte. Les Distributeurs Automatiques de Billets (DAB) sont des appareils permettant de retirer une somme d'argent de son compte avec ou sans carte et d'un code confidentiel, dans les limites fixées contractuellement au préalable tandis que les Guichets Automatique de Banque (GAB) ou encore Automated Teller Machine

(ATM en anglais) disposent des fonctionnalités d'un DAB, mais il permet aussi de réaliser d'autres opérations bancaires (consultation de solde, demande de chéquier, mini relevé, versements...).

- Les Terminaux de paiement électroniques : ce sont des appareils disposant de plusieurs fonctions dont la lecture de données contenues dans une carte bancaire, la demande d'autorisation de transaction au serveur distant d'un émetteur, le stockage des transactions effectuées.

1.3.1.3. Les acteurs de la monétique

Le tableau ci-dessous comprend les principaux acteurs intervenants dans une transaction financière en monétique et leurs rôles.

Tableau 2 : Les acteurs de la monétique

Acteur	Fonction
Le porteur (client)	C'est celui à qui l'établissement financier remet la carte bancaire qu'elle a émise. Il doit souscrire à un contrat porteur carte bancaire où sont indiquées les conditions générales portant, entre autres, sur la délivrance, l'utilisation, la sécurité et le renouvellement de la carte
L'émetteur (Banque)	Il crée la carte et la met à disposition de son client. Il reste propriétaire de la carte. C'est généralement une banque ou un établissement financier. Quelques fonctions de l'émetteur : - Gestion du contrat et des relations avec le porteur - Gestion de la fabrication de carte

	- Traitement des autorisations - Traitement des oppositions - Recouvrements des litiges - Lutte contre la fraude
L'accepteur (commerçant)	L'accepteur est le professionnel qui accepte l'utilisation d'une carte bancaire pour le paiement d'un produit ou d'un service. Il est lié à une banque appelé acquéreur. Il doit respecter ses engagements vis-à-vis de sa banque et s'assurer de la régularité des paiements par carte
L'acquéreur (Banque)	C'est la structure gérant de l'accepteur. L'acquéreur est la banque qui met à disposition des machines ou dispositifs qui remplissent deux fonctions : a. Permettre au porteur de réaliser des transactions électroniques (après introduction de sa carte) b. Réaliser l'acquisition et le traitement de ces transactions quand elles sont finalisées Les principales responsabilités de l'acquéreur sont les suivantes : - Gestion des contrats et des relations avec les accepteurs - Gestion du matériel d'acceptation - Acquisition et traitement des paiements/retraits

	- Traitement des litiges - Lutte contre la fraude - Etc.
L'organisme interbancaire (réseaux)	Il peut regrouper les banques au sein d'un même pays, d'une même région, ou des banques appartenant à des zones géographiques différents. La relation entre l'émetteur et l'organisme interbancaire est matérialisée par un contrat d'adhésion et de services ; idem pour la relation acquéreur et organisme interbancaire On dispose de réseaux comme : - Réseaux nationaux ou régionaux ; tel que GIM-UEMOA, GUINEE INTERBANCAIRE MONETIQUE, MONETIQUE TUNISIE, etc. - Réseaux internationaux ; tel que Visa International, MasterCard, Union pay International, etc.

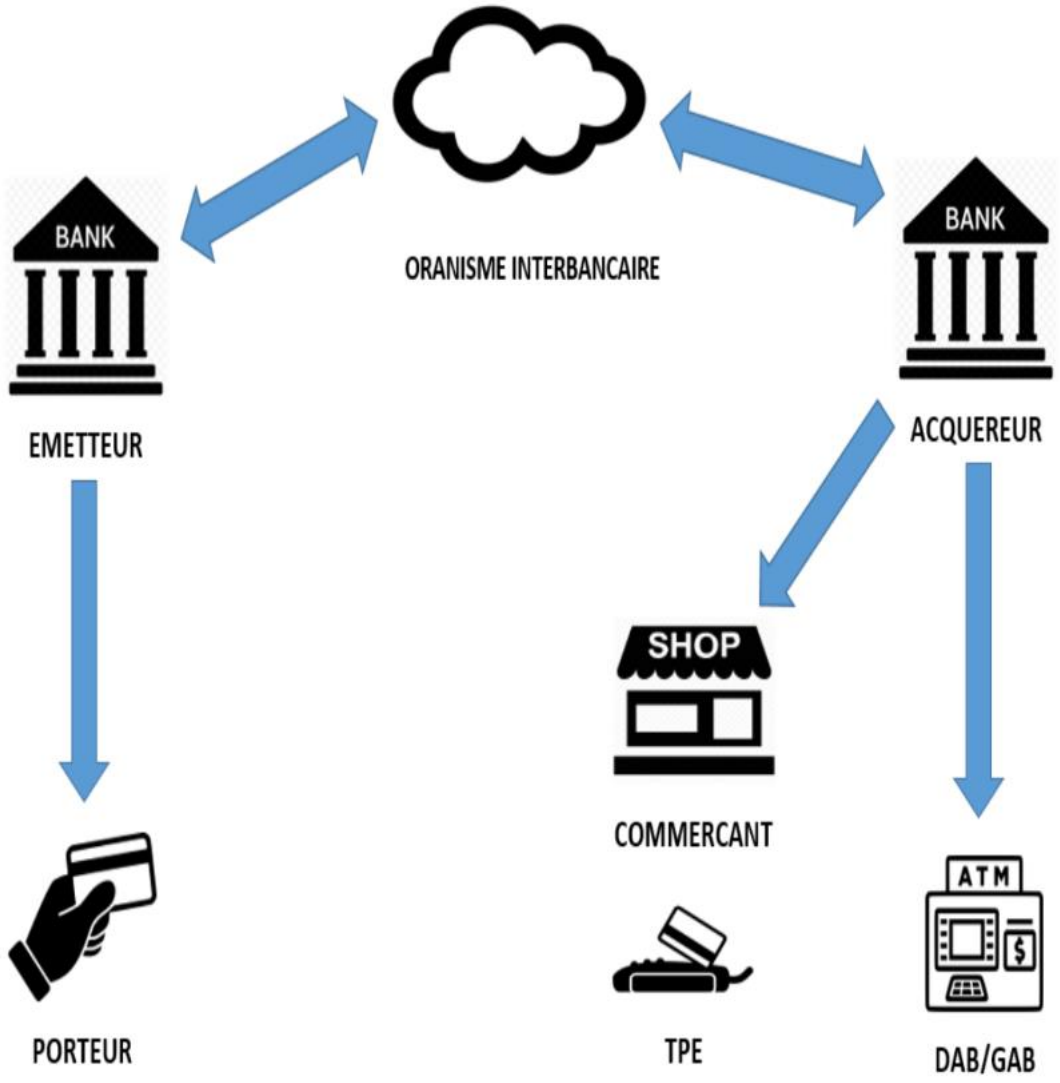

Figure 2 : Acteurs d'une transaction dans un réseau

1.3.1.4. Protocole de monétique ISO 8583

Le protocole ISO 8583 est une norme internationale relative aux spécifications d'échange de messages. Elle définit une interface commune par laquelle les acquéreurs et émetteurs de carte peuvent échanger des transactions financières initiées par carte. Elle précise la structure du message, son format et son contenu, les éléments d'information et les valeurs des éléments d'information. Les échanges onlines dans le cadre de la monétique s'effectuent généralement à l'aide du protocole ISO 8583. Il définit le format des messages ainsi que la cinématique des flux pour permettre à des systèmes distincts de communiquer. Le GIM-

UEMOA, VISA International, Mastercard International comme la plupart des systèmes monétiques dans le monde utilisent des protocoles basés sur l'ISO 8583.

1.3.2. Généralités sur le machine learning

Nous ne pouvons aborder le concept de machine learning sans parler au préalable de l'intelligence artificielle. Le machine learning est un sous-ensemble de l'intelligence artificielle. Qu'est-ce que l'intelligence artificielle (IA) ?

1.3.2.1. Définition de l'intelligence artificielle

La réponse à la question de savoir ce qu'est l'intelligence artificielle reste délicate. L'intelligence humaine étant en soit plutôt difficile à appréhender, que dire de celle d'entités artificielles.

L'intelligence artificielle (IA) est l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence. Elle est aussi définie par l'un de ses créateurs, Marvin Lee Minsky, comme : " la construction de programmes informatiques qui s'adonnent à des tâches qui sont pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique. ".

Il existe plusieurs domaines dans l'intelligence artificielle dont le machine learning qui sera le point à aborder dans la prochaine section.

1.3.2.2. Définition du machine learning

L'apprentissage automatique (en anglais machine learning) ou apprentissage statistique est un champ d'étude de l'intelligence artificielle qui se fonde sur des approches statistiques pour donner aux ordinateurs la capacité d' « apprendre » à partir de données. Il permet aux ordinateurs d'apprendre sans avoir été programmés explicitement à cet effet. Concrètement, il s'agit d'une science moderne permettant de découvrir des patterns et d'effectuer des prédictions

à partir de données en se basant sur des statistiques, sur du forage de données, sur la reconnaissance de patterns et sur les analyses prédictives. Le machine learning est une méthode d'analyse des données qui automatise la création de modèles analytiques. C'est une branche de l'intelligence artificielle qui repose sur l'idée que les systèmes peuvent apprendre des données, identifier des tendances et prendre des décisions avec un minimum d'intervention humaine. Pour apprendre et se développer, les ordinateurs ont toutefois besoin de données à analyser et sur lesquelles s'entraîner. De ce fait, le Big Data est essentiel au Machine Learning, et c'est la technologie qui permet d'exploiter pleinement le potentiel du Big Data.

Grâce aux nouvelles technologies informatiques, le machine learning a énormément progressé. Les chercheurs s'intéressant à l'intelligence artificielle voulaient en effet savoir si les ordinateurs étaient capables d'apprendre des données. La dimension itérative du machine learning est importante car les modèles s'adaptent d'eux-mêmes lorsqu'ils sont exposés à de nouvelles données. Ils apprennent de calculs précédents afin de produire des décisions et résultats fiables et reproductibles. La science n'est donc pas nouvelle, mais elle connaît un nouvel élan.

1.3.2.3. Intérêt et Utilisation du machine learning

L'intérêt pour le machine learning aujourd'hui s'explique par des facteurs qui sont également à l'origine du Big data :

- Multiplication et la diversification des données disponibles
- Puissance de calcul plus importante et moins coûteuse
- Stockage des données plus économique
- Etc.

Autant de facteurs qui permettent désormais de créer automatiquement des modèles capables d'analyser des données complexes et volumineuses, et d'obtenir rapidement des résultats précis, y compris à très grande échelle.

- Services financiers : Les banques et autres entreprises du secteur financier utilisent principalement le machine learning dans le but d'interpréter les données et de prévenir la

fraude. Les informations obtenues servent à identifier des opportunités d'investissement ou indiquent aux investisseurs le meilleur moment pour réaliser leurs opérations

- Administration : Les administrations en charge de la sécurité et des services publics ont particulièrement besoin du machine learning car elles ont accès à de nombreuses sources de données pouvant receler de précieuses informations. Ainsi, l'analyse des données des capteurs fait apparaître des pistes pour améliorer l'efficacité et réaliser des économies. Le machine learning permet également de détecter la fraude et d'éviter l'usurpation d'identité
- Santé : Le machine learning a le vent en poupe dans le secteur de la santé, grâce à l'avènement d'accessoires connectés et de capteurs qui évaluent en temps réel l'état de santé d'un patient. La technologie sert également à analyser les données pour dégager des tendances ou des indicateurs d'alerte permettant d'améliorer les diagnostics et les traitements
- Marketing et vente : Les recommandations d'achat des sites web utilisent le machine learning pour analyser votre historique d'achats et vous proposer des articles susceptibles de vous intéresser. L'avenir de la grande distribution réside dans cette capacité à capturer des données, à les analyser et à les mettre à profit pour personnaliser une expérience d'achat (ou déployer une campagne marketing)
- Energie : Trouver de nouvelles sources d'énergie. Analyser les minerais présents dans le sol. Prévoir les pannes des capteurs d'une raffinerie. Simplifier l'approvisionnement en pétrole pour optimiser l'efficacité et la rentabilité. Les cas d'usage dans ce secteur sont nombreux et ne cessent de se multiplier.
- Internet des objets connectés
- Etc.

Nous avons constaté que le phénomène de fraudes à la carte bancaire est bel et bien réel et ne cesse de prendre de l'ampleur malgré les moyens mis en place. C'est dans l'optique d'apporter une approche de solution que nous avons pensé à des solutions nouvelles basé sur le machine learning. Pour ce faire, le choix de la proposition d'approche de solutions pour laquelle on a opté, est de type native, dans le but de comprendre et de d'exposer les différentes options que peut nous fournir le machine learning dans notre lutte contre la fraude.

II. Cadre conceptuel

II. Cadre conceptuel : Choix d'une méthode de détection basée sur le machine learning

Il serait question ici d'aborder la fraude proprement dite. Nous allons exposer la fraude, et les différents algorithmes de machine learning pouvant nous aider à le résoudre.

2.1. Les fraudes monétiques

2.1.1. Définition

La fraude monétique se définie comme étant l'ensemble des achats effectués au moyen de payements électroniques, sans le consentement du titulaire du moyen de paiement. Elle se manifeste par des achats effectués en ligne sur Internet ou en magasin, notamment par le biais de cartes contrefaites clonées par les as du carding. La fraude à la carte bancaire explose depuis l'avènement des paiements électroniques et du développement d'Internet.

2.1.2. Types de fraudes

Plusieurs méthodes et moyens sont utilisées pour réaliser des fraudes. Autant ces méthodes sont diversifiées qu'il en existe de fraudes. Cependant, nous pouvons catégoriser ces fraudes de sortes de retrouver les différents types de fraudes. Nous pouvons citer trois (03) grandes catégories de fraudes :

- Les fraudes physiques
    - L'utilisation par un tiers de sa carte bancaire suite à une perte ou un vol
    - Le skimming : piratage des terminaux : cette copie est souvent réalisée grâce à un dispositif de copie placé sur un terminal de paiement ou un distributeur de billets.
    - Le card trapping : variante du skimming. Les fraudeurs ne copient pas seulement les données de la carte, mais font en sorte qu'elle reste bloqué dans le distributeur. Lorsque le porteur va demander de l'aide, ils récupèrent la carte.
    - Les racleurs de RAM où les fraudeurs installent un virus au niveau des GAB et/ou TPE
- Les fraudes en ligne
    - L'utilisation des identifiants correspondants à sa carte bancaire : suite à un paiement en ligne, des « pirates » récupèrent les identifiants de la carte et s'en servent pour effectuer des achats

- L'hameçonnage ou phishing : c'est une méthode souvent utilisée par des fraudeurs pour tenter de collecter des données sensibles. Le phishing prend souvent la forme d'un envoi de mail en usurpant l'identité d'une entreprise, le plus fréquemment des banques ou bien des sociétés habituées à recourir au prélèvement automatique ; dans ce faux mail, copie conforme du site original de la société, celle-ci vous informe qu'elle a besoin de mettre à jour diverses coordonnées (bancaires ou d'accès à votre messagerie).
- Moulinage : technique de fraude consistant à utiliser les règles, propres à un émetteur, pour générer des numéros de carte et effectuer des paiements.

- Les fraudes à l'exécution
  - Répudiation abusive : contestation par le porteur, de mauvaise foi, d'un ordre de paiement valide dont il est l'initiateur

2.1.3. Techniques de détection de fraudes

Plusieurs méthodes et moyens sont utilisées pour réaliser des fraudes. Autant ces méthodes sont diversifiées qu'il en existe de fraudes. Aussi, plusieurs techniques de détection de fraudes existent à nos jours pour contrer les différentes fraudes existantes.

L'approche traditionnelle pour la protection des clients repose sur des solutions de protection multiples telles que l'authentification (simple ou à double facteur) qui constitue une première couche de sécurité essentielle. L'authentification à double facteur offre des gages de sécurité bien plus importants que les autres types d'authentification, son utilisation se démocratise largement, notamment pour les paiements et les opérations sensibles. Pour autant, l'authentification étant bien souvent le premier niveau de sécurité rencontré par un client, il est aussi le premier à être attaqué. Des moyens de protection additionnels existent (limitation des options proposées, temporisation des opérations…), mais ils dégraderaient l'expérience client, ce qui va à l'encontre de la simplification et de la fluidification des parcours client recherchées par les entreprises.

L'approche classique de détection de fraude, largement répandue aujourd'hui, consiste à détecter des schémas de fraude déjà rencontrés par le passé (Les schémas de fraude classiques

font l'objets de règles simples du type « si . . . alors . . . »). Cette approche est principalement fondée sur l'application de règles préétablies, simples ou avancées, sur le flux de transactions :

- La détection unitaire, qui consiste à définir une règle métier où le non-respect d'un critère peut générer une alerte (virement vers un compte/IBAN sous surveillance…)
- La corrélation d'événements, qui consiste à mettre en œuvre des règles métiers plus avancées corrélant plusieurs types de données (réalisation d'une opération depuis un pays sous surveillance et dépassement d'un seuil cumulé sur 24h…)

Le système 3D Secure (3 Domain Secure) est un système de sécurisation des paiements en ligne, créé par les émetteurs internationaux Visa et MasterCard. Il s'agit d'un protocole de sécurité qui vise à prévenir la fraude dans les transactions en ligne effectuées par carte de paiement ou carte de crédit. Il permet d'assurer que c'est bien le porteur qui effectue le paiement sur internet. Le 3D Secure est aujourd'hui la méthode la plus moderne, mais aussi la plus répandue, pour la prévention de la fraude dans les paiements en ligne.

Un paiement sur internet nécessite généralement trois informations basiques, le numéro de carte bancaire, la date d'expiration ainsi que le cryptogramme visuel (code à 3 chiffres situé au dos de votre carte). Pourquoi l'utilisation du numéro de la carte, de la date d'expiration et du cryptogramme visuel ne suffisent plus aujourd'hui ? Parce que n'importe qui ayant acquis votre carte (peu importe la façon) pourra faire autant d'achat qu'il souhaite sur internet (jusqu'à atteindre votre plafond).

Le service 3D Secure n'a pas de caractère obligatoire, pour les cybercommerçants comme pour les banques.

L'authentification du payeur est un processus qui inclut trois parties, qui sont :

- La banque et le commerçant qui recevra directement les fonds versés – Le récepteur
- L'émetteur, autrement dit la banque qui a délivré la carte de paiement (comme Visa ou MasterCard par exemple) – L'acquéreur.

- Et ce qu'on appelle le « Interoperability Domain », c'est à dire l'infrastructure fournie par le système de cartes pour prendre en charge le protocole 3D Secure – L'intermédiaire.

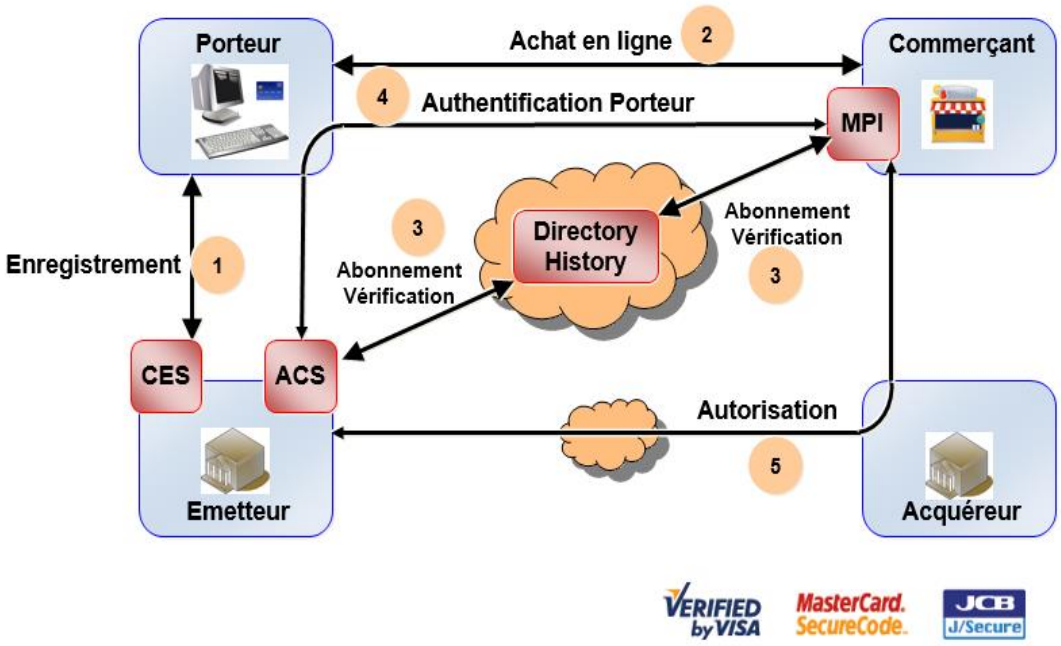

Figure 3 : Système 3D Secure

2.2. Les modèles mathématiques

2.2.1. Les types d'algorithmes

Ils existent plusieurs types d'algorithmes qu'utilise la machine learning pour son fonctionnement :

- Apprentissage supervisé :

Les algorithmes d'apprentissage supervisé sont entraînés sur des exemples étiquetés, par exemple une entrée dont le résultat attendu est connu. Un expert doit préalablement étiqueter des exemples. L'algorithme d'apprentissage reçoit une série de données en entrée avec les sorties correctes correspondantes, et apprend en comparant la sortie réelle avec les sorties correctes. Cette méthode d'apprentissage est couramment utilisée dans les applications où les données historiques servent à prévoir des événements futurs probables. Par exemple, elle permet de savoir dans quels cas des transactions de carte de crédit risquent d'être frauduleuses ou quel assuré est susceptible de soumettre une demande d'indemnisation.

- Apprentissage non-supervisé :

Les algorithmes d'apprentissage non-supervisé s'appliquent aux données sans étiquettes historiques. Dans ce cas, le système ne connaît pas la « bonne réponse ». L'algorithme s'applique dans ce cas à trouver seul les similarités et distinctions au sein de ces données, et à regrouper ensemble celles qui partagent des caractéristiques communes. C'est à l'algorithme de déterminer le modèle présenté. Le but est d'explorer les données et d'en découvrir la structure. Aucun expert n'est requis. L'algorithme doit découvrir par lui-même la structure plus ou moins cachée des données.

- Apprentissage semi-supervisé :

L'apprentissage semi-supervisé est comme l'apprentissage supervisé mais à la différence que son entraînement repose à la fois sur des données étiquetées et non étiquetées.

- Apprentissage par renforcement :

Il est souvent utilisé en robotique, dans les jeux vidéo et la navigation. Avec l'apprentissage par renforcement, l'algorithme découvre par tâtonnements les actions qui donnent les meilleurs résultats. Cette méthode vise à utiliser les observations recueillies lors de l'interaction avec l'environnement pour prendre des mesures qui maximiseraient les avantages ou réduiraient les risques.

2.2.2. Quelques exemples de modèles mathématiques

Il existe plusieurs modèles mathématiques, qui sont utilisés, de par les résultats attendus ou de par les données utilisées. Voici quelques-uns de ces modèles :

- Les modèles linéaires : c'est un modèle avec lequel on essaie de trouver une relation entre la variable à prédire et les variables qui décrivent nos données. C'est le cas, par exemple, d'un modèle qui prédit un loyer à partir de la surface de l'appartement.

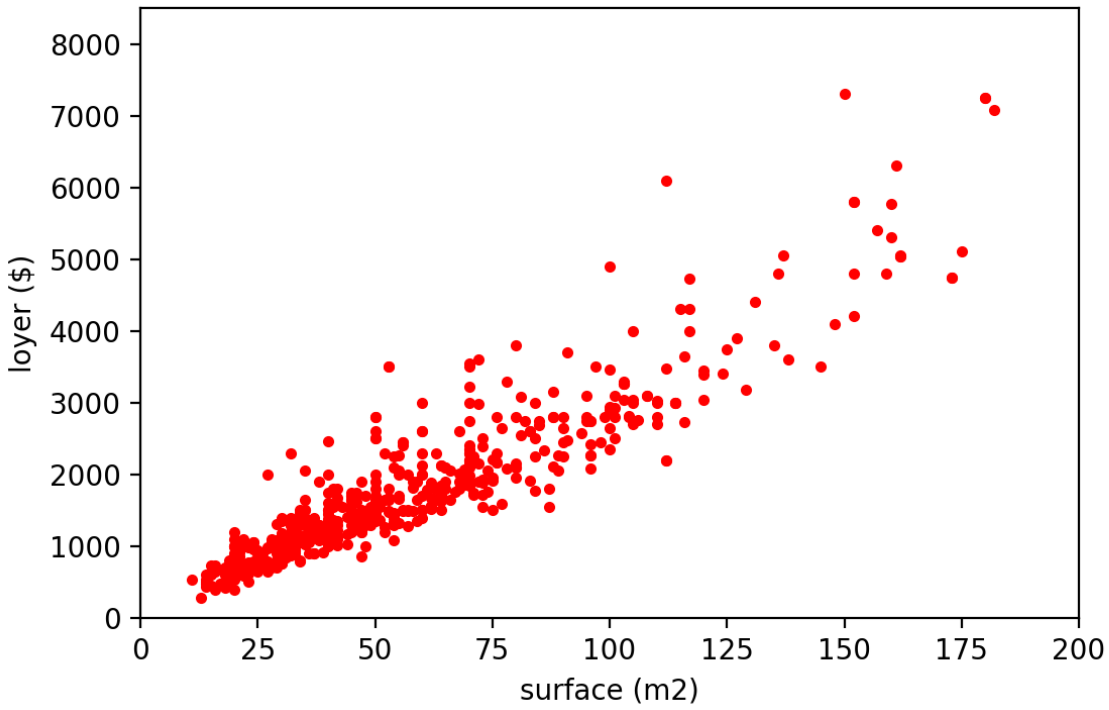

Figure 4 : Modèle linéaire

- La régression logistique est une méthode d'apprentissage automatique où la variable cible est catégorique. Dans le cas d'une classification binaire, la variable cible ne prend que deux valeurs, succès ou échec, généralement codées comme 0 ou 1. Des modèles de régression logistique binaire sont utilisés pour estimer la probabilité d'une réponse binaire sur la base d'un ensemble de variables indépendantes. La régression logistique simple est à bien des égards analogues à la régression linéaire, sauf que la variable cible est nominale et non une mesure. Divers domaines utilisent la régression logistique, en science médicale, la régression logistisque est couramment utilisé pour prédire les taux de mortalité des patients blessés ou pour prédire la probabilité qu'un patient ait un décès donné en fonction des caractéristiques observées du patient. Différente de la régression linéaire, où une équation linéaire est utilisée, la régression logistique peut être obtenue en utilisant la fonction sigmoïde,

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

  où le t représente toute valeur réelle entrée et la sortie est toujours des valeurs comprises entre 0 et 1 et peut être interprétée comme une probabilité.
- Les arbres de décisions : ils permettent de classifier des objets en effectuant des décisions successives sur la base de leur variable. Les nœuds de l'arbre représentent ces décisions alors que les feuilles représentent les valeurs de la variable cible (à prédire). La figure ci-dessous présente un exemple d'arbre de décision permettant de prédire si une personne possède ou pas un ordinateur

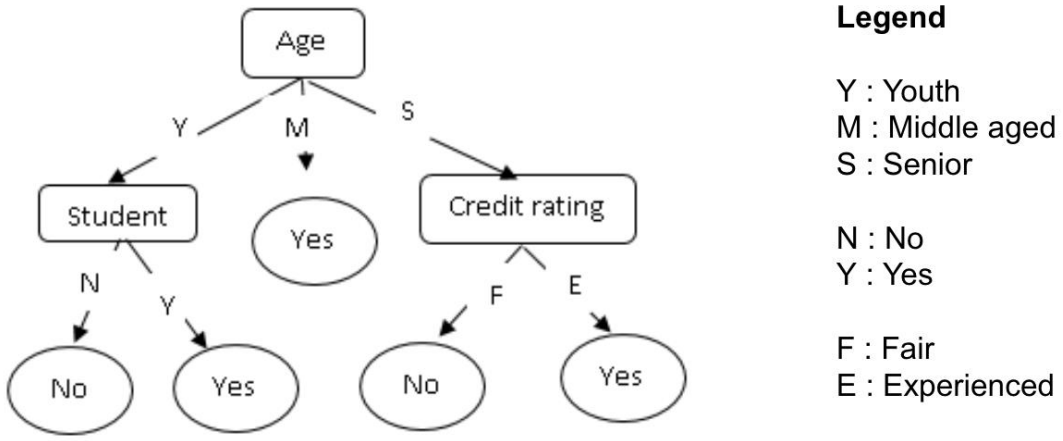

Figure 5: Arbre de décision

- Les SVMs (Support Vector Machines) : ils sont beaucoup plus utilisés sur des données avec beaucoup de variables. L'idée est de chercher un séparateur qui sépare au mieux les objets de chaque catégorie. En effet, il peut y avoir une infinité de séparateurs possibles. Le meilleur hyperplan selon les SVMs est celui qui maximise les marges avec les objets de chaque catégorie.

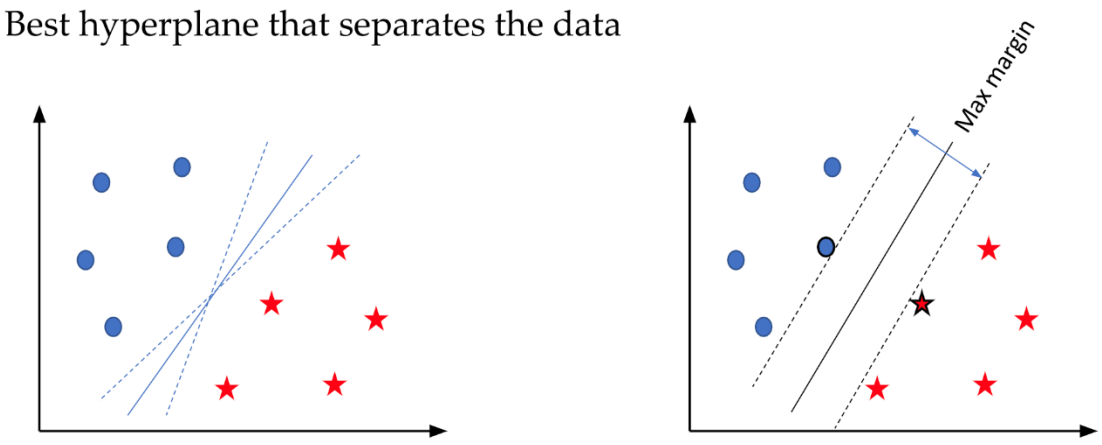

Figure 6 : Support Vector Machines

- Les méthodes ensemblistes : ils permettent de combiner plusieurs classifieurs afin d'améliorer les résultats. On peut en distinguer deux sous-familles :

- Le bagging : il permet d'entraîner plusieurs modèles (classifieurs) en parallèle pour ensuite les regrouper afin de prendre une décision. L'algorithme de bagging le plus connu est le Random Forest dont le principe est d'apprendre plusieurs arbres de manière indépendante sur des échantillons (avec replacement) des données d'apprentissage. La prédiction finale sera agrégée par un vote majoritaire (ou une moyenne des probabilités) issu des prédictions des différents arbres appris

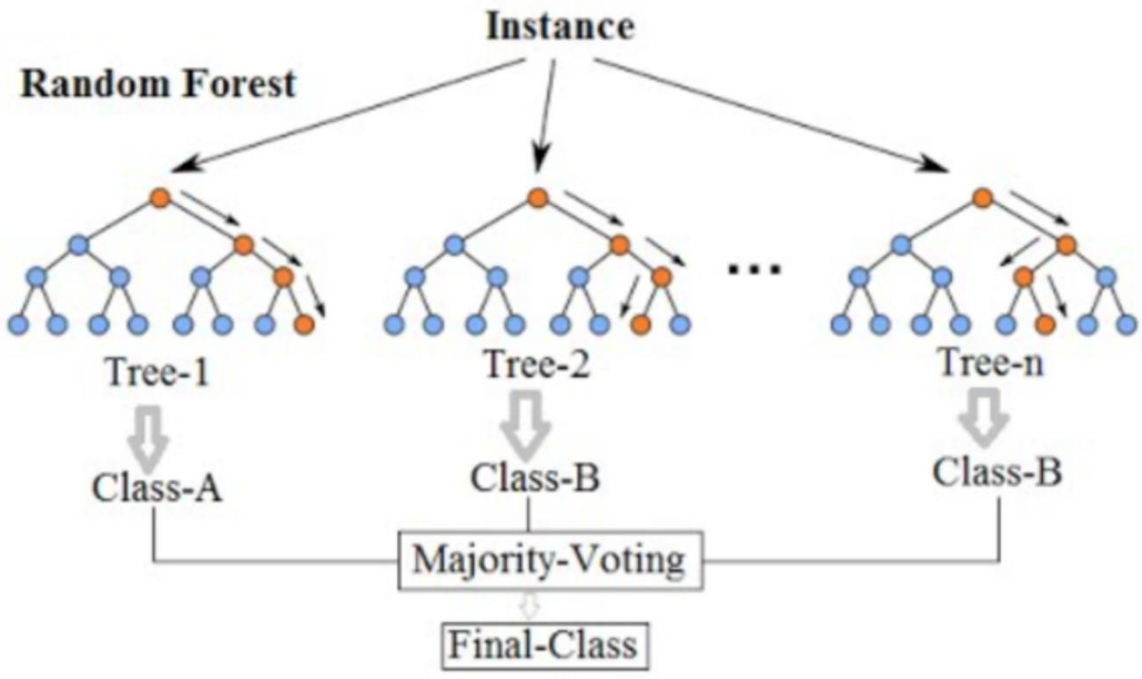

Figure 7: Méthode ensembliste : le boosting : random forest

- Le boosting : il consiste aussi à combiner des classificateurs mais de manière séquentielle. L'idée est que le classificateur numéro N se concentre sur les items mal classés par le classificateur numéro N-1. En effet chaque classificateur est entraîné sur un jeu de données pondéré qui accorde plus d'importance aux observations qui ont été mal classées par le classificateur précédent. La prédiction finale est aussi une combinaison pondérée des prédictions de tous les classificateurs appris en fonction de leur performance. Même si on peut théoriquement choisir n'importe quel type de classificateurs, les arbres de décision représentent le type le plus souvent utilisé.

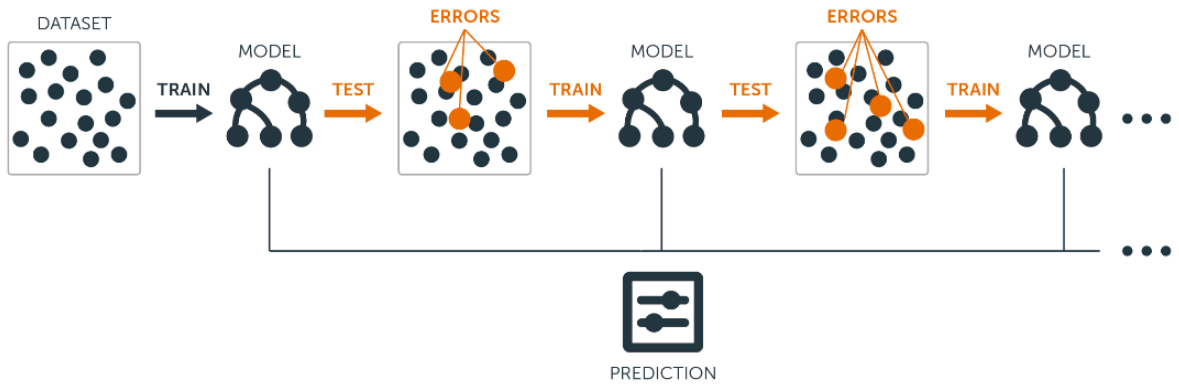

Figure 8 : Méthode ensembliste : le Boosting

- Les réseaux de neurones : un exemple de réseau de neurones est le multi-layer perceptron. Ce modèle est composé de plusieurs perceptrons organisés sous forme de couches. Chaque perceptron reçoit en entrée un certain nombre de valeurs à travers ses connexions entrantes et associe un poids à chacune de ces connexions. Le perceptron effectue une somme pondérée des valeurs en entrée à laquelle il applique une fonction d'activation. Le but de la fonction d'activation est de rajouter de la non-linéarité au modèle lui permettant d'apprendre des comportements plus complexes.

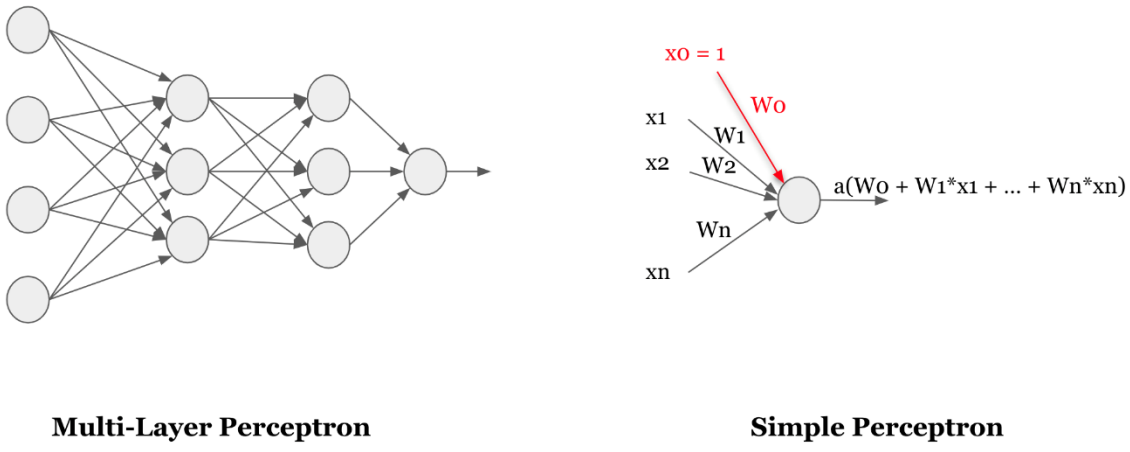

Figure 9:Exemple de réseau de neurones : Multi-Layer Perceptron

2.3. Choix d'un modèle appliqué à la fraude choisie

Précisons d'abord qu'un système de détection de fraude peut être source de deux types d'erreur. D'une part, il peut produire de fausses alertes sur des transactions légitimes – on parle alors de « faux positif » : le système détecte une fraude alors qu'il n'y en a pas – ce qui empêche un client de réaliser une transaction. D'autre part, il peut ne pas produire d'alerte sur un cas de fraude – on parle de « faux négatif ». Dans les deux cas, les mécanismes décisionnels du système de détection opèrent dans un contexte critique, avec un impact direct sur le client.

- Transactions de paiements

En considérant des données de transactions de paiement en ligne, nous pouvons dégager des caractéristiques spécifiques pouvant nous indiquer un choix d'algorithme donné. Par essence, les données des transactions sont séquentielles et en grande quantité. Pour les traiter, le modèle choisi doit pouvoir utiliser une infrastructure dotée d'une mémoire interne conséquente et capable d'intégrer et de gérer très rapidement de gros volumes de données. Par ailleurs, ces données sont « non-stationnaires », pour deux raisons : soit du fait que les fraudeurs adaptent leurs stratégies aux méthodes de détection, soit du fait que les habitudes d'achat des clients changent au cours de l'année (fêtes, évènements notables, soldes, vacances…). Ces deux sources de non-stationnarité n'ont pas le même impact sur les performances des algorithmes. La première augmente plutôt les faux négatifs car les nouveaux types de fraude ne sont pas encore détectés par le système, alors que la seconde augmente plutôt les faux positifs car les variations de comportements d'achat d'un client peuvent être considérées par le système comme un comportement inhabituel et donc à plus haut risque de fraude. Pour garantir les meilleures performances, les systèmes de détection de fraude reposent donc sur des algorithmes séquentiels qui ont l'avantage de s'adapter en temps réel.

La fraude est un évènement rare en proportion des volumes de transactions légitimes – l'ordre de grandeur est, habituellement, d'une transaction frauduleuse sur mille. Tant mieux ! Mais, du point de vue du concepteur du système de détection de fraudes, cela a pour conséquence que l'ensemble des données d'apprentissage utilisées pour construire les modèles de détection est très fortement déséquilibré : le faible ratio de paiements frauduleux en regard de la quantité de transactions rend la fraude difficile à prédire avec des algorithmes classiques

d'apprentissage. Deux solutions s'offrent au concepteur d'algorithmes : utiliser des méthodes de ré-échantillonnage des données pour réduire artificiellement ce déséquilibre afin de pouvoir utiliser des algorithmes d'apprentissage classiques ou garder l'ensemble d'apprentissage tel quel mais utiliser une famille d'algorithmes d'apprentissage capable de traiter ce type de données fortement déséquilibré.

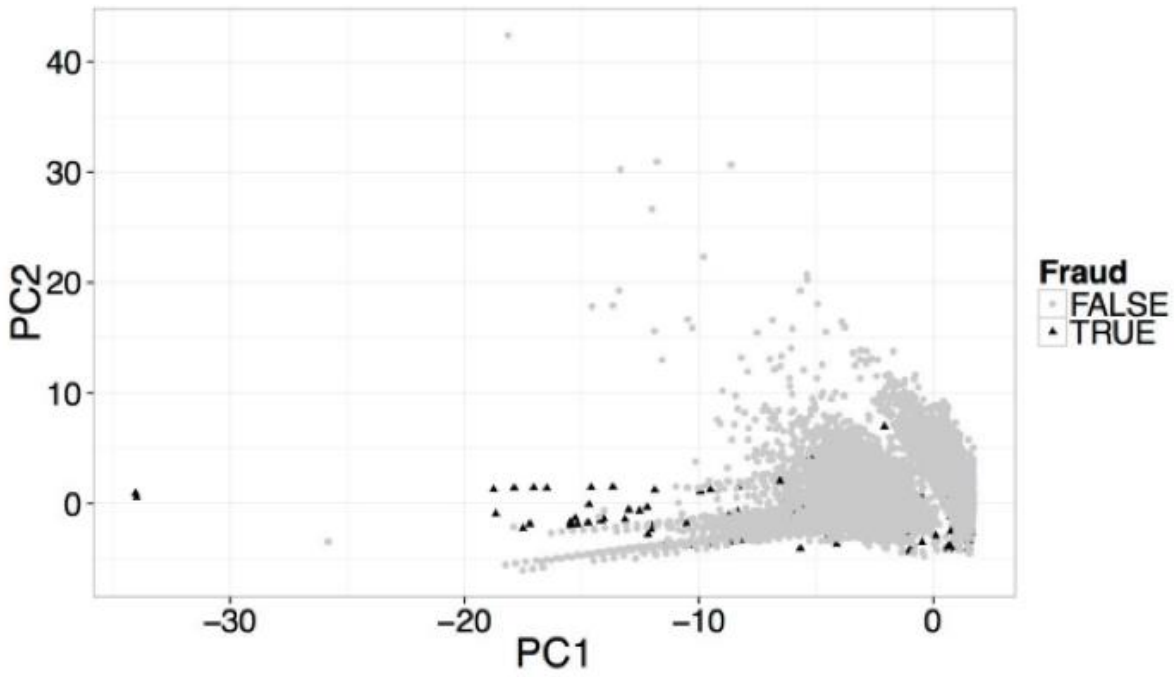

Figure 10: Exemple de graphe montrant des transactions frauduleuse et non frauduleuse

A l'aide de techniques de visualisation comme sur la figure ci-dessus, il est possible de mettre en lumière une dernière caractéristique des données de transaction de paiement qui complexifie parfois le problème : le « chevauchement » entre certains cas de fraudes et les transactions électroniques légitimes. Cela provient du fait que les fraudeurs, pour ne pas être détectés afin d'utiliser les informations de la carte volée le plus longtemps possible, s'efforcent d'adopter un comportement apparaissant comme « normal » du point de vue du système. Ce type de comportement demeure évidemment difficile à détecter. Pour y parvenir, le concepteur s'aide de méthodes telle que la création de nouvelles variables, qui permet, en quelques sortes, d'enrichir le contexte de nouvelles informations participant à l'identification de la fraude. Parmi les techniques permettant d'améliorer significativement les performances de détection, signalons des travaux de recherche récents sur l'injection d'informations de contexte

sémantique dans un modèle de détection de fraude ou la représentation des données de paiement sous la forme d'un graphe pour ensuite en extraire des variables supplémentaires.

Dans un contexte, où sont susceptibles d'apparaître des disruptions de comportement des porteurs, la balance penche plus pour la mise en œuvre des algorithmes d'apprentissage non supervisé, qui ont la capacité de « découvrir » des classements transaction frauduleuse / transaction légitime sans se baser sur des exemples de fraude passée. Malheureusement, ils sont aujourd'hui encore très difficiles à configurer car il existe une immense variété de comportements des porteurs de cartes. L'un des principaux défauts de ces algorithmes est qu'ils produisent des alertes dès qu'un individu semble avoir un comportement inhabituel, ce qu'il est très difficile à qualifier lorsque l'information collectée sur ces mêmes individus est faible. Ainsi, des milliers de "comportements inhabituels", mais néanmoins légitimes, se produisent chaque jour. Les algorithmes non supervisés produisent par conséquent énormément de faux positifs, ce qui les disqualifie en première approche pour un système de détection.

Sur la base de l'historique des transactions frauduleuses, les algorithmes d'apprentissage supervisé généralisent, quant à eux, un classement qui leur permet de considérer chaque nouvelle transaction comme légitime ou frauduleuse. Ils ne sont pas pour autant la panacée non plus. Pour comprendre pourquoi, nous devons nous demander comment est constitué cet historique, autrement dit, comment est détectée une fraude. En règle générale, il existe deux sources : soit des experts humains qui analysent les alertes manuellement (et trouvent de vrais positifs), soit un porteur de carte lorsqu'il découvre un usage frauduleux de sa carte (un faux négatif). Dans le premier cas, la correction de supervision est relativement rapide mais il faut que la transaction ait fait l'objet d'une alerte. Les détections frauduleuses sont donc détectées mais seulement pour un sous-ensemble de transactions (on parle de jeu de données « semi-supervisé »). Dans le second cas, le temps d'identifier une transaction comme frauduleuse peut être plus long car il faut que le porteur de carte la détecte. Les algorithmes d'apprentissage doivent donc être capables de comprendre ces deux mécanismes : être en mesure de tirer profit de données semi-supervisées obtenues à court terme mais aussi des données supervisées obtenues à long terme, ce que les algorithmes supervisés classiques ne savent pas faire.

- Détection de fraudes dans le e-commerce

Un autre système est celui du scoring qui permet d'établir un score dans le but d'évaluer le risque pris par le commerçant en fonction du produit ou du service acheté ainsi que du moyen de payement utilisé par le client. Cette évaluation repose essentiellement sur des modèles de calculs complexes issus du système bancaire qui prennent en compte un ensemble de paramètres spécifiques à l'achat ainsi qu'au secteur d'activité de l'entreprise. Le score est établi dès que l'achat est effectué, et ce de manière instantanée, sans que le client en n'ait connaissance. Dans le cas d'un score élevé, les deux cas de figure les plus courant sont :

- La commande est automatiquement annulée (la vente est refusée par le commerçant)
- La commande aboutie mais peut subir des contrôles à posteriori (notamment par des analystes formés à la prévention de la fraude).

Ainsi le e-commerce doit incorporer un niveau de pertes financières difficiles à maîtriser ou à évaluer, que ce soit en amont par le blocage systématique de la transaction qui peuvent s'avérer légitimes soit en aval, sur l'ensemble des fraudes non détectées.

Le 3D Secure mis en place par les banques pour sécuriser les achats en ligne est un bon exemple : les contraintes sont importantes puisque le taux d'abandon de transaction est élevé du fait de la dépendance d'un SMS envoyé par la banque sur le numéro de téléphone lié au compte du client. Si ce dernier n'est pas à jour ou que le téléphone n'est pas opérationnel ou à portée de main de l'acheteur, ce dernier sera dans l'impossibilité de finaliser sa commande… Du coup, ce système n'est pas forcément bien vu par les e-commerçants. Il apporte toutefois une sécurité puisqu'il permet de vérifier que le client est bien le propriétaire de la carte bancaire avec laquelle il souhaite régler.

III. Cadre analytique et quelques recommandations

III. Conception et réalisation du système

Il s'agira pour nous d'évaluer et de concevoir un algorithme basé sur le scoring afin d'inmplémenter un système capable de détecter des fraudes à posteori ou en temps réel.

3.1. Conception et architecture logicielle

3.1.1. Scoring

L'évaluation des risques-clients (credit scoring en anglais) désigne aujourd'hui un ensemble d'outils financiers d'aide à la décision utilisés pour évaluer automatiquement (par un algorithme) la solvabilité d'un "tiers" ainsi que le risque de non-remboursement de prêts ou de traites d'assurance, de loyer, etc. Il peut-être aussi utilisé dans la détection de fraudes, en évaluant le risque qu'un achat soit frauduleux ou non, selon le type de client, ou selon d'autres paramètres qui peuvent être mises en jeu.

3.1.2. Les types de scoring

- Le score d'octroi évalue le risque de défaut de paiement pour décider d'accorder ou de refuser le crédit.
- Le score de comportement évalue le risque de défaut de paiement d'un client existant lors des prises de décisions concernant la gestion de son compte, telles que la limite de crédit, la gestion des dépassements de limite, l'offre de nouveaux produits et autres.
- Le score de recouvrement est utilisé dans les stratégies de recouvrement pour évaluer la probabilité du remboursement de la dette par le client

3.1.3. Choix du type de scoring et du modèle à utiliser

Plusieurs types de scores ou modèles peuvent être construits. Si nous nous positionnons au niveau de la carte comme unité statistique, chaque ligne de la base d'étude étant un numéro de carte, nous pouvons construire une classification des différents comportements de la carte. On utilisera alors des indicateurs de type Récence Fréquence Montant classiques puis des indicateurs de localisation géographique et type d'achat afin d'identifier les profils des porteurs de carte. Ce type de modèle peu réactif au niveau de la détection de fraude a surtout un intérêt pour mieux connaitre ses clients vis-à-vis de l'utilisation de leur carte bancaire.

Une seconde possibilité est de se positionner sur une granularité carte * mois. La carte sera présente autant de fois dans la base d'étude que de mois ou elle aura réalisé une transaction. Cette seconde alternative permet de réaliser des classifications des comportements sur des historiques « mensuels ». Elle permet également de mesurer des transferts ou changement de comportements d'un mois sur l'autre et d'observer les effets saisonniers (périodes de fêtes, vacances, ...).

Une troisième alternative est d'être avec une granularité carte * jour. Dans ce cas, la base d'étude contiendra une même carte autant de fois que de journées ou cette carte aura été utilisée. Des indicateurs d'historique d'utilisation de la carte peuvent alors être construits telles que le nombre et le montant des transactions à J, à J-1, par pays et type d'achat. Il s'agit de scores ne permettant pas d'arrêter à proprement parler la fraude mais permettant à la banque d'identifier avant ses clients les tentatives et / ou transactions frauduleuses afin de rembourser le client et mettre en opposition sa carte. Ce type de modèle, peu couteux à mettre en place permet à la banque une gestion positive de la fraude pour ses clients. En détectant en premier la fraude, la banque peut procéder à des remboursements des comptes fraudés et renforcer son image de banquier protecteur.

Une quatrième alternative, technologiquement plus difficile à mettre en œuvre est une analyse en temps réel des transactions afin de détecter, toujours en temps réel, les comportements de fraude. Dans ce cas, l'unité statistique est la transaction. La difficulté technologique réside en la capacité du système de lutte contre la fraude, et du modèle statistique à disposer et calculer en temps réel des indicateurs basés sur l'historique du comportement de fraude. A ce niveau de granularité, la transaction doit être enrichie d'indicateurs d'analyse des historiques glissants d'une carte, d'un commerçant, d'un pays, ... afin d'augmenter l'information portée par la transaction et mieux discriminer un comportement de fraude. L'utilisation d'historiques glissants permet d'éviter des effets de seuils et rend la détection beaucoup plus efficace. Pour chaque transaction, nous calculerons donc des indicateurs comme par exemple le nombre de transactions acceptées pour la carte sur les 2 dernières heures, 24 dernières heures. Le montant des transactions réalisées sur internet pour la carte sur les 2 dernières heures, 24 dernières heures. Le nombre de cartes différentes utilisées chez un même commerçant pendant les 5 dernières heures.

Par rapport aux différentes stratégies de modélisation liées à la granularité de la base d'apprentissage, la stratégie la plus efficace est de pouvoir calculer pour chaque transaction la probabilité que celle-ci soit frauduleuse. Cela nécessite d'avoir comme granularité la transaction. Hors à ce niveau, l'information disponible n'est pas d'une grande richesse. Une transaction se caractérise par une date/heure, un montant, un code réponse un code pays et une catégorie de commerçant (grande distribution, carburant, restaurant, ...).

Le fraudeur n'est pas forcément une personne isolée, il s'agit majoritairement d'une personne ou organisation ayant pour objectif de récupérer le maximum d'argent dans le minimum de temps. Cela dit, tous les indicateurs que nous devrons créer devront avoir ce point commun. Pour ce faire, nous calculerons par exemple des indicateurs de type vélocité. Pour chaque transaction, nous calculerons le nombre de transactions réalisées par la même carte sur un historique glissant de 5 minutes, 10 minutes, 15 minutes, 30 minutes, 1 heure, ... Nous pourrons également calculer toujours pour chaque transaction, le nombre de transactions réalisées sur 1h pour la carte en retrait, en paiement, sur internet, ... Ce type d'indicateur permettra de mettre en évidence une carte qui soudainement réalise un nombre élevé de transactions dans un délai de temps court. Ces indicateurs, compliqués à calculer au niveau algorithmique car utilisant des historiques glissants sont une des clefs de l'efficacité des modèles de détection de fraude à la carte bancaire.

Comme pour tout score, le nombre d'indicateurs à créer en phase de modélisation ne doit pas être limitant afin d'accroitre l'espace de recherche du comportement de fraude. La fraude n'étant pas systématiquement linéaire, ces indicateurs devront être discrétisés afin de valoriser les effets de seuils ou les effets quadratiques.

En se basant sur les indicateurs de type Récence Fréquence Montant classique, on peut définir une formule selon laquelle le scoring sera effectif. On aura, par exemple, la formule suivante :

Score RFM = Récence * (poids accordé à la Récence) + Fréquence * (poids accordé à la Fréquence) + Montant * (poids accordé au Montant)

Où, chaque poids est défini arbitrairement. Notons que le score RFM se calcule sur une période donnée que l'on définit.

Ici la récence correspond à la date du dernier achat effectué soit le nombre de jours écoulé depuis le dernier achat. La fréquence, elle, correspond au nombre d'achats effectués, et le montant est la moyenne de la somme dépensée. Dans notre cas, plus le score sera élevé, plus douteuse serait la transaction.

Pour la réalisation et la faisabilité de notre projet, nous devons utiliser des jeux de données

3.1.4. Architecture Matérielle

Afin que notre système soit effectif et qu'il puisse être implémenter, il faudrait qu'il repose sur un support. Notre système sera utilisé par des banques ou des organismes interbancaires. Notre système pourra être hébergé sur des plateformes propres à chaque entreprise. Dans notre cas, notre système aura comme support le cloud[5], pour permettre une gestion centralisée et rationnelle des fraudes. Ceci permettra aussi un accès facile et garantie.

En effet, notre système se connectera à des systèmes monétiques. Ces systèmes monétiques feront office de sources d'informations au système. La sécurité qui tournera autour de notre système serait celle du cloud. Des critères de sécurités qui seront mis en exergue ici seront celles utilisés par les fournisseurs de cloud dans le cas de systèmes critiques tel que par exemple :

- Le chiffrement des données qui peut être géré selon le bon vouloir du client
- Les options de connectivité autorisant les connexions privées, ou dédiées
- Des services permettant de gérer des attaques de types ddos
- Des services permettant de gérer les accès et identités
- Etc.

[5] *« Le cloud computing est un modèle qui permet un accès réseau à la demande et pratique à un pool partagé des ressources informatiques configurables (telles que réseaux, serveurs, stockage, applications et services) qui peuvent être provisionnées rapidement et distribuées avec un minimum de gestion ou d'interaction avec le fournisseur de services »*

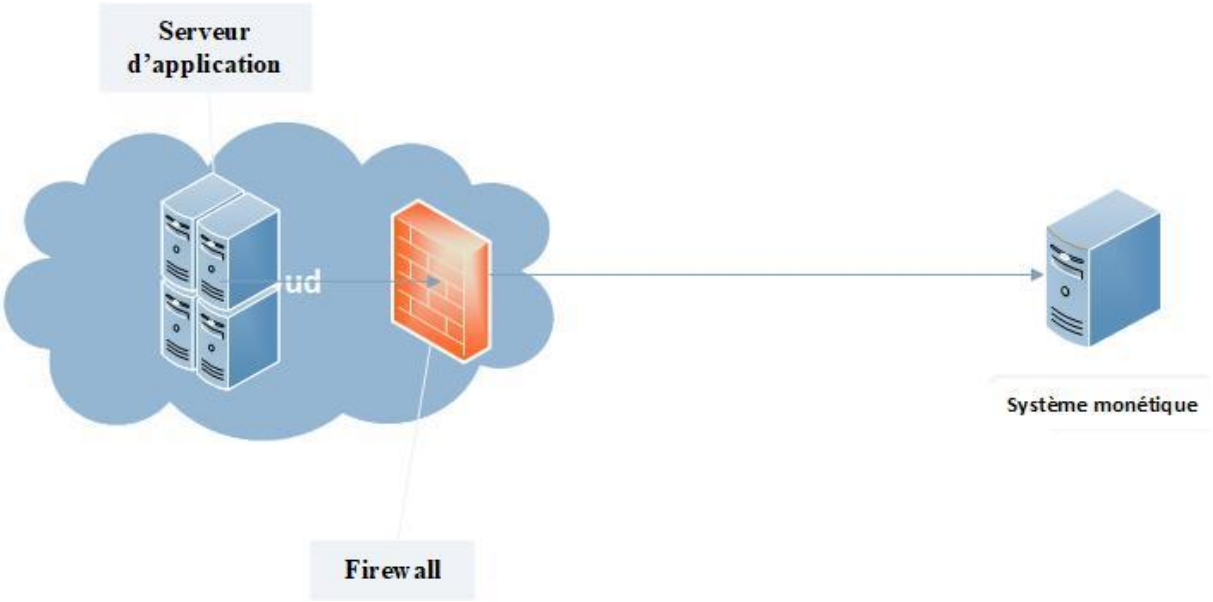

Figure 11 :Architecture système

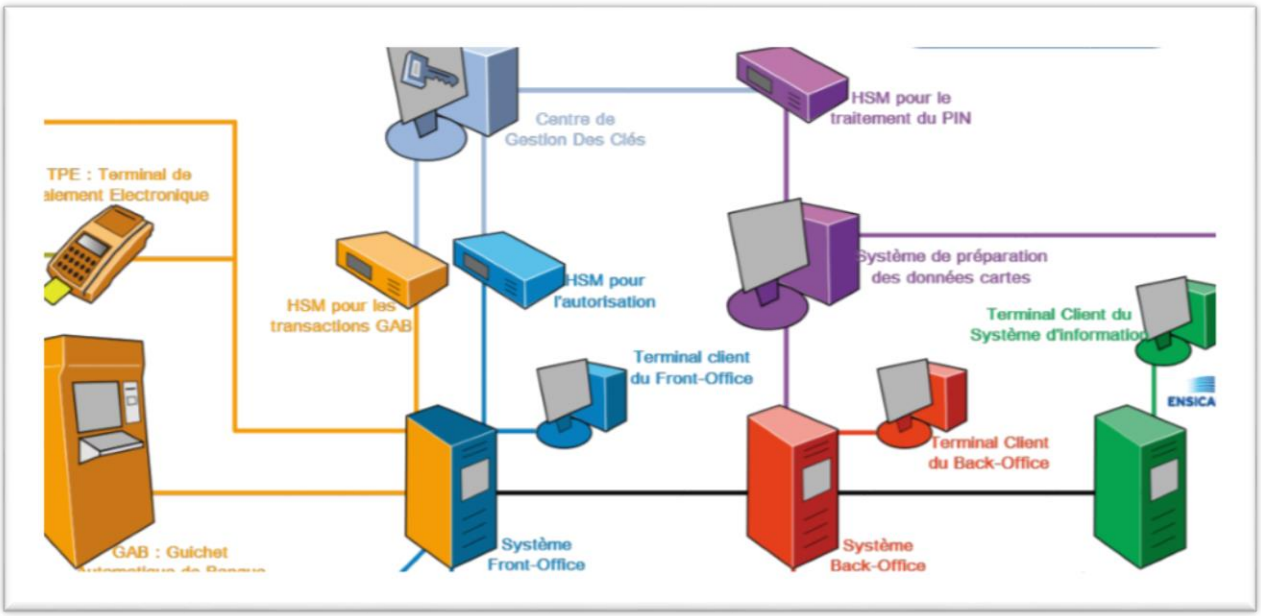

Figure 12 : Architecture d'un système monétique

3.2. Outils de développement et réalisation

3.2.1. Choix du langage de développement

Tableau 3 : Présentation des différents langages de programmation de l'intelligence artificielle

Langages	Avantages	Inconvénients
Python	- Simplicité de syntaxe - Adapté à tout sorte de plateforme tel Windows, mac, linux, etc. - Python supporte le développement fonctionnel, orienté objet, et procédural - Temps de développement relativement court par rapport à d'autres langages - Grande communauté - Grande bibliothèque de librairies	- Compilation et exécution plus lentes que C++ ou Java - Pas adapté à l'intelligence artificielle liée aux application mobiles
Java	- Java convient bien au traitement automatique du langage naturel (NLP, Natural Language Processing) et aux recherches algorithmiques mais aussi au Neural Network (des séries d'algorithmes qui reproduisent une intelligence humaine) - Contrairement à C++, Java est relativement simple à utiliser et à débugger - Adapté à de nombreuses plateformes - Pas besoin de compilation	- Java est plus lent que C++ et Python. Il souffre d'un temps de réponse plus long et d'une exécution moins rapide
C++	- Outil bien adapté pour résoudre des problèmes complexes d'intelligence artificielle.	- Syntaxe de développement relativement compliqué par rapport à python

	- On peut trouver des librairies de fonctions assez fournies - C++ est un outil de développement multi-paradigmes qui supporte les principes de l'orienté objet	- Compliqué pour des développeurs néophytes
LISP	- LISP est utilisé en IA pour sa grande flexibilité, pour sa rapidité à sortir des prototypes - Rapide et efficace dans la phase de développement, comme le langage est supporté par des compileurs à a place des interpréteurs	- Communauté petite pour trouver des informations facilement vu son ancienneté - Ne marche pas dans tous les environnements, en raison de son ancienneté
R	- Langage de programmation efficace pour analyser et manipuler des données dans un but statistique	Souvent utilisé que pour analyser, présenter des statistiques

Au vu des avantages qu'offre python par rapport aux autres langages, nous avons opté le développement de notre système dans ce langage.

3.2.2. Les Framework du machine learning

Plusieurs langages de programmation existent à nos jours, pour le développement des applications. Et aujourd'hui, la programmation de l'intelligence artificielle est un nouveau secteur où les langages font rage. D'après plusieurs recherches Python reste un langage phare de l'intelligence artificielle. Le tableau ci-après met en exergue les différents langages et avantages pour le développement d'application d'intelligence artificielle.

Un Framework peut être défini comme étant une collection de bibliothèques constituant les fondations de base pour le développement, la maintenance et le suivi d'une application. Le Framework oblige ainsi les développeurs à suivre des conventions dans leur manière et façon de coder. L'avantage de cette obligation réside dans le fait que n'importe quel développeur serait en mesure de prendre la main sur l'ensemble du projet rapidement. Les Frameworks de

Machine Learning aident à s'affranchir des tâches fastidieuses lors de l'expérimentation, de l'optimisation ou de la mise en production d'une Intelligence Artificielle (IA).

Mais différents frameworks ont, logiquement, des atouts et des défauts différents. Parmi les frameworks les plus populaires aujourd'hui, citons TensorFlow, MXNet, scikit-learn, Keras et PyTorch. Parmi ces frameworks, scikit-learn est celui que nous avons retenu à cause de sa facilité d'accès. Il permet de rapidement tester les modèles de base utilisés en machine learning, et dans un outil pédagogique, il est très adapté.

3.3. Présentation du système

3.3.1. Jeu de données

Cette section décrit les données utilisées pour générer les jeux de données. Il décrit également l'analyse utilisée pour trouver des exemples de fraudes à utiliser pour les méthodes d'apprentissage supervisé.

Notre étude s'appuie principalement sur les données collectées lors des transactions, sur internet, des porteurs de carte. Pour pouvoir former un algorithme d'apprentissage supervisé, des données avec exemple de fraude sont nécessaires. Avec des exemples de fraudes, des méthodes sont utilisées pour trouver des modèles sous-jacents dans le comportement du fraudeur. Comme indiqué précédemment, les fraudeurs sont motivés par un gain maximal dans les plus brefs délais.

La définition d'un comportement frauduleux est une tâche difficile en ce qui concerne le comportement des utilisateurs. Qu'est-ce qui sera considéré comme anormal et frauduleux ? plusieurs utilisateurs, avec différents profils existent. Il est presque impossible de regarder un profil d'utilisateur sur une certaine période et de le qualifier de frauduleux, mais en se basant sur ses habitudes, son comportement sur une certaine période déterminante, on peut utiliser des méthodes pour détecter des anomalies dans ses habitudes futures.

3.3.2. Tests

Avant de pouvoir utiliser scikit-learn, il faut préalablement installer plusieurs pré-requis.il faut préalablement avoir python installé, et les librairies NumPy, Scipy, Joblib.

La figure ci-dessous montre le code qui permet de charger des données. Ceci permettra de faire une analyse des jeux de données.

temp.py | essai_memo.py*

```
# -*- coding: utf-8 -*-
"""

@author: pc
"""
#librairies
import pandas as pd
import numpy as np

#visualisation de données
import matplotlib.pyplot as plt
#visualisation de données sous forme d'histogramme
import seaborn as sns

#charger les données
transaction_data = pd.read_csv('C:/Users/willt/Downloads/credit-card-data/

#pour avoir des informations sur les données chargées
transaction_data.info()
#un histogramme des données chargées
sns.pairplot(transaction_data)
#pour voir un peu un histogramme qui voir une moyenne
sns.distplot(transaction_data['Transaction_Value'])
#pour afficher une certaine corrélation entre les différents champs de la
sns.heatmap(transaction_data.corr(),annot=True)

```

Figure 13 :Chargement des données

La figure ci-dessous montre des extraits d'informations d'un jeu de données chargées. Dans notre cas, le jeu de données concerne des informations de transactions.

Entrée [3]:
```
#pour montrer des extraits d'informations
transaction_data.head()
```

Out[3]:

	Transaction_ID	Transaction_Date	Credit_Card_ID	Transaction_Value	Transaction_Segment
0	CTID28830551	24-Apr-16	1629-9566-3285-2123	23649	SEG25
1	CTID45504917	11-Feb-16	3697-6001-4909-5350	26726	SEG16
2	CTID47312290	1-Nov-16	5864-4475-3659-1440	22012	SEG14
3	CTID25637718	28-Jan-16	5991-4421-8476-3804	37637	SEG17
4	CTID66743960	17-Mar-16	1893-8853-9900-8478	5113	SEG14

Figure 14 : Exemple de Jeu de données utilisé

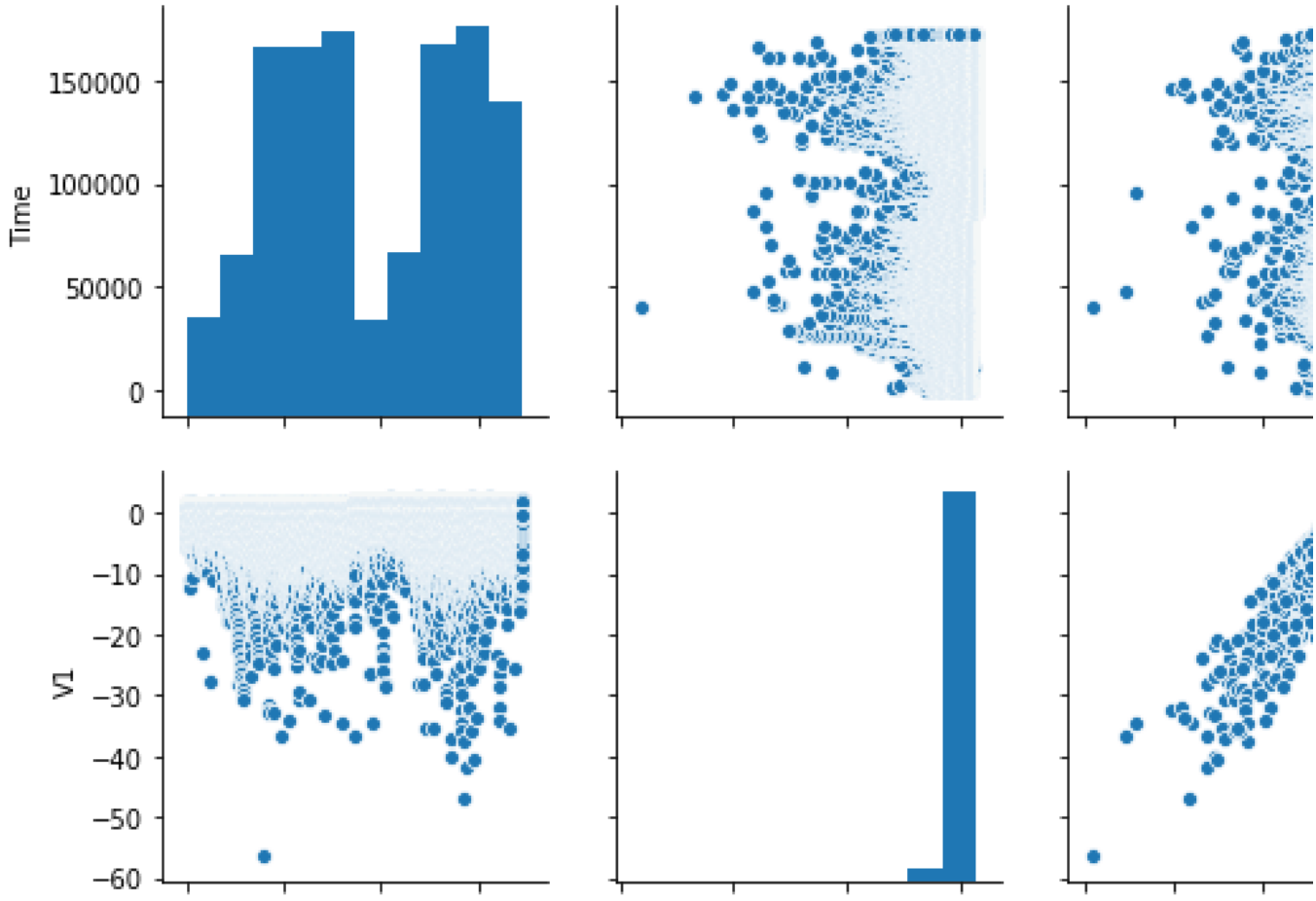

Figure 15 : Exemple d'histogramme de données bancaires

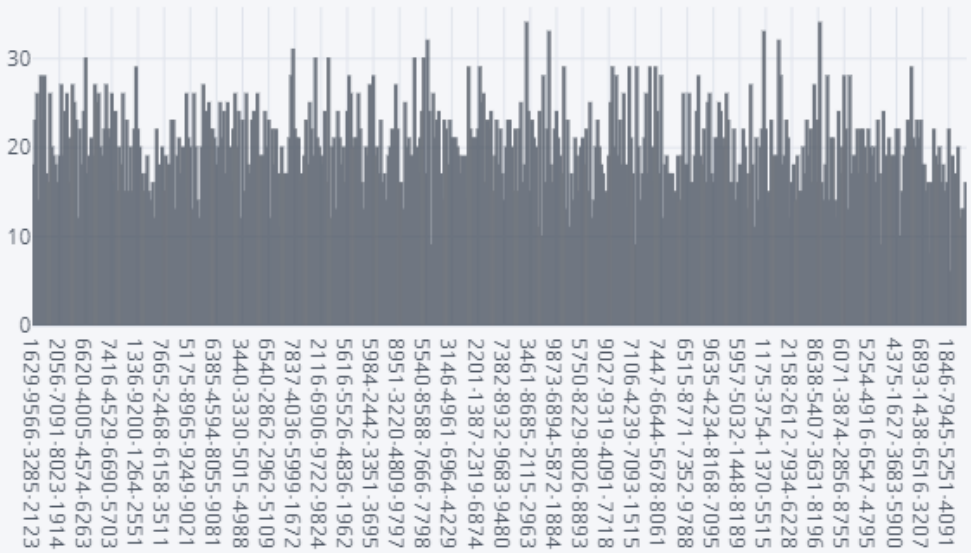

Figure 16 :*Graphe montrant le nombre de transaction par carte bancaire*

Conclusion

L'avancée continue et fulgurante des NTICS nous a poussé à adopter et à adapter à un nouveau style de vie dans notre quotidien. Le secteur de la monétique, caractérisé par l'utilisation des cartes, contribue à ce nouveau style de vie, et offre beaucoup d'avantages pratiques. Néanmoins, ceci n'est pas sans inconvénients ; en effet, le développement d'internet n'a pas été sans entrainer des piratages, fraudes en tout genre. C'est le cas des fraudes monétiques qui ne cesse de prendre de l'ampleur malgré les moyens mises en place. Notre travail a donc été d'étudier et de présenter des approches de solutions basé sur le machine learning pouvant palier aux nouvelles techniques de fraudes qui ne cessent de voir le jour. Pour mener à bien notre projet, il nous a fallu comprendre les différentes techniques de fraudes existantes ainsi que le fonctionnement du machine learning, avant de pouvoir mettre en place un système basé sur un type d'algorithme donné du machine learning pouvant détecter des fraudes monétiques.

Détecter ou prévenir une fraude suit un processus cyclique d'identifications, d'évaluations et de traitements de risques, en vue de mener des actions pour la contrer. Le processus reste le même avec l'introduction des statistiques et du machine learning. En effet, le nouveau système de détection de fraudes basé sur le machine learning se base sur des algorithmes qui lui permet d'apprendre des informations qu'il reçoit. Ainsi il pourrait, par exemple selon le profil d'un porteur, déterminer si la transaction a belle et bien été faite par lui. La réalisation de ce projet nous a permis de découvrir et d'explorer les horizons du machine learning. Plus encore, elle nous a permis d'en apprendre plus sur la monétique. Un secteur caractérisé par les cartes bancaires. Ces domaines présentent beaucoup d'intérêts, et suscitent beaucoup d'émerveillement ; encore loin d'être à leur apogée, nous restons convaincus qu'il en découlera beaucoup d'options.

Il est important de notifier que nous avons été heurté à beaucoup de problèmes dans la réalisation de notre projet, notamment en ce qui concerne l'obtention d'informations. En effet, à cause de la professionnalisation et de l'enjeu des domaines, très peu d'informations sont disponibles.

Le temps de travail étant relativement court, pour nous, il ne fait aucun doute que notre travail peut être amélioré. En effet, notre système peut être mise à la disposition des banques

ou des organismes comme le GIM-UEOMOA pour combattre les fraudes. Notre système, en plus de détecter des fraudes, peut être proactif selon les algorithmes mises en place, et avec la participation des différents acteurs qui entrent dans le processus de lutte contre la fraude.

Bibliographie

KINGNIDE A.B.H.L. Rody, *Etude et mise en place d'une application web de génération et de validation de fichiers de compensation LIS,* Institut Supérieur d'Informatique (ISI), 2017-2018, 104

KODJO Noé Gérald, *Etude et mise en place d'un système de paiement sécurisé sur internet,* Institut Supérieur d'Informatique (ISI), 2017-2018, 117

NIANG Fatime, *Etude et mise en place d'une plateforme de monitoring des transactions et de la lutte contre la fraude monétique,* Institut Supérieur d'Informatique (ISI), 2017-2018, 117

Webographie

- http://www.gim-uemoa.org/fr/politique-securitaire/lutte-contre-la-fraude-la-cybercriminalite
- https://www.cartes-bancaires.com/agrement-et-securite/securite-des-donnees/
- https://fr.wikipedia.org/wiki/Mon%C3%A9tique
- https://www.wavestone.com/app/uploads/2017/03/lutte-fraude-bancaire-en-ligne-nouvelles-methodes.pdf
- https://www.undernews.fr/banque-cartes-bancaires/etude-independante-lutte-contre-la-fraude-monetique-carding.html
- https://www.jurifiable.com/conseil-juridique/droit-penal/types-de-fraudes
- http://www.anz.com/vanuatu/fr/personal/ways-bank/internet-banking/protect-banking/types-fraud/
- https://banque.ooreka.fr/astuce/voir/746109/fraude-a-la-carte-bancaire
- http://www.comprendrelespaiements.com/abc-de-la-monetique-les-acteurs-et-leurs-roles/
- https://business.lesechos.fr/directions-financieres/comptabilite-et-gestion/gestion-des-risques/0211917750477-fraude-les-10-scenarios-les-plus-inquietants-307893.php
- https://monetiques.wordpress.com/2015/06/16/la-fraude-emetteur-et-acquereur/
- https://deontofi.com/fraude-aux-cartes-de-credit-avec-code-sms-les-banques-doivent-rembourser/
- https://www.soprabanking.com/actualites/-ro-intelligence-artificiel-et-du-big-data-prevention-fraudes-17
- https://www.lemagit.fr/definition/Authentification-a-double-facteur le 13/08/19 à 23:25
- https://docassas.u-paris2.fr/nuxeo/site/esupversions/1d014ad6-e51f-4538-a3ce-7ab895e1c831?inline le 28/08/19 à 02:47
- https://fr.calameo.com/read/005557832401e2c902dfc le 29/08/19 à 00:13
- https://fr.calameo.com/books/005557832401e2c902dfc le 29/08/19 à 00:13
- https://www.banque-france.fr/sites/default/files/medias/documents/oscp-rapport-annuel-2009.pdf
- https://tel.archives-ouvertes.fr/tel-01009369/document le 29/08/19 à 00:24
- https://afrique.latribune.fr/africa-tech/2019-04-08/afrique-de-l-ouest-les-banques-ripostent-face-a-la-cybermenace-813496.html 05/09/19 22:12

- https://www.lebigdata.fr/machine-learning-et-big-data le 17/09/19 à 20:15
- https://www.sas.com/fr_ch/insights/analytics/machine-learning.html le 17/09/19 à 20:15
- https://openclassrooms.com/fr/courses/4011851-initiez-vous-au-machine-learning/4020611-identifiez-les-differents-types-dapprentissage-automatiques le 17/09/19 à 20:21
- https://blog.stack-labs.com/code/la-grande-famille-des-mod%C3%A8les-de-machine-learning/ le 11/10/2019
- http://www.statsoft.fr/industries/detection-de-fraudes.php le 15/10/19 à 02:00
- https://startuppers.club/fr/code/quel-langage-apprendre-pour-coder-lintelligence-artificielle-ai/ le 16/10/19 à 00:06
- https://github.com/awesomedata/awesome-public-datasets le 28/09/19 à 21:16
- https://www.kaggle.com/datasets le 17/10/19 à 23:09

Annexes

Table des matières

Résumé

L'avancée continue et fulgurante des NTICS nous a poussé à adopter et à adapter à un nouveau style de vie dans notre quotidien. Le secteur de la monétique, caractérisé par l'utilisation des cartes, contribue à ce nouveau style de vie, et offre beaucoup d'avantages pratiques. Néanmoins, ceci n'est pas sans inconvénients ; en effet, le développement d'internet n'a pas été sans entrainer des piratages, fraudes en tout genre. C'est le cas des fraudes monétiques qui ne cesse de prendre de l'ampleur malgré les moyens mises en place. Notre travail a donc été d'étudier et de présenter des approches de solutions basé sur le machine learning pouvant palier aux nouvelles techniques de fraudes qui ne cessent de voir le jour. Pour mener à bien notre projet, il nous a fallu comprendre les différentes techniques de fraudes existantes ainsi que le fonctionnement du machine learning, avant de pouvoir mettre en place un système basé sur un type d'algorithme donné du machine learning pouvant détecter des fraudes monétiques.

Détecter ou prévenir une fraude suit un processus cyclique d'identifications, d'évaluations et de traitements de risques, en vue de mener des actions pour la contrer. Le processus reste le même avec l'introduction des statistiques et du machine learning. En effet, le nouveau système de détection de fraudes basé sur le machine learning se doit d'aller chercher des mines d'informations dans des banques d'informations. L'information étant le cœur de notre système, elle est cruciale, donc elle se doit d'être pertinente. C'est dans ce sens, que le machine learning sélectionne et utilise les informations en leur créant une valeur ajoutée, afin de permettre d'identifier, avec un taux d'échec le plus bas possible, des comportements frauduleux. En effet, basé sur des méthodes d'apprentissages, le machine learning apprend des données qu'il reçoit. Grâce à cela, il arrive à déduire, à s'adapter et à réaliser des tâches tel que définir le profil d'un porteur de carte par exemple, ou encore définir une transaction frauduleuse.

Mots-clés : Machine learning, Monétique, Intelligence artificielle, Algorithme

Abstract

The continuous and dazzling progress of the NTICS has pushed us to adopt and adapt to a new lifestyle in our daily lives. The electronic banking sector, characterized by the use of cards, contributes to this new lifestyle, and offers many practical advantages. Nevertheless, this is not without drawbacks; in fact, the development of the internet has not been without causing piracy, fraud of any kind. This is the case of electronic fraud, which is growing in spite of the means put in place. Our work has been to study and present solutions approaches based on machine learning that can overcome the new fraud techniques that continue to emerge. To carry out our project, we had to understand the different techniques of fraud existing and the functioning of machine learning, before being able to set up a system based on a given type of machine learning algorithm that can detect electronic fraud.

Detecting or preventing fraud follows a cyclical process of identifying, assessing and treating risks in order to take action to counter it. The process remains the same with the introduction of statistics and machine learning. Indeed, the new fraud detection system based on machine learning has to search for information mines in information banks. Information being the heart of our system, it is crucial, so it must be relevant. It is in this sense that the machine learning selects and uses the information by creating added value, in order to identify fraudulent behavior with the lowest possible failure rate. Indeed, based on learning methods, machine learning learns the data it receives. With this, he can deduce, adapt and perform tasks such as define the profile of a cardholder for example, or define a fraudulent transaction.

Keywords: Machine learning, Electronic payment, Artificial intelligence, Algorithm