REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université de Mohamed El-Bachir El-Ibrahimi - Bordj Bou Arreridj

Institut d'Electronique et des Télécommunications (IET)





Mémoire

Présenté pour obtenir

LE DIPLOME DE MASTER

Filière : Télécommunications

Spécialité : Systèmes des TELECOMMUNICATION

Par

BENGUEDDOUDJ MERWA

AREZKI CHIRAZ

Thème :

Détection de la fraude dans les transactions bancaires par l'intelligence artificielle

Soutenu le 23/06/2025 :

Devant le Jury composé de :

Nom & Prénom	Grade	Qualité	Etablissement
Dr. LAKHLEF Nora	**M.C. B**	**Président**	**Univ-BBA**
Dr. SID AHMED Soumia	**M.C . A**	**Examinateur**	**Univ-BBA**
Pr. ASBAI Nassim	**Professeur**	**Encadrant**	**Univ-BBA**

Année Universitaire 2025/2026

Remerciements

Nous remercions tout d'abord Dieu le tout puissant qui nous avoir accordé la volonté et le courage pour l'accomplissement de ce travail.

Nous exprimons nos remerciements avec grand plaisir et grand respect à notre cher encadrant : Pr.ASBAI .NASSIM professeur de l'université de Mohamed El-Bachir El-Ibrahimi - Bordj Bou Arreridj à celui qui nous avons fourni le sujet de cette thèse, pour sa disponibilité, sa gentillesse, son soutien et ses encouragements, nous ne le remercions jamais assez. Grâce à lui nous avons pu réaliser ce travail.

Nous remercions également tous nos professeurs pendant nos cursus universitaires au sein de l'université de Bordj Bou Arreridj pour leur soutien et leur accueil chaleureux et sympathique.

Nous remercions également les membres du jury d'avoir accepté d'étudier ce travail.
Nous adressons nos plus sincères remerciements à nos chers parents pour leur dévouement, leur patience, leurs sacrifices et leur soutien indéfectible tout au long de notre parcours.
Encore un dernier remerciement à toute personne qui a contribué de près ou de loin dans la réussite de ce modeste travail.

Dédicace

Je dédie ce modeste travail :

Tout d'abord, à Allah le Tout-Puissant,pour m'avoir accordé la santé, la force, la patience et la volonté nécessaires afin de mener à bien ce travail.

À moi-même,en reconnaissance de ma patience, de ma persévérance et des efforts que j'ai fournis tout au long de ce parcours, malgré les difficultés rencontrées.

À mes très chers parents,pour leur amour inconditionnel, leur soutien constant et leurs sacrifices inestimables. Les mots ne sauraient exprimer toute ma gratitude et mon profond respect à leur égard.

À mes chers frères **DJAMEL, ABDERRAHMANE** et **NEDJMEUDDINE**, ainsi qu'à mes chères sœurs **AYA, SOUHIR** et **IMANE**,pour leurs encouragements, leur présence à mes côtés et leur soutien moral tout au long de mes études.

À mon honorable professeur, **Dr ASBAI NASSIM**,pour son encadrement, ses précieux conseils et sa disponibilité tout au long de la réalisation de ce mémoire.

À toutes les personnes qui m'ont aidé de près ou de loin à réaliser ce travail,et qui ont contribué, chacun à sa manière, à l'aboutissement de ce projet.

Que ce travail soit le témoignage de ma reconnaissance, de mon respect et de ma sincère affection envers toutes ces personnes.

BENGUEDDOUDJ MERWA

Dédicace

Je dédie ce modeste travail :

Tout d'abord, à Allah le Tout-Puissant,pour m'avoir accordé la santé, la force, la patience et la volonté nécessaires afin de mener à bien ce travail.

À moi-même, en reconnaissance de tous les efforts, de la patience et de la persévérance que j'ai fournis tout au long de mes années d'études, ainsi que pour tous les défis que j'ai pu surmonter grâce à Dieu et à ma détermination.

À mes chers parents, symbole d'amour, de sacrifice et de générosité, qui ont été mon soutien à chaque étape, et sans leurs prières et leur soutien, je n'aurais pas pu atteindre ce résultat.

À mes frères **ABDESSALAM** et **ABDELGHANI**, qui ont été pour moi une source de soutien et d'encouragement, et qui m'ont accompagné dans différentes circonstances.

À mon honorable professeur **Dr ASBAI NASSIM**, qui a supervisé mon travail et m'a apporté conseils et orientations tout au long de la réalisation de cette recherche, je lui exprime mes sincères remerciements et ma profonde gratitude.

Et à toutes les personnes qui m'ont aidé de près ou de loin, qui m'ont apporté leur aide ou un mot d'encouragement, et qui ont eu un grand impact dans l'accomplissement de ce travail.

À vous tous, je dédie ce travail, avec tout mon amour et ma gratitude.

AREZKI CHIRAZ

Résumé

Le secteur bancaire connaît ces dernières années une augmentation importante des risques de fraude en raison du développement continu des services bancaires numériques et de la hausse du volume des transactions électroniques. Dans ce contexte, ce mémoire vise à étudier et à développer un système intelligent de détection de fraude dans les transactions bancaires en utilisant les techniques de l'intelligence artificielle et de l'apprentissage automatique, avec un intérêt particulier pour l'algorithme **AdaBoost**. La partie théorique présente le concept de la fraude bancaire, ses principaux types et les méthodes de lutte contre ce phénomène, tout en mettant en évidence le rôle des techniques d'apprentissage automatique dans l'amélioration des systèmes de détection des opérations suspectes. Dans la partie pratique, le jeu de données **Credit Card Fraud Detection** a été utilisé. Les données ont subi plusieurs étapes de prétraitement, notamment le nettoyage, la normalisation et le rééquilibrage des classes à l'aide des techniques **SMOTE** et **Random Under Sampling** afin de résoudre le problème du déséquilibre des données. Ensuite, le modèle **AdaBoost** a été entraîné, ses paramètres ont été optimisés et ses performances ont été évaluées à l'aide de plusieurs mesures statistiques, notamment **Accuracy, Precision , Recall , F1-Score** et **ROC-AUC**. Les résultats obtenus ont montré de très bonnes performances et une capacité efficace à distinguer les transactions légitimes des transactions frauduleuses, ce qui confirme l'importance des techniques d'intelligence artificielle dans le renforcement de la sécurité des transactions bancaires et dans le développement de systèmes de détection de fraude plus efficaces et plus fiables.

Mots-clés *:* Fraude bancaire ; Intelligence artificielle ; Machine Learning ; Détection de fraude ; AdaBoost**.**

Abstract

In recent years, the banking sector has experienced a significant increase in fraud risks due to the continuous expansion of digital banking services and the growing volume of electronic transactions. In this context, this dissertation aims to study and develop an intelligent system for detecting fraud in banking transactions using artificial intelligence and machine learning techniques, with a particular focus on the **AdaBoost** algorithm. The theoretical part reviews the concept of banking fraud, its main types, and the methods used to combat it, while highlighting the role of machine learning techniques in improving suspicious transaction detection systems. In the practical part, the **Credit Card Fraud Detection** dataset was used. The data underwent several preprocessing stages, including cleaning, normalization, and class rebalancing using the **SMOTE** and **Random Under Sampling** techniques to address the problem of data imbalance. After that, the **AdaBoost** model was trained, its parameters were optimized, and its performance was evaluated using several statistical metrics, including **Accuracy**, **Precision**, **Recall**, **F1-Score**, and **ROC-AUC**. The obtained results demonstrated high performance and an effective ability to distinguish between legitimate and fraudulent transactions, confirming the importance of artificial intelligence techniques in enhancing the security of banking transactions and developing more efficient and reliable fraud detection systems.

Keywords *:* Banking Fraud ; Artificial Intelligence ; Machine Learning ; Fraud Detection ; AdaBoost.

الملخص

يشهد القطاع المصرفي في السنوات الأخيرة تزايدًا ملحوظًا في مخاطر الاحتيال نتيجة التوسع المستمر في الخدمات البنكية الرقمية وارتفاع حجم المعاملات الإلكترونية. وفي هذا السياق، تهدف هذه المذكرة إلى دراسة وتطوير نظام ذكي للكشف عن الاحتيال في المعاملات المصرفية بالاعتماد على تقنيات الذكاء الاصطناعي وتعلم الآلة، مع التركيز على خوارزمية **AdaBoost** وقد تم في الجانب النظري استعراض مفهوم الاحتيال المصرفي وأهم أنواعه وأساليب مكافحته، إضافة إلى إبراز دور تقنيات تعلم الآلة في تحسين أنظمة الكشف عن العمليات المشبوهة.

أما في الجانب التطبيقي، فقد تم استخدام مجموعة البيانات **Credit Card Fraud Detection** حيث خضعت البيانات لعدة مراحل من المعالجة المسبقة شملت التنظيف والتطبيع وإعادة توازن الفئات باستعمال تقنيتي **SMOTE** و **Under Sampling Random** لمعالجة مشكلة عدم توازن البيانات.

بعد ذلك تم تدريب نموذج **AdaBoost** وتحسين معلماته وتقييم أدائه باستخدام مجموعة من المقاييس الإحصائية، من بينها **Accuracy** و **ROC-AUC** و **F1-Score** و **Recall** و **Precision** و قد أظهرت النتائج المحصل عليها أداءً مرتفعًا وقدرة فعالة على التمييز بين المعاملات السليمة والاحتيالية، مما يؤكد أهمية تقنيات الذكاء الاصطناعي في تعزيز أمن المعاملات المصرفية وتطوير أنظمة أكثر كفاءة وموثوقية للكشف عن الاحتيال.

الكلمات المفتاحية :الاحتيال المصرفي ؛ الذكاء الاصطناعي ؛ تعلم الآلة ؛ كشف الاحتيال ؛**AdaBoost**

Sommaire

Table de matière :

Remerciements ii
Dédicace iii
Dédicace iv
Résumé v
Abstract vi
Table de matière ix
Liste des figures xii
Liste des tableaux xiii
Liste des abréviations xiv
Introduction Générale 1

chapitre 01:

1.1. Introduction 3
1.2. Qu'est-ce que la fraude bancaire ? 3
1.3. Pourquoi la fraude bancaire est-elle en hausse ? 4
1.4. Typologies de la fraude bancaire ? 6
1.5. Facteurs et impacts de la fraude bancaire 9
1.5.1. Facteurs de la fraude bancaire 9
1.5.2. Impacts de la fraude bancaire 11
1.6. Évolution des tendances de fraude 12
1.7. Cadre réglementaire et normatif 13
1.7.1. Réglementations internationales 13
1.7.2. Cadre réglementaire national 14
1.7.3. Normes techniques 15
1.8. Acteurs du contrôle de la fraude bancaire 16
1.8.1 Acteurs internes 16
1.8.2 Acteurs externes 16
1.9. Conclusion 17

Chapitre02:

2.1. Introduction 18
2.2. Les méthodes traditionnelles de contrôle et leurs limites 18
2.2.1. Les dispositifs classiques de contrôle bancaire 18
2.2.2. Les limites des méthodes traditionnelles 22
2.3. L'apport des nouvelles technologies dans la détection de la fraude 24
2.3.1. La transformation numérique du secteur bancaire 24
2.3.2. Le rôle du Data Mining dans la détection de fraude 25
2.3.3. Le Machine Learning dans la détection de fraude 26
2.3.4. Les avantages des systèmes intelligents 28
2.4. Présentation de l'algorithme AdaBoost et son application à la détection de fraude 28
2.4.1. Présentation générale d'AdaBoost 28
2.4.2. Principe mathématique d'AdaBoost 30
2.4.3. Application d'AdaBoost à la fraude bancaire 34
2.4.4. Les métriques d'évaluation du modèle 35
2.4.5. Importance des variables (Feature Importance) 40
2.5. Conclusion 41

Chapitre 03:

3.1. Introduction 41
3.2. Présentation des données et prétraitement 41
3.2.1. Présentation du jeu de données 41
3.2.2. Prétraitement des données 43
3.3. Gestion du déséquilibre des classes 44
3.4. Entraînement et optimisation du modèle AdaBoost 46
3.4.1. Construction du modèle 46
3.4.2. Optimisation des hyperparamètres 47
3.4.3. Optimisation du seuil de classification 48
3.5. Résultats et évaluation des performances 50
3.5.1. Métriques d'évaluation 50
3.5.2. Résultats expérimentaux 50
3.5.3. Analyse des figures obtenues 50

3.6. Conclusion 62
Conclusion Générale et Perspectives 65
Liste des références 68

Liste des figures

Figure 1.1 Les composantes essentielles de la fraude bancaire et leurs interactions 4
Figure.1.2 Évolution reconstituée de la fraude bancaire en Algérie (2018–2023) 5
Figure.1.3 Types de cyberattaques bancaires..... 7
Figure.1.4 Représentation visuelle des principales typologies de la fraude bancaire 9
Figure.2.1 Structure traditionnelle du contrôle bancaire 21
Figure.2.2 Elément déclencheur de la fraude..... 21
Figure.2.3. Délai de découverte de la fraude 23
Figure.2.4 Architecture d'un système de détection de fraude bancaire en temps réel 24
Figure.2.5 Architecture moderne d'un système bancaire numérique..... 25
Figure.2.6 Processus général de détection de fraude par Machine Learning..... 27
Figure.2.7 Fonctionnement général de l'algorithme AdaBoost..... 29
Figure.2.8 Illustration de l'algorithme AdaBoost 31
Figure.2.9 Déséquilibre des classes dans un dataset de fraude 35
Figure.2.10 Sous-apprentissage, apprentissage approprié et sur-apprentissage 37
Figure.2.11 Matrice de confusion (valeurs réelles vs prédites) 37
Figure.2.12 Schéma de courbe ROC en orange et AUC en gris. 40

Figure.3.1 Distribution des classes dans le dataset 43
Figure.3.2 Architecture générale du système de détection de fraude bancaire basé sur AdaBoost… 46
Figure.3.3 Matrice de confusion normalisée du modèle AdaBoost 51
Figure.3.4 Courbe ROC du modèle AdaBoost 52
Figure.3.5 Courbe Precision-Recall du modèle AdaBoost..... 53
Figure.3.6 Distribution des probabilités par classe (KDE)..... 54
Figure.3.7 Comparaison des métriques d'évaluation 55
Figure.3.8 Top 10 des caractéristiques les plus importantes 56
Figure.3.9 Évolution des métriques en fonction du seuil de classification..... 57
Figure.3.10 Trade-off Precision-Recall..... 58
Figure.3.11 Évolution de l'Accuracy pendant l'entraînement 59
Figure.3.12 Évolution de la Loss pendant l'entraînement 60
Figure.3.13 Performance du modèle par classe 61

Liste des tableaux

Tableau.1.1 :récapitulatif des typologies……8
Tableau.1.2 :Synthèse des impacts de la fraude bancaire.....12
Tableau.1.3 :Acteurs externes de la lutte contre la fraud 16
Tableau.2.1 :Limites des approches traditionnelles.....23
Tableau 2.2 :Comparaison des techniques de Machine Learning.....27
Tableau 2.3 :Exemple d'importance des variables.....40
Tableau 3.1 : Résultats de la séparation des données d'apprentissage et de test……42
Tableau 3.2 : Distribution des classes après le rééquilibrage du dataset…….... 45
Tableau 3.3 :Optimisation des hyperparamètres par GridSearchCV……48
Tableau 3.4 :Seuil de classifications optimal……48
Tableau 3.5 :Performances obtenues avec le seuil optimal 49
Tableau 3.6 :Résultats expérimentaux du modèle AdaBoost 50

Liste des abréviations

- **AI** : Artificial Intelligence
- **API** : Application Programming Interface
- **AUC** : Area Under the Curve
- **CERT** : Computer Emergency Response Team
- **CERT-Algeria** : Centre Algérien de Réponse aux Incidents Informatiques
- **CIB** : Carte Interbancaire
- **CNP** : Card Not Present
- **COSO** : Committee of Sponsoring Organizations
- **CSV** : Comma-Separated Values
- **DL** : Deep Learning
- **DSP2 / PSD2** : Directive sur les Services de Paiement 2
- **F1-Score** : Harmonic Mean of Precision and Recall
- **FN** : False Negative
- **FP** : False Positive
- **FPR** : False Positive Rate
- **IA** : Intelligence Artificielle
- **ISO** : International Organization for Standardization
- **ISO 27001** : Norme de management de la sécurité de l'information
- **KDE** : Kernel Density Estimation
- **KPI** : Key Performance Indicator
- **ML** : Machine Learning
- **NFC** : Near Field Communication
- **NIS2** : Network and Information Security 2
- **PCA** : Principal Component Analysis
- **PCI DSS** : Payment Card Industry Data Security Standard
- **Precision** : Précision des prédictions positives
- **Recall** : Taux de détection des fraudes
- **RF** : Random Forest
- **RGPD** : Règlement Général sur la Protection des Données

- **ROC** : Receiver Operating Characteristic
- **SCA** : Strong Customer Authentication
- **SMOTE** : Synthetic Minority Over-sampling Technique
- **SSI** : Sécurité des Systèmes d'Information
- **SVM** : Support Vector Machine
- **TN** : True Negative
- **TP** : True Positive
- **VS Code** : Visual Studio Code
- **XGBoost** : Extreme Gradient Boosting

Introduction générale

Introduction Générale

Le secteur bancaire connaît une transformation profonde sous l'effet de la digitalisation des services financiers, du développement des paiements électroniques et de l'essor des technologies numériques. Cette évolution a permis d'améliorer la rapidité des transactions, l'accessibilité des services bancaires ainsi que la qualité des services offerts aux clients. Dans ce contexte, les technologies émergentes, notamment l'intelligence artificielle, le Machine Learning, le Big Data et l'analyse comportementale, occupent une place croissante dans les systèmes financiers modernes grâce à leur capacité à analyser de grands volumes de données et à détecter des comportements anormaux [1].

Cependant, cette transformation numérique a également favorisé l'apparition de nouvelles formes de criminalité financière, notamment la fraude dans les transactions bancaires [2] . Les fraudeurs exploitent aujourd'hui les technologies numériques pour contourner les dispositifs de sécurité traditionnels à travers diverses techniques telles que le phishing, le skimming, l'usurpation d'identité et les cyberattaques [3] .Selon plusieurs rapports internationaux, les pertes financières liées à la fraude et à la cybercriminalité continuent d'augmenter en raison de la sophistication croissante des méthodes employées par les fraudeurs [4] .En Algérie, le développement des cartes CIB, des services bancaires en ligne et des solutions de paiement électronique a renforcé l'exposition du secteur bancaire aux risques de fraude, ce qui exige des mécanismes de sécurité plus performants [5].

Face à l'augmentation du volume des transactions électroniques et à la complexité des schémas frauduleux, les systèmes traditionnels de détection montrent certaines limites. Les méthodes fondées sur des règles fixes et les contrôles manuels peinent à identifier efficacement les comportements suspects en temps réel [6] .De plus, les fraudeurs exploitent continuellement les failles des systèmes d'information afin de contourner les dispositifs de sécurité existants [7] .Le déséquilibre entre les transactions légitimes et frauduleuses constitue également un défi majeur pour les systèmes de détection [8] .Cette situation soulève la problématique suivante : comment améliorer l'efficacité des systèmes de détection de fraude dans les transactions bancaires afin d'identifier rapidement les opérations frauduleuses tout en réduisant les erreurs de classification ?

Dans ce cadre, ce mémoire a pour objectif d'étudier l'utilisation de l'intelligence artificielle et du Machine. Learning pour la détection de la fraude bancaire, Il vise à analyser les différentes formes de fraude, à présenter les limites des approches traditionnelles et à mettre en évidence l'intérêt des méthodes de Machine Learning pour la détection des transactions frauduleuses. Une attention particulière est accordée à l'algorithme AdaBoost**,** utilisé comme modèle de classification. Pour atteindre ces objectifs, une méthodologie analytique et expérimentale a été adoptée, reposant sur une étude bibliographique ainsi que sur l'implémentation et l'évaluation d'un modèle appliqué à un jeu de données de transactions bancaires. Les performances du modèle sont évaluées à l'aide de plusieurs indicateurs tels que l'Accuracy, la Precision, le Recall, le F1-Score et l'AUC [9] . Les données font également l'objet d'un prétraitement comprenant la normalisation et la gestion du déséquilibre des classes à l'aide de techniques adaptées telles que SMOTE [01].

Chapitre 1:

Le cadre théorique et réglementaire de la fraude bancaire

.1. Introduction :

La fraude bancaire représente aujourd'hui un défi important pour les institutions financières dans le monde entier [11]. Avec le développement des technologies numériques et des services bancaires en ligne, les méthodes de fraude sont devenues plus variées et plus sophistiquées [12]. Cette situation oblige les banques à renforcer leurs systèmes de sécurité et de contrôle afin de protéger les données et les transactions des clients.

Le secteur bancaire est particulièrement exposé à ces risques en raison de la forte utilisation des moyens de paiement électroniques et de la digitalisation croissante des opérations financières [13]. La fraude bancaire peut avoir des conséquences économiques importantes et affecter la confiance des utilisateurs dans les services financiers numériques.

Dans ce contexte, ce chapitre a pour objectif de présenter les notions générales liées à la fraude bancaire. Il abordera les principales formes de fraude, leurs facteurs et leurs impacts, ainsi que le cadre réglementaire et les acteurs impliqués dans la lutte contre ce phénomène [14].

.2. Qu'est-ce que la fraude bancaire ?

La fraude bancaire désigne tout acte intentionnel et illégal commis par une personne physique ou morale dans le but d'obtenir un avantage financier ou un bien au détriment d'une banque, d'un client ou d'un système de paiement **[15].**

Dans le contexte algérien, elle est considérée comme une composante du risque opérationnel pouvant affecter la stabilité du système bancaire et la confiance des utilisateurs.
Trois éléments sont généralement réunis pour qualifier une fraude bancaire :

- L'intention : l'acteur agit en connaissance du caractère illégal.
- L'acte matériel : une action concrète (piratage, falsification, manipulation).
- Le préjudice : perte financière pour la banque ou le client.

Il est important de distinguer la fraude bancaire d'autres infractions proches comme le blanchiment d'argent (qui consiste à dissimuler l'origine de fonds déjà illicites) ou le simple non- paiement d'un prêt (qui relève plutôt du risque de crédit) **[16].**



Figure.1.1 : Les composantes essentielles de la fraude bancaire et leurs interactions

.3. Pourquoi la fraude bancaire est-elle en hausse ?

Plusieurs facteurs expliquent l'augmentation constante de la fraude bancaire ces dernières années. Comme le souligne une étude récente, les cyberattaques deviennent de plus en plus sophistiquées et leurs impacts plus graves, ce qui oblige les organisations à renforcer continuellement leurs stratégies de défense **[17].**

Parmi les facteurs principaux :

- **La numérisation des services bancaires :** augmentation des transactions en ligne et donc des surfaces d'attaque.
- **L'anonymat relatif sur Internet :** les fraudeurs peuvent agir depuis différents pays.
- **La faiblesse humaine :** le facteur humain reste le maillon faible (phishing, mots de passe faibles).
- **L'économie souterraine numérique :** vente de données volées et outils de hacking sur le darknet **[18].**

l'extension progressive de la carte interbancaire (CIB) et des services de paiement électronique a également augmenté l'exposition aux fraudes numériques.

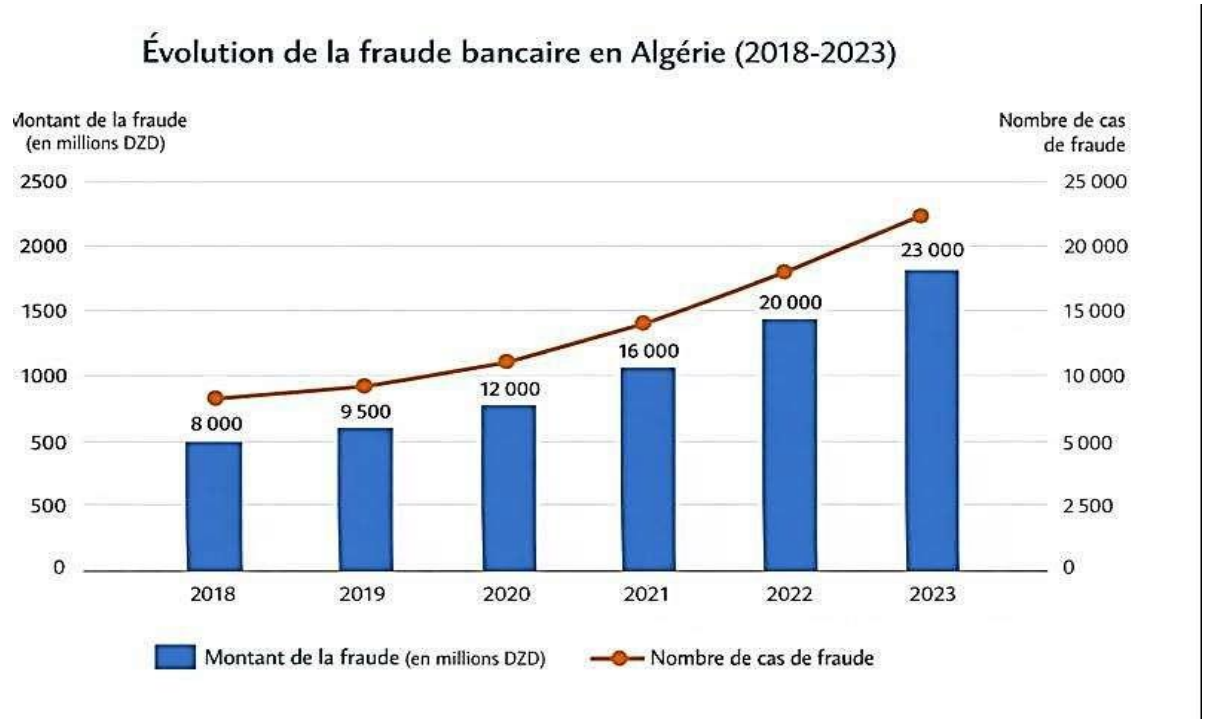


Figure.1.2 Évolution reconstituée de la fraude bancaire en Algérie **(2018–2023)**

Source : élaborée par l'auteur à partir des rapports annuels de la Banque d'Algérie, des statistiques du GIE Monétique et des rapports du CERT-Algeria sur la cybersécurité bancaire**[29]**

"Les données de fraude bancaire en Algérie ne sont pas publiées sous forme de séries statistiques complètes, donc j'ai utilisé une reconstitution basée sur les indicateurs indirects comme l'évolution des paiements électroniques et les rapports de la Banque d'Algérie."

1 .4. Typologies de la fraude bancaire ?

Les fraudes bancaires sont très variées. Elles peuvent être classées selon le canal d'attaque, la cible ou la technique utilisée.

A) Fraude par carte bancaire

C'est l'une des formes les plus répandues dans le monde ainsi que dans les systèmes de paiement modernes **[15].** Elle comprend :

- **Skimming :** copie illégale de la piste magnétique ou des données de la carte.
- **Carding :** utilisation de cartes volées pour effectuer des achats frauduleux.
- **Blanchiment de cartes :** achat de biens facilement revendables.
- **Vol de données sans contact :** capture via des terminaux NFC frauduleux.

En Algérie, cela concerne principalement les cartes CIB et les cartes Edahabia dans certains cas.

B) Fraude par chèque

Malgré la diminution progressive de son utilisation, elle reste présente :

- Faux chèques
- Chèque sans provision
- Chèque volé ou falsifié

Ce type de fraude reste fréquent dans certaines opérations commerciales.

C) Fraude sur les prêts et crédits

Le fraudeur trompe la banque pour obtenir un financement :

- Fausses déclarations de revenus
- Utilisation d'identités fictives
- Multiples crédits obtenus frauduleusement

D) Fraude sur les virements

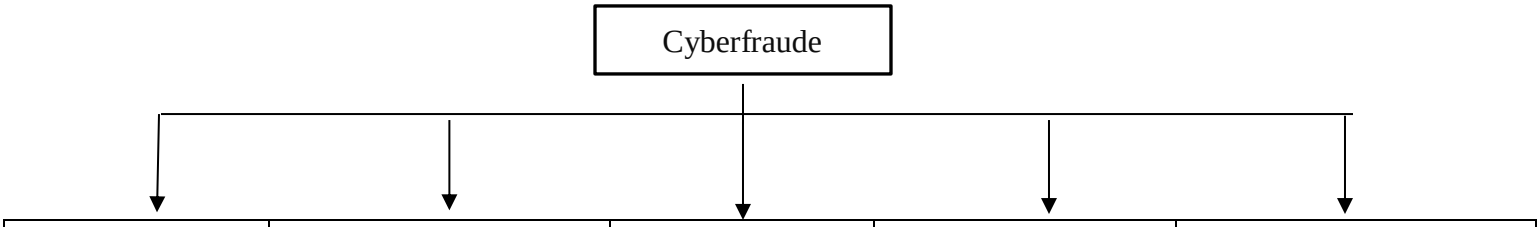
- Usurpation d'identité
- Détournement de virements
- Prise de contrôle de compte via phishing ou malware **[9]**

Cette fraude devient plus fréquente avec le développement des services bancaires à distance.

E) Cyberfraude bancaire

- Phishing
- Smishing
- Vishing
- Malware bancaire
- Man-in-the-browser

Les cyberattaques représentent aujourd'hui une menace importante pour les banques algériennes en raison de la digitalisation progressive du secteur.



Pishing	**Smising**	**Vishing**	**Malware bancaire**	**Man in the Browser**
Emails de hameçonnage	Messages frauduleux	Appels frauduleux	Logiciels malveillants	Man in the Browser 1

Figure.1.3 Types de cyberattaques bancaires

F) Fraude interne

- Création de faux comptes
- Suppression de traces de transactions
- Vente d'informations confidentielles

Cette fraude est particulièrement sensible car elle implique des employés ayant un accès direct au système bancaire.

G) Fraude à l'identité

Elle consiste à utiliser des documents personnels volés ou falsifiés pour ouvrir des comptes ou obtenir des crédits.

H) Fraude mobile et paiement sans contact

- Faux applications bancaires
- Attaques NFC
- Détournement de sessions

Avec le développement des applications bancaires mobiles en Algérie, ce type de fraude devient de plus en plus important.

Tableau.1.1 Récapitulatif des typologies

Typologie	Moyen utilisé	Exemple
Carte bancaire	Skimming, carding	Copie de carte CIB
Chèque	Faux chèque	Chèque falsifié
Crédit	Faux revenus	Prêt frauduleux
Virement	Usurpation	Virement falsifié
Cyberattaque	Phishing, malware	Email frauduleux
Interne	Employé malveillant	Création de faux comptes
Identité	Vol de documents	CNI falsifiée
Mobile	Fake apps	Application bancaire frauduleuse

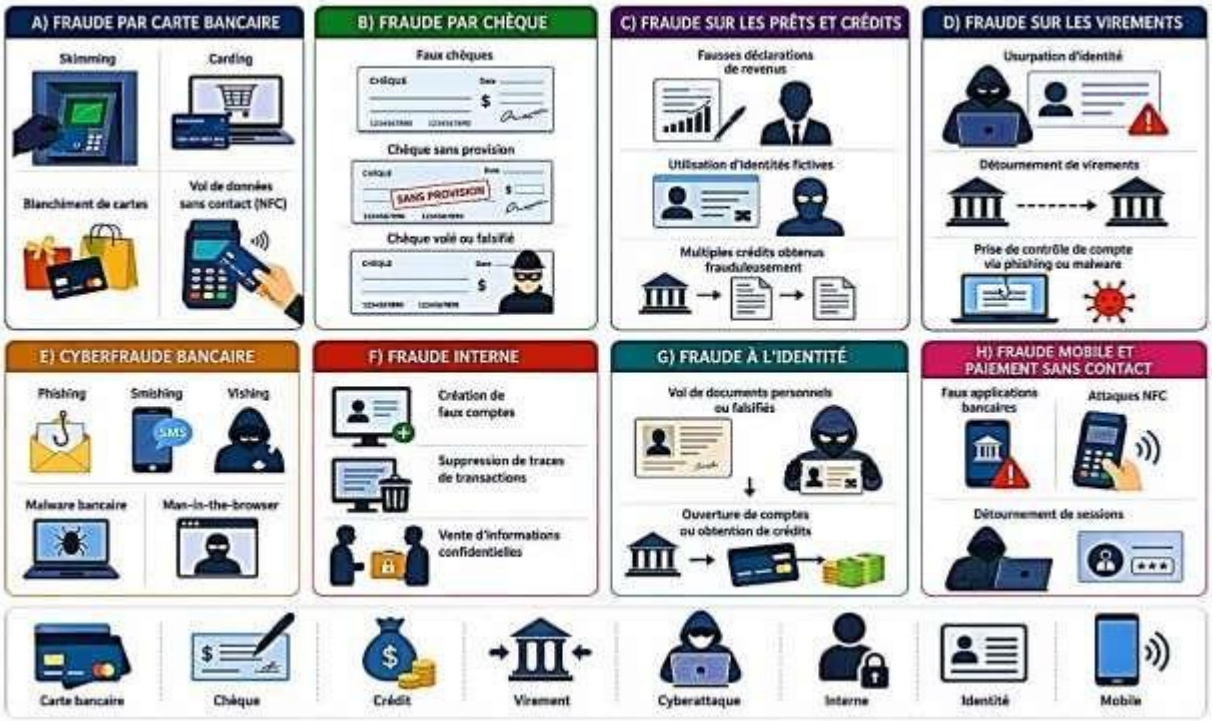


Figure .1.4 Représentation visuelle des principales typologies de la fraude bancaire

1.5. Facteurs et impacts de la fraude bancaire

1.5.1. Facteurs de la fraude bancaire

La fraude bancaire résulte d'un ensemble de facteurs interdépendants qui évoluent avec la transformation numérique du secteur financier. Ces facteurs peuvent être regroupés en quatre catégories principales.

A) Facteurs technologiques

L'évolution rapide des systèmes bancaires numériques a introduit de nouvelles vulnérabilités exploitables par les fraudeurs :

La **complexité croissante des infrastructures informatiques** augmente la surface d'attaque.

- Les **transactions en temps réel** réduisent le temps de vérification et facilitent les fraudes instantanées.

- L'**interconnexion internationale des systèmes bancaires** complique la traçabilité des opérations frauduleuses.
- L'**utilisation de systèmes legacy (anciens systèmes)** augmente les risques liés aux failles de sécurité non corrigées. **[24]**

B) Facteurs humains

Le facteur humain reste universellement l'un des principaux vecteurs de fraude :

- Manque de formation en cybersécurité.
- Utilisation de mots de passe faibles.
- Techniques d'ingénierie sociale (phishing, smishing).
- Erreurs humaines internes.

Selon plusieurs études récentes, plus de 70 % des incidents de fraude impliquent une erreur humaine ou une manipulation sociale **[2].**

- **Cas de l'Algérie :**

  Selon plusieurs rapports du **CERT-Algeria**, les attaques par phishing visant les utilisateurs de services bancaires mobiles sont en augmentation, notamment à travers de faux SMS imitant des banques locales.

C) Facteurs organisationnels

Les structures internes des institutions bancaires peuvent également favoriser les fraudes :

- Absence de coordination entre les différents services (silos d'information).
- Faible maturité des politiques de gouvernance des risques.
- Pression commerciale pouvant réduire les contrôles de sécurité.
- Insuffisance des mécanismes de signalement interne. **[21]**

D) Facteurs externes

Les menaces externes jouent un rôle majeur dans la prolifération de la fraude :

- Organisation de réseaux criminels structurés à l'échelle internationale.
- Développement du **Dark Web** facilitant l'achat de données bancaires volées.
- Failles réglementaires entre pays.
- Périodes de crise économique augmentant les comportements frauduleux. **[24]**

E) Facteurs externes

Les menaces externes jouent un rôle majeur dans la prolifération de la fraude :

- Organisation de réseaux criminels structurés à l'échelle internationale.
- Développement du **Dark Web** facilitant l'achat de données bancaires volées.
- Failles réglementaires entre pays.
- Périodes de crise économique augmentant les comportements frauduleux. **[24]**

1.5.2. Impacts de la fraude bancaire

La fraude bancaire génère des conséquences multidimensionnelles affectant les individus, les institutions et l'économie globale.

A) Impacts financiers

- **Clients :** pertes financières directes et procédures de remboursement complexes.
- **Banques :** coûts opérationnels élevés (enquêtes, remboursements, sanctions).
- **Économie :** pertes estimées à plusieurs milliards de dollars annuellement dans le secteur bancaire mondial.

Les systèmes modernes basés sur l'IA atteignent aujourd'hui des performances élevées, avec des taux de rappel supérieurs à 90 % dans certaines approches avancées **[20]**.

B) Impacts sur la réputation

- Perte de confiance des clients.
- Dégradation de l'image de marque.
- Migration des clients vers des institutions plus sécurisées.

C) Impacts juridiques et réglementaires

- Sanctions financières imposées par les autorités de régulation.
- Obligation de mise en conformité (RGPD, DSP2, NIS2). **[22]**
- Responsabilité juridique en cas de négligence avérée.

- **Cas de l'Algérie :**

  La régulation bancaire est assurée principalement par :

  - Banque d'Algérie
  - Commission Bancaire
  - Loi algérienne sur la cybersécurité et la protection des données personnelles (Loi 18-07)

D) Impacts sociaux et psychologiques

- Stress et anxiété des victimes.
- Augmentation de la charge de travail pour les services bancaires.
- Diminution de la confiance dans le système financier.

Tableau.1.2 Synthèse des impacts de la fraude bancaire

Partie prenante	Impact principal	Conséquence
Clients	Perte financière	Stress, méfiance
Banques	Coûts élevés	Réputation affectée
Régulateurs	Contrôle renforcé	Pression institutionnelle
État	Instabilité économique	Perte de confiance

1.6. Évolution des tendances de fraude

Les tendances récentes montrent une transformation des techniques frauduleuses :

- Forte augmentation des fraudes **Card-Not-Present (CNP)**.
- Utilisation de micro-transactions pour éviter la détection.
- Exploitation des périodes de crise (ex : COVID-19).
- Automatisation des attaques via des scripts et bots. **[24]**

1.7. Cadre réglementaire et normatif :

La lutte contre la fraude bancaire repose sur un cadre réglementaire international strict et en constante évolution.

1.7.1. Réglementations internationales :

A. Accords de Bâle (Bâle III) [14]

Les accords de Bâle sont des normes internationales élaborées par le Comité de Bâle sur le contrôle bancaire. Ils imposent aux banques :

- La gestion du risque opérationnel
- La constitution de fonds propres pour couvrir les pertes liées à la fraude
- L'amélioration des systèmes de contrôle interne

B. Directive DSP2 (PSD2) [22]

La DSP2 (Deuxième Directive sur les Services de Paiement) est une directive européenne qui vise à renforcer la sécurité des paiements électroniques. Elle introduit :

- L'authentification forte du client (SCA)
- La surveillance des transactions à risque
- L'obligation de reporting des incidents

C. RGPD [23]

Le RGPD (Règlement Général sur la Protection des Données) est un règlement européen qui protège les données personnelles. Il impose :

- La minimisation des données collectées
- La transparence des traitements automatisés
- La protection des données personnelles (anonymisées ou pseudonymisées pour l'IA

D. NIS2 [25]

La directive NIS2 (Network and Information Security) est une directive européenne sur la sécurité des réseaux et de l'information. Elle renforce :

- La cybersécurité des infrastructures critiques
- L'obligation de déclaration des incidents
- Les audits de sécurité réguliers

1.7.2. Cadre réglementaire national

En Algérie, la régulation du secteur bancaire et de la cybersécurité repose sur plusieurs institutions et textes officiels :

A) Banque d'Algérie

Elle constitue l'autorité centrale de régulation bancaire et impose :

- Normes de sécurité des systèmes d'information.
- Contrôle des transactions électroniques.
- Supervision des banques commerciales.

B) Commission Bancaire

Elle assure :

- Le contrôle prudentiel des banques.
- La surveillance des risques opérationnels incluant la fraude.

C) Loi 18-07 sur la protection des données personnelles [26]

Cette loi encadre :

- La collecte des données clients.
- L'utilisation des données dans les systèmes informatiques.
- La protection contre les abus et accès non autorisés.

D) CERT-Algeria [27]

Le CERT-Algeria (Computer Emergency Response Team) est un organisme national chargé de la surveillance, de la détection et de la réponse aux incidents de cybersécurité affectant les infrastructures numériques nationales.

1.7.3. Normes techniques

- **ISO 27001 [24]**

La norme ISO 27001 est une norme internationale de gestion de la sécurité des systèmes d'information permettant aux organisations de protéger leurs données sensibles à travers des politiques de sécurité et des contrôles adaptés**.**

- **PCI DSS [28]**

Le standard PCI DSS (Payment Card Industry Data Security Standard) est un ensemble de règles de sécurité destiné à protéger les données des cartes bancaires et sécuriser les transactions électroniques

Cas Algérie :

Les grandes banques algériennes commencent progressivement à adopter :

- ISO 27001 pour la sécurité des SI
- PCI DSS pour les paiements par carte CIB

1.8. Acteurs du contrôle de la fraude bancaire

La lutte contre la fraude repose sur une collaboration multi-acteurs.

1.8.1 Acteurs internes

- Direction conformité et risque.
- Service sécurité des systèmes d'information (SSI).
- Cellule antifraude.
- Service juridique.
- Service client.
- Direction générale.

1.8.2 Acteurs externes

- Banque d'Algérie.
- Commission Bancaire.
- CERT-Algeria.
- Police spécialisée en cybercriminalité.

Tableau.1.3. Acteurs externes de la lutte contre la fraud

Catégorie	Monde	Algérie	Rôle
Régulation bancaire	Régulateurs bancaires	Banque d'Algérie	Contrôle des banques
Justice	Autorités judiciaires	Tribunaux	Traitement des fraudes
Sécurité	Interpol / Europol	Police / Gendarmerie	Enquête
Technologie	Fournisseurs technologiques	Entreprises IT	Détection de fraude
Consommateurs	Associations de consommateurs	Associations locales	Protection des clients

1.9. Conclusion

La fraude bancaire représente un phénomène complexe et évolutif, qui touche l'ensemble des systèmes financiers modernes. À travers ce chapitre, nous avons présenté les principales définitions de la fraude, ses différentes typologies ainsi que les facteurs qui favorisent son développement.

Nous avons également montré que la fraude bancaire engendre des impacts importants, aussi bien sur les institutions financières que sur les clients et l'économie. Par ailleurs, la lutte contre ce phénomène repose sur un cadre réglementaire strict et sur la collaboration de plusieurs acteurs, aux niveaux national et international.

Cependant, malgré les efforts mis en place, les méthodes traditionnelles de détection montrent aujourd'hui certaines limites face à la sophistication croissante des techniques frauduleuses. Cette situation nécessite l'intégration de solutions technologiques plus avancées, notamment basées sur l'intelligence artificielle. Des travaux récents confirment que les approches d'IA, bien que prometteuses, doivent encore être améliorées pour faire face à la sophistication croissante des fraudes dans les réseaux financiers [11]..

Dans ce contexte, le chapitre suivant sera consacré à l'étude des différentes méthodes de détection de la fraude, en mettant l'accent sur les approches modernes et les techniques d'apprentissage automatique, telles que l'algorithme AdaBoost.

Chapitre 02:

Les Méthodes de détection de la fraude : approches traditionnelles et technologiques

2.1. Introduction

L'évolution des technologies numériques a profondément transformé le secteur bancaire. Les services bancaires en ligne et les paiements électroniques ont facilité les transactions financières, mais ils ont également augmenté les risques de fraude bancaire **[2].** Face à cette évolution, les banques doivent mettre en place des systèmes efficaces capables de détecter rapidement les opérations suspectes et de limiter les pertes financières.

Les méthodes traditionnelles de contrôle, basées principalement sur la vérification humaine et les procédures internes, présentent aujourd'hui plusieurs limites face au volume important des transactions et à la sophistication des techniques frauduleuses **[30]**. Pour cette raison, les institutions financières utilisent de plus en plus les technologies de l'intelligence artificielle et du Machine Learning (ML) afin d'améliorer les performances des systèmes de détection de fraude **[31]**.

Ainsi, ce chapitre présente d'abord les méthodes traditionnelles de contrôle et leurs limites. Ensuite, il expose l'apport des nouvelles technologies dans la détection de fraude bancaire. Enfin, il présente l'algorithme AdaBoost et son application dans les systèmes intelligents de détection de fraude.

2.2. Les méthodes traditionnelles de contrôle et leurs limites

2.2.1. Les dispositifs classiques de contrôle bancaire

Avant l'apparition des systèmes intelligents, les banques utilisaient principalement des mécanismes de contrôle interne reposant sur la supervision humaine, les procédures administratives et les audits périodiques. Ces dispositifs avaient pour objectif de réduire les risques opérationnels, les erreurs comptables et les actes de fraude **[32]**.

Le dispositif traditionnel de contrôle bancaire repose généralement sur plusieurs niveaux complémentaires :

A. Le contrôle hiérarchique

Le contrôle hiérarchique représente la première ligne de défense contre la fraude bancaire.

Il est assuré quotidiennement par les responsables d'agence et les chefs de services **[30].** Ce type de contrôle comprend notamment :

- La vérification des opérations sensibles.
- La validation des transactions importantes.
- Le suivi des activités des employés.
- La surveillance des mouvements financiers inhabituels.

Cette approche permet de détecter certaines anomalies simples, notamment les erreurs administratives et les opérations suspectes réalisées au sein de l'agence.

B. Le contrôle permanent

Le contrôle permanent correspond à un ensemble de vérifications réalisées de manière continue afin d'assurer le respect des procédures internes et des règles réglementaires **[33]**.

Il permet :

- D'évaluer les risques opérationnels.
- De contrôler la conformité des opérations.
- De surveiller les transactions financières.
- De limiter les violations des procédures internes.

Dans la plupart des établissements financiers, cette fonction est assurée par des départements spécialisés dans la gestion des risques et la conformité bancaire.

C. L'audit interne

L'audit interne constitue une fonction indépendante chargée d'évaluer l'efficacité des systèmes de contrôle et de gestion des risques **[34]**.

Les auditeurs examinent périodiquement les opérations bancaires afin d'identifier les faiblesses organisationnelles, les anomalies comptables et les risques de fraude.

L'audit interne contribue notamment à :

- Vérifier l'application des procédures.
- Analyser les risques opérationnels.
- Détecter les anomalies comptables.
- Proposer des mesures correctives.

D. Le contrôle comptable

Le contrôle comptable vise à garantir la fiabilité des données financières et des écritures comptables **[30]**.

Il permet de détecter certaines irrégularités à travers :

- Les rapprochements bancaires.
- La vérification des soldes.
- Le suivi des comptes internes.
- L'analyse des écarts comptables.

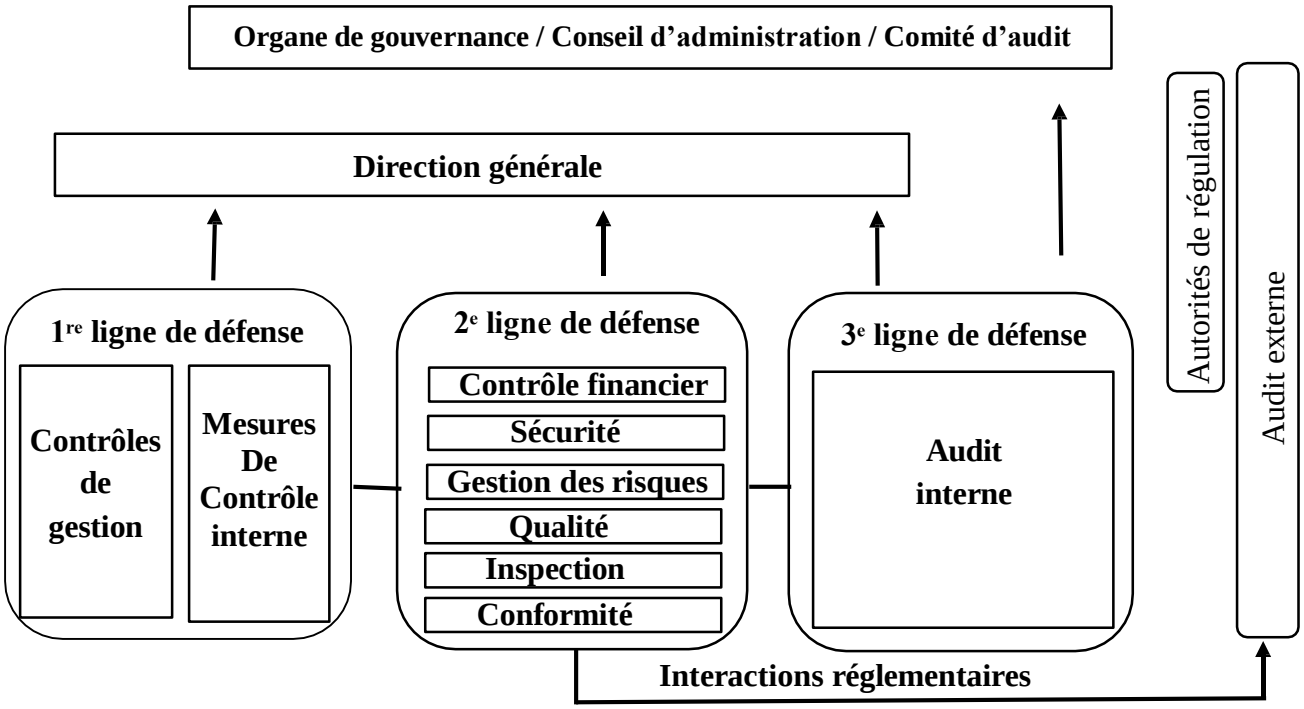


Figure.2.1 Structure traditionnelle du contrôle bancaire. [35]

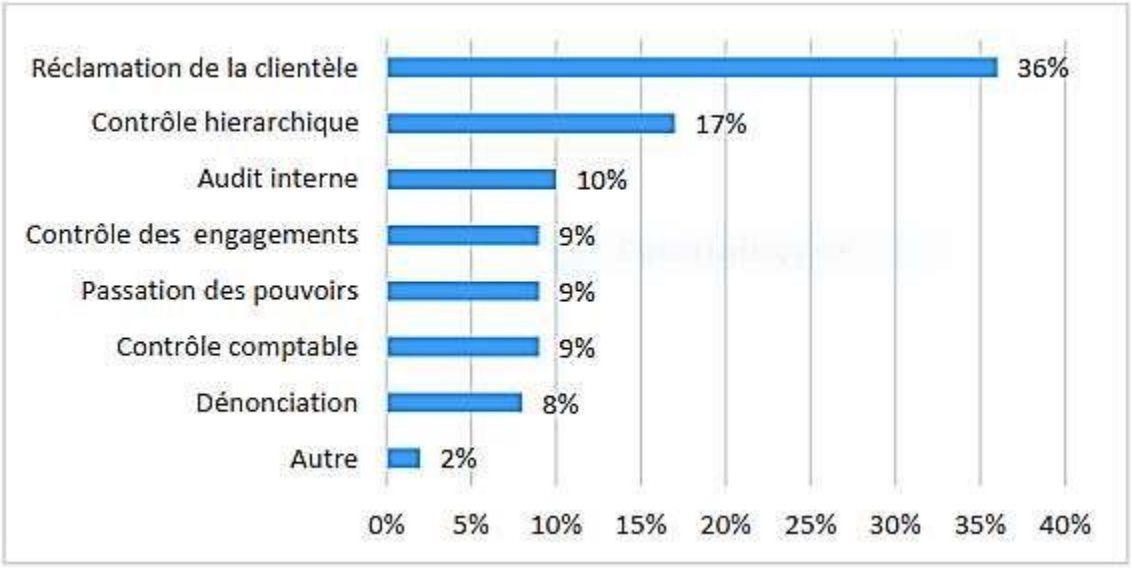


Figure.2.2 Elément déclencheur de la fraude **[36]**

2.2.2. Les limites des méthodes traditionnelles :

Malgré leur importance dans les établissements financiers, les approches traditionnelles de contrôle présentent plusieurs insuffisances face à l'évolution de la fraude bancaire moderne **[30]**.

A. La lenteur des procédures de détection

Les contrôles manuels nécessitent souvent un temps important pour analyser les opérations bancaires. Le délai de découverte des fraudes peut parfois atteindre plusieurs mois **[32],** ce qui permet aux fraudeurs de dissimuler leurs activités et de multiplier les transactions frauduleuses.

Cette lenteur réduit considérablement la capacité des banques à réagir rapidement aux attaques financières.

B. L'incapacité à traiter de grands volumes de données

Les banques modernes traitent quotidiennement des millions de transactions. Les méthodes classiques ne peuvent pas analyser efficacement un volume aussi important de données en temps réel **[9]**. Cette limitation réduit fortement la capacité des contrôleurs à détecter rapidement les comportements suspects et les schémas frauduleux complexes.

C. Le contrôle par échantillonnage

Les audits internes reposent souvent sur des techniques d'échantillonnage qui ne couvrent qu'une partie limitée des transactions, Par conséquent, certaines fraudes peuvent ne pas être détectées. De plus, les systèmes manuels dépendent fortement de l'intervention humaine, ce qui augmente les risques d'erreurs, de négligence ou de collusion interne. **[37]**.

D. L'absence de détection automatisée

Les approches classiques reposent principalement sur des contrôles humains et des procédures statiques. Elles ne permettent pas d'identifier automatiquement les comportements suspects ou les anomalies complexes en temps réel. Cette absence d'automatisation représente une faiblesse majeure face à la sophistication croissante des techniques de fraude bancaire.

E. L'évolution des techniques de fraude

Les fraudeurs utilisent aujourd'hui des technologies avancées telles que le phishing, les logiciels malveillants, le vol de données bancaires et les attaques informatiques sophistiquées **[26]**. Face à ces nouvelles menaces, les approches traditionnelles deviennent insuffisantes pour assurer une protection efficace des systèmes bancaires modernes.

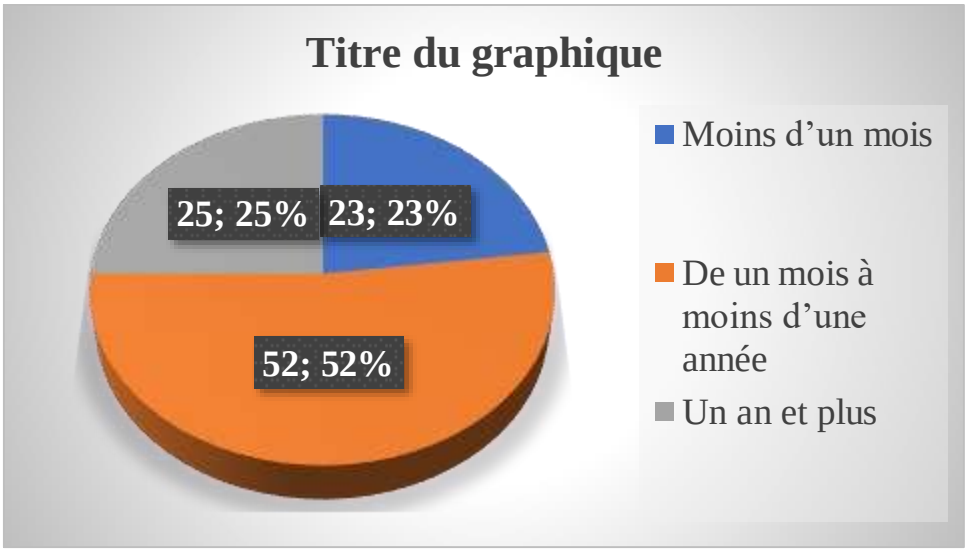


Figure.2.3. Délai de découverte de la fraude **[36]**

Tableau.2.1 Limites des approches traditionnelles

Méthode traditionnelle	Limites principales
Contrôle hiérarchique	Dépend fortement de l'intervention humaine
Audit interne	Vérification périodique et non continue
Contrôle comptable	Détection tardive des anomalies
Contrôle permanent	Difficulté à analyser de grands volumes de données

2.3. L'apport des nouvelles technologies dans la détection de la fraude

2.3.1. La transformation numérique du secteur bancaire

La digitalisation du secteur bancaire a profondément modifié les méthodes de gestion des opérations financières **[38]**.

Les banques utilisent aujourd'hui :

- les plateformes numériques ;
- les applications mobiles ;
- les paiements électroniques ;
- les services cloud ;
- les API bancaires.

Cette transformation améliore la rapidité et l'accessibilité des services financiers. Toutefois, elle augmente également les risques liés à la fraude et à la cyber sécurité.

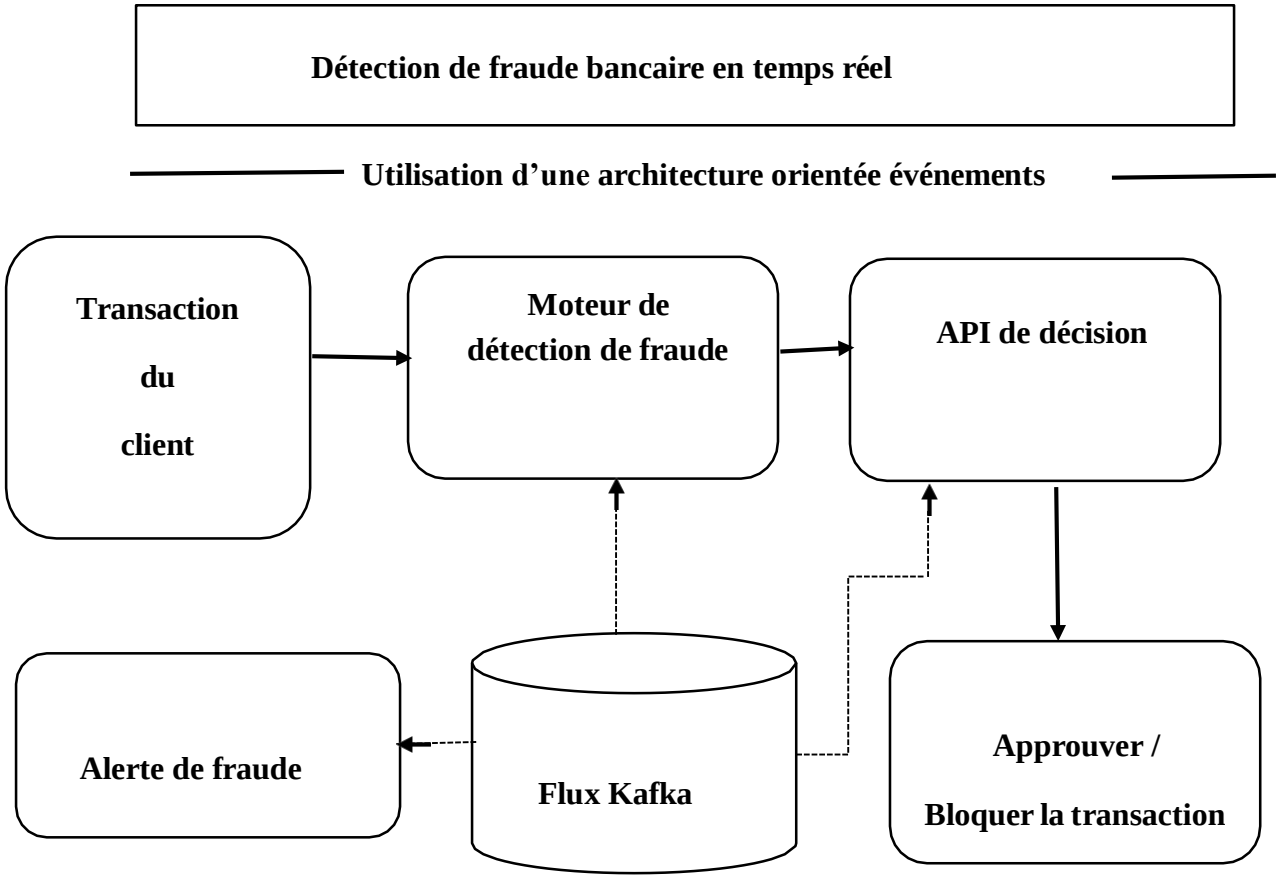


Figure.2.4 Architecture d'un système de détection de fraude bancaire en temps réel. **[39]**

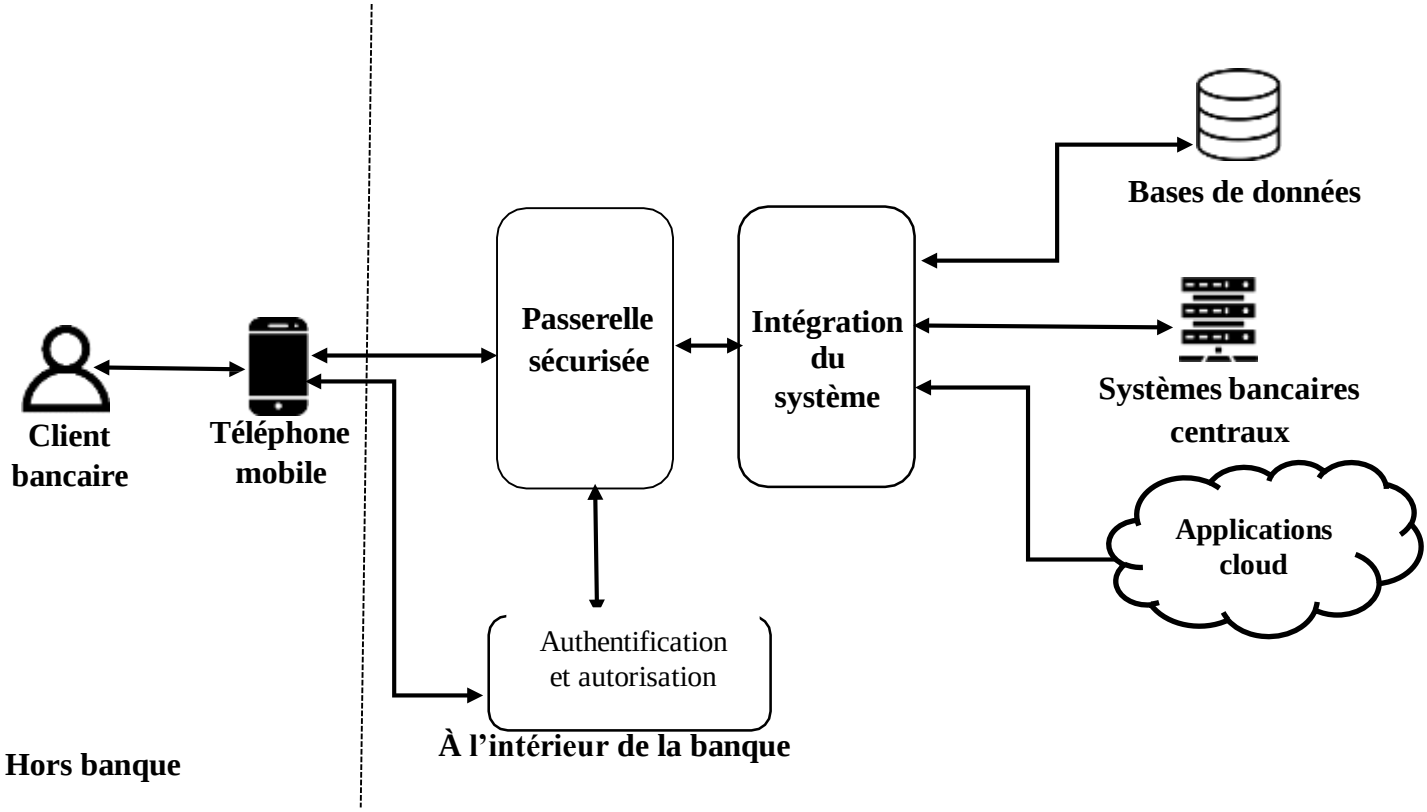


Figure.2.5 Architecture moderne d'un système bancaire numérique **[40]**

2.3.2. Le rôle du Data Mining dans la détection de fraude

Le Data Mining désigne l'ensemble des techniques permettant d'extraire automatiquement des connaissances à partir de grandes bases de données **[41]**.

Dans le domaine bancaire, il permet :

- D'identifier des comportements suspects.
- De détecter des modèles cachés.
- D'analyser les habitudes des clients.
- De prévoir les risques de fraude.

Les banques utilisent ces techniques pour construire des systèmes intelligents capables de surveiller les transactions en temps réel.

2.3.3. Le Machine Learning dans la détection de fraude

Le Machine Learning représente une branche de l'intelligence artificielle permettant aux systèmes informatiques d'apprendre automatiquement à partir des données **[42]**.

Contrairement aux systèmes basés uniquement sur des règles fixes, les algorithmes de Machine Learning peuvent :

- Améliorer progressivement leurs performances.
- S'adapter aux nouveaux comportements frauduleux.
- Détecter des anomalies complexes.

Les principales techniques utilisées dans la détection de fraude sont les méthodes supervisées, non supervisées et les modèles de Deep Learning.

A. Les méthodes supervisées

Les méthodes supervisées utilisent des données étiquetées contenant des transactions normales et des transactions frauduleuses.

Parmi les algorithmes les plus utilisés :

- Régression logistique **[43]**
- Decision Tree **[44]**
- Random Forest **[45]**
- AdaBoost **[46]**
- XGBoost **[47]**
- Support Vector Machine (SVM) **[48]**

B. Les méthodes non supervisées

Ces techniques sont utilisées lorsque les données frauduleuses ne sont pas clairement identifiées **[37]**.

Elles permettent principalement :

- La détection d'anomalies.
- L'identification des comportements inhabituels.
- L'analyse des clusters suspects.

C. Le Deep Learning

Les réseaux de neurones profonds sont capables d'analyser des structures complexes et des volumes massifs de données **[49]**.

Ils sont particulièrement utilisés dans :

- La détection de fraude en ligne.
- L'analyse comportementale.
- La reconnaissance de schémas complexes.

Tableau 2.2 Comparaison des techniques de Machine Learning

Technique	Avantages	Limites
Régression logistique	Simple et interprétable	Performance limitée
Random Forest	Bonne précision	Complexité élevée
AdaBoost	Bonne gestion des classes déséquilibrées	Sensible au bruit
Deep Learning	Très performant	Nécessite beaucoup de données

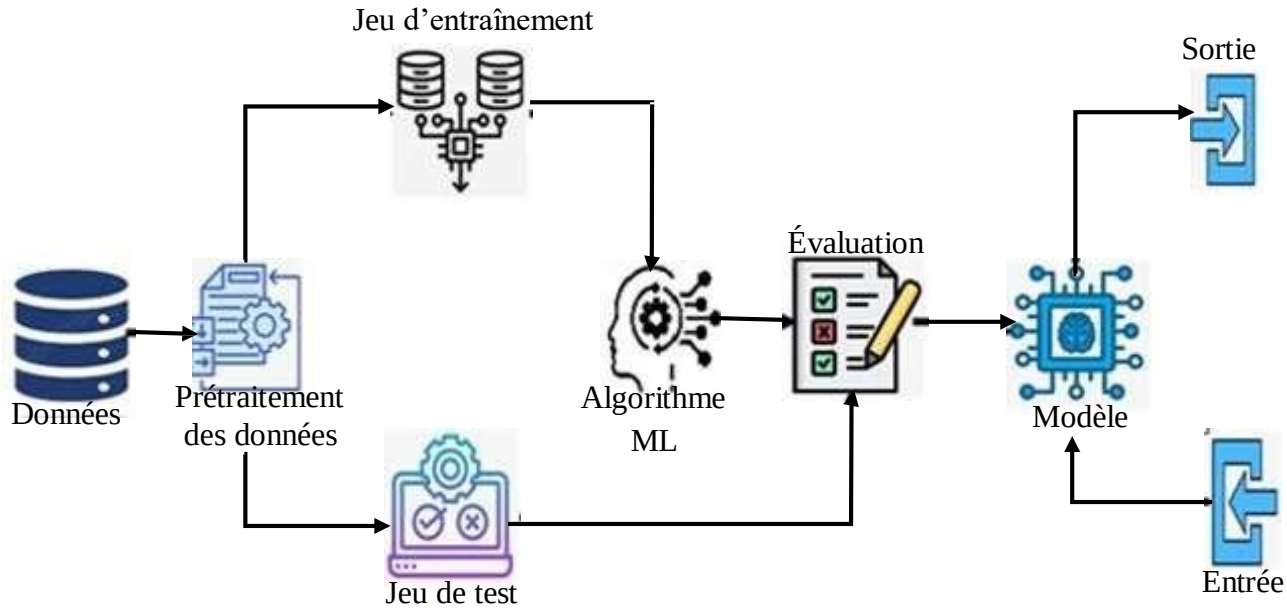


Figure.2.6 Processus général de détection de fraude par Machine Learning **[50]**

2.3.4. Les avantages des systèmes intelligents :

Les systèmes modernes de détection de fraude présentent plusieurs avantages importants :

A. Détection en temps réel

Les algorithmes intelligents peuvent analyser instantanément les transactions bancaires et générer des alertes automatiques **[51]**.

B. Réduction des faux positifs

Les modèles avancés permettent d'améliorer la précision des alertes et de limiter les blocages inutiles des transactions légitimes.

C. Adaptation continue

Les modèles de Machine Learning peuvent être réentraînés régulièrement afin de s'adapter aux nouvelles stratégies des fraudeurs.

D. Automatisation des analyses

Les systèmes intelligents réduisent considérablement la charge de travail des analystes humains.

2.4. Présentation de l'algorithme AdaBoost et son application à la détection de fraude

2.4.1. Présentation générale d'AdaBoost :

AdaBoost (Adaptive Boosting) est un algorithme de Machine Learning proposé par Freund et Schapire en 1997 **[52]**.

Il appartient à la famille des méthodes d'apprentissage ensembliste (Ensemble Learning).

Le principe d'AdaBoost consiste à combiner plusieurs classificateurs simples afin de construire un modèle plus performant.

Le fonctionnement général de l'algorithme repose sur :

1. L'entraînement d'un premier modèle simple.
2. L'identification des erreurs de classification.
3. L'augmentation du poids des observations mal classées.

4. La création d'un nouveau modèle concentré sur les erreurs précédentes.
5. La combinaison finale des différents modèles.

Cette approche permet d'améliorer progressivement les performances du système de classification.

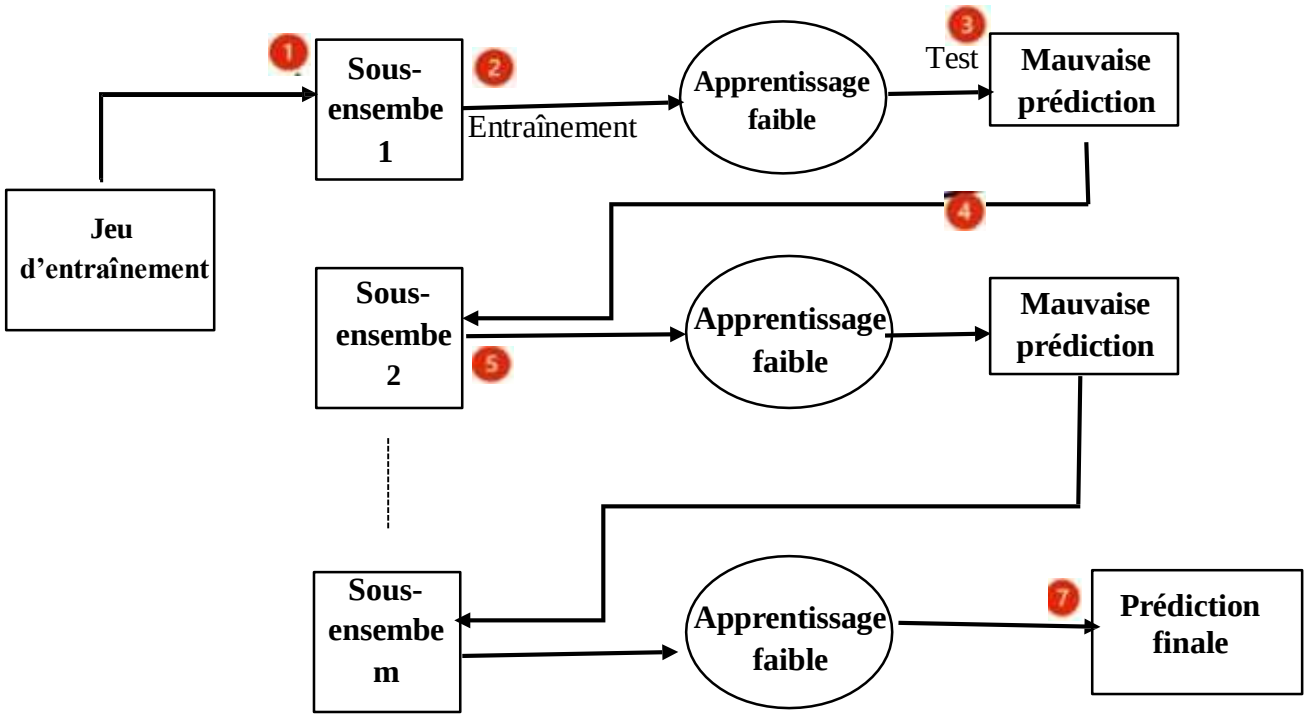


Figure.2.7 Fonctionnement général de l'algorithme AdaBoost**[53]**

2.4.2. Principe mathématique d'AdaBoost

Le modèle AdaBoost attribue un poids à chaque observation du dataset. Les transactions mal classées reçoivent progressivement des poids plus importants afin d'attirer l'attention des classificateurs suivants.

Le classificateur final est obtenu par combinaison pondérée des classificateurs faibles.

Le calcul du poids du classificateur est donné par :

$$\alpha_t = \frac{1}{2} ln \left(\frac{1 - s_t}{s_t}\right) \quad \textbf{(1)}$$

où :

- ε_t représente le taux d'erreur du classificateur ;
- α_t représente l'importance du classificateur dans la décision finale.

La prédiction finale du modèle est donnée par :

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t \, h_t(x)\right) \quad \textbf{(2)}$$

Cette stratégie permet d'améliorer progressivement la capacité du modèle à détecter les transactions frauduleuses.

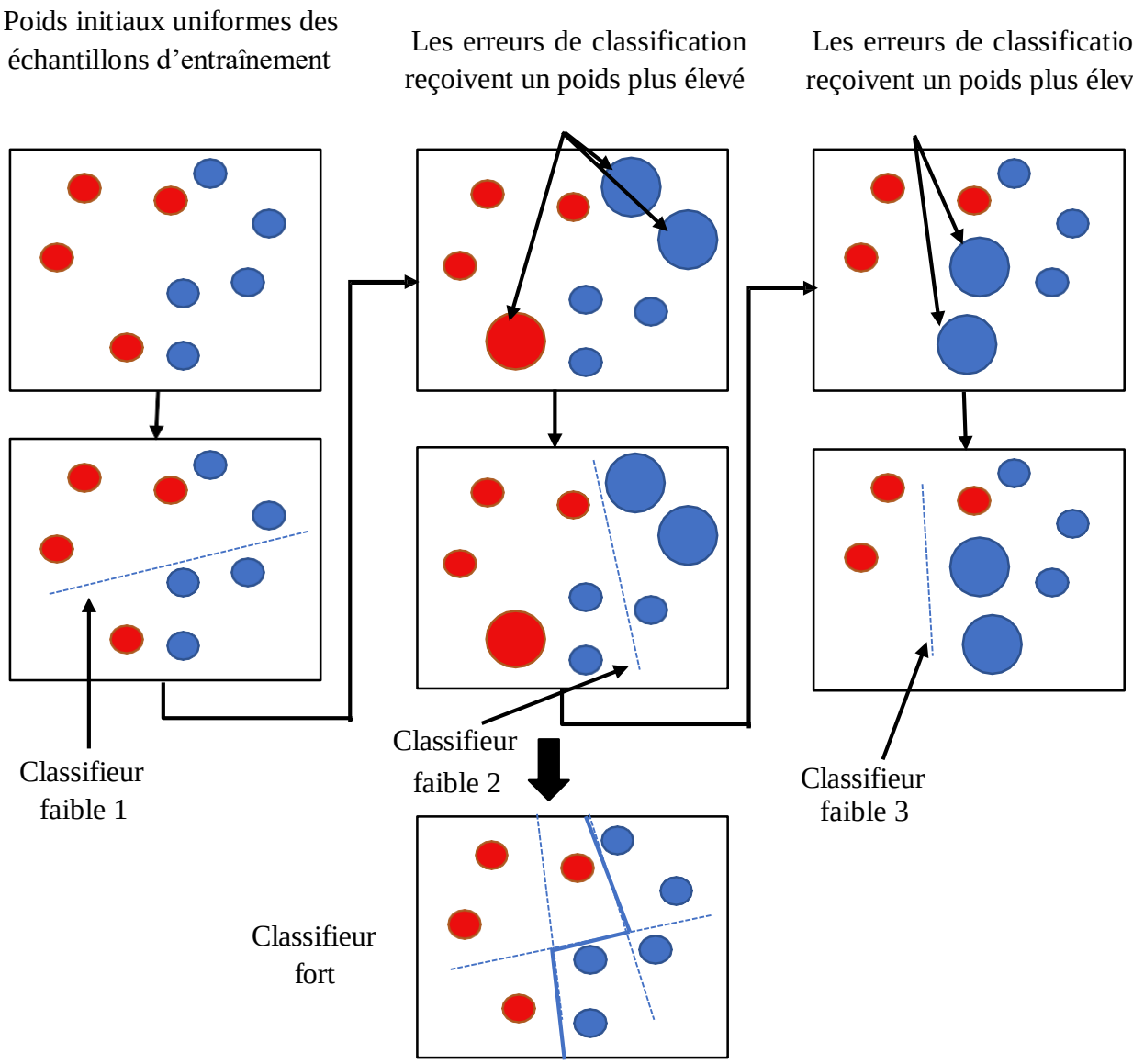


Figure.2.8 Illustration de l'algorithme AdaBoost **[54]**

L'algorithme AdaBoost peut être résumé comme suit :

Algorithme

Entrée : Étant donné N échantillons d'apprentissage :

$X = \{(x_1, y_1), (x_2, y_2) \ldots, (x_N, y_N)\}$

où x_i représente le i ème échantillon et $y_i \in \{-1, 1\}$ est l'étiquette de classe pour l'échantillon i .

Sortie : Le classifieur final est défini par :

$$sign\ (H_{final}(x))\ , \begin{cases} if\ sign\ (H_{final}(x)) < 0, & h_t(x_i)(y_{i_{pred}}) = -1 \\ if\ sign\ (H_{final}(x)) > 0, & h_t(x_i)(y_{i_{pred}}) = +1 \end{cases}$$

Étapes de l'algorithme

1. Initialiser les paramètres du classifieur AdaBoost :
2. $w_i^1 = \frac{1}{N}, j = 1, \ldots, N$

2. Pour $t = 1$ à T , effectuer les étapes suivantes :

3. Prendre un échantillon D_t à partir de (X) en utilisant la distribution w^t

4. Utiliser la distribution D_t pour entraîner le classifieur faible ($h_t(x_i)$) avec une erreur minimale :

$$\epsilon_t = \sum_{i=1}^{N} w^t I(h_t(x_i) \neq y_i)$$

où :

$$h_t(x_i) \in \{-1,1\}$$

et I représente la fonction de perte $0-1$ définie par :

$I(h_t(x_i), y_i) = 1,\ si \quad h_t(x_i) = y_i$ et -1 sinon.

5. Tant que :

$\epsilon_t >= 0.5$

6. Réinitialiser les poids à :

$w_i^t = \frac{1}{N}, j = 1, \dots, N.$

7. Recalculer ϵ_t

8. Fin Tant que

9. Calculer le poids de chaque classifieur faible :

$$\alpha_t = \frac{1}{2} \ln(\frac{1 - \epsilon_t}{\epsilon_t})$$

10. Mettre à jour les poids des échantillons d'apprentissage :

$$w^{t+1} = \frac{w_i^t \, exp^{-\alpha_t y_i h_t(x_i)}}{\sum_{i=1}^{N} w^t \, exp^{-\alpha_t y_i h_t(x_i)}},$$

où :

w_i^{t+1} représente le poids de l'échantillon i dans l'itération suivante.

11. Fin pour .

12. Le classifieur final AdaBoost est défini par :

$H_T = sign\,(H_{final}(x)) = sign(\sum_{t=1}^{T} \alpha_t h_t(x))$

Ht(x) représente le classificateur faible (Weak Learner) à l'itération t

2.4.3. Application d'AdaBoost à la fraude bancaire :

Dans le domaine bancaire, l'algorithme AdaBoost est principalement utilisé pour classifier les transactions financières en deux catégories distinctes : les transactions normales et les transactions frauduleuses. Grâce à sa capacité d'apprentissage adaptatif, ce modèle améliore progressivement la détection des comportements suspects en accordant davantage d'importance aux observations difficiles à classifier.

Chaque transaction bancaire est représentée par plusieurs variables descriptives telles que le montant de la transaction, l'heure de l'opération, la localisation géographique, la fréquence des opérations, le type de transaction ainsi que le comportement habituel du client. Ces caractéristiques

permettent au modèle d'identifier les schémas normaux et les comportements potentiellement frauduleux. **[31]**.

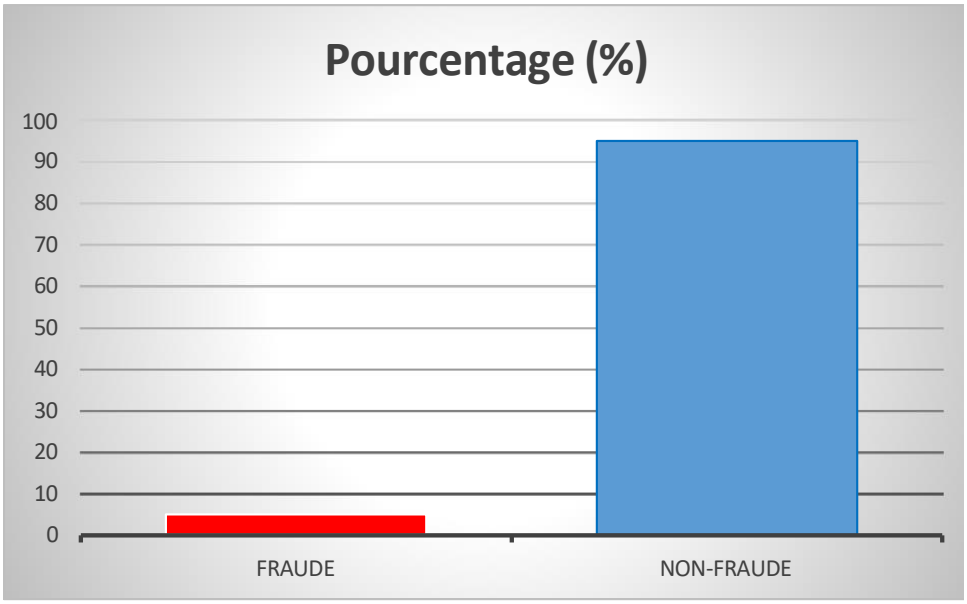


Figure.2.9 Déséquilibre des classes dans un dataset de fraude **[55]**

Le déséquilibre des classes représente l'un des principaux défis dans les systèmes de détection de fraude bancaire. En effet, les transactions frauduleuses représentent généralement une proportion très faible par rapport au nombre total des transactions. Cette situation rend la tâche de classification plus complexe pour les modèles de Machine Learning traditionnels **[56]**.

AdaBoost est particulièrement adapté à cette problématique puisqu'il accorde davantage d'importance aux observations mal classées au cours des différentes itérations d'apprentissage.

2.5. Les métriques d'évaluation du modèle

L'évaluation d'un modèle (Model Evaluation) est le processus qui consiste à utiliser différentes métriques afin d'analyser les performances d'un modèle de machine learning, ainsi que ses points forts et ses limites. Cette étape est essentielle pour vérifier si le modèle est capable de produire des prédictions fiables sur de nouvelles données non utilisées pendant l'apprentissage.

L'évaluation est particulièrement importante dans les systèmes de détection de fraude bancaire, car elle permet de mesurer la capacité du modèle à distinguer correctement les transactions normales des transactions frauduleuses. Elle aide également à détecter les problèmes de surapprentissage (Overfitting) et de sous-apprentissage (Underfitting), qui représentent deux causes majeures des mauvaises performances des algorithmes de Machine Learning.

Dans les projets de Machine Learning, les performances d'un modèle peuvent être influencées par plusieurs problèmes d'apprentissage. Le surapprentissage (Overfitting) apparaît lorsque le modèle apprend excessivement les données d'entraînement, y compris le bruit et les détails inutiles, ce qui réduit sa capacité de généralisation sur de nouvelles données. À l'inverse, le sous-apprentissage (Underfitting) correspond à une situation dans laquelle le modèle ne parvient pas à capturer correctement la structure des données, entraînant ainsi de faibles performances aussi bien sur les données d'apprentissage que sur les données de test. La situation idéale, appelée Right Fit, est atteinte lorsque le modèle présente un bon équilibre entre apprentissage et généralisation.

Dans les problèmes de classification, plusieurs métriques sont utilisées afin d'évaluer les performances du modèle.

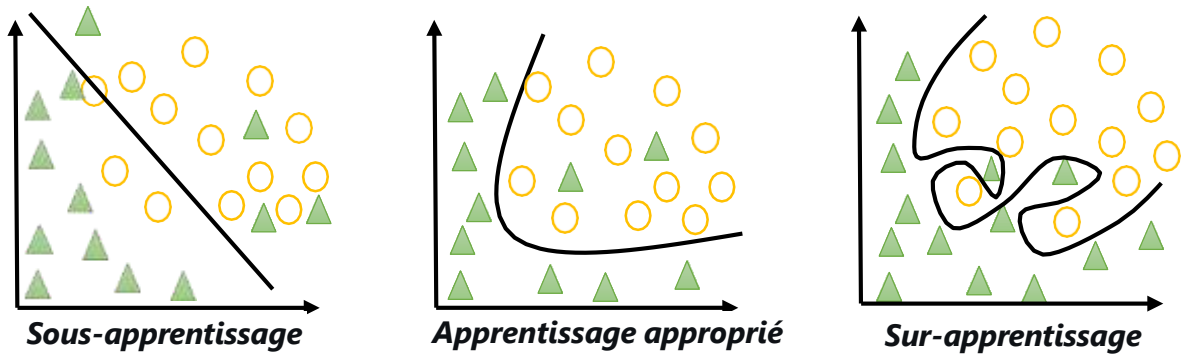


Figure.2.10 Sous-apprentissage, apprentissage approprié et sur-apprentissage **[58]**.

A. La matrice de confusion

La matrice de confusion (Confusion Matrix) est un tableau permettant de comparer les prédictions du modèle avec les valeurs réelles. Elle repose sur quatre situations principales :

Valeurs prédites \ Valeurs réelles	Positif (1)	Négatif (*0*)
Positif (1)	VP	FP
Négatif (*0*)	FN	VN

Figure.2.11 Matrice de confusion (valeurs réelles vs prédites) **[59]**.

- **Vrais Positifs (VP)** : fraude correctement détectée ;
- **Vrais Négatifs (VN)** : transaction normale correctement classifiée ;
- **Faux Positifs (FP)** : transaction normale considérée à tort comme frauduleuse ;

- **Faux Négatifs (FN)** : fraude non détectée par le modèle.

Cette matrice constitue la base de calcul des principales métriques d'évaluation.

Dans les systèmes de détection de fraude bancaire, les erreurs de type Faux Négatifs sont particulièrement critiques, car elles correspondent à des fraudes réelles non détectées par le modèle,

ce qui peut entraîner des pertes financières importantes pour les institutions bancaires.

B. Accuracy

L'Accuracy mesure le pourcentage global de prédictions correctes réalisées par le modèle.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad \textbf{(3)}$$

Bien que cette métrique soit facile à interpréter, elle peut devenir trompeuse dans les datasets déséquilibrés.

En effet, un modèle peut obtenir une Accuracy élevée simplement en classant la majorité des transactions comme normales, sans détecter efficacement les fraudes.

C. Precision

La Precision mesure la proportion des transactions détectées comme frauduleuses qui sont réellement des fraudes.

Dans les systèmes bancaires, une faible précision entraîne une augmentation des faux positifs, ce qui peut provoquer des blocages inutiles de transactions légitimes et réduire la satisfaction des clients.

La Precision est calculée par la formule suivante :

$$\text{Precision} = \frac{VP}{VP+FP} \quad \textbf{(4)}$$

D. Recall

- Le Recall, également appelé taux de vrais positifs (Taux de vrais positifs (TVP)), mesure la capacité du modèle à détecter les véritables fraudes.

$$\text{Recall} = \frac{VP}{VP+FN}$$

Dans le domaine bancaire, cette métrique est particulièrement importante, car une fraude non détectée peut entraîner des pertes financières importantes

F1-Score

Le F1-Score combine la Precision et le Recall afin d'obtenir une mesure équilibrée des performances du modèle.

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \textbf{(6)}$$

Cette métrique est très utile lorsque les données sont déséquilibrées et qu'il est nécessaire de maintenir un équilibre entre détection des fraudes et réduction des fausses alertes.

3 La courbe ROC et l'AUC

La courbe ROC (Receiver Operating Characteristic) permet d'évaluer la capacité d'un modèle à distinguer les classes positives et négatives. Elle représente la relation entre le True Positive Rate (Recall) le False Positive Rate (FPR).

La courbe ROC aide à comprendre le compromis entre la sensibilité et la spécificité selon différents seuils de décision. L'aire sous la courbe, appelée AUC(Area Under the Curve), fournit une mesure globale des performances du modèle.

Une valeur proche de 1 indique un excellent modèle, tandis qu'une valeur proche de 0,5 indique un modèle peu performant. Plus la valeur de l'AUC est élevée, plus le modèle possède une bonne capacité de séparation entre les transactions normales et les transactions frauduleuse.

Dans les datasets déséquilibrés utilisés pour la détection de fraude bancaire, le Recall, le F1-Score et l'AUC sont généralement considérés comme plus pertinents que l'Accuracy seule **[60]**.

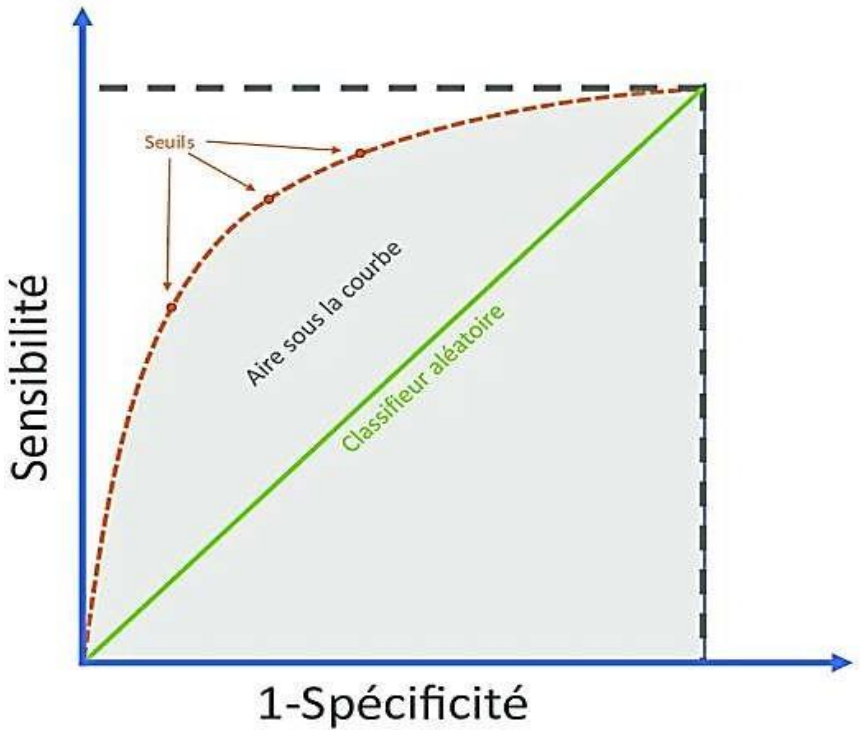


Figure.2.12 Schéma de courbe ROC en orange et AUC en gris. **[61]**.

3.1.1. Importance des variables (Feature Importance)

L'un des principaux avantages de l'algorithme AdaBoost est sa capacité à évaluer l'importance des variables utilisées dans le processus de classification.

Cette importance est généralement déterminée par la contribution de chaque variable à la réduction de l'erreur de classification au sein des modèles utilisés, notamment les arbres de décision **[62]**.

Ainsi, plus une variable contribue à améliorer la séparation entre les classes, plus son poids et son importance dans le modèle deviennent élevée.

Dans le domaine de la détection de fraude bancaire, plusieurs variables jouent un rôle essentiel dans l'identification des comportements suspects, notamment :

• le montant de la transaction ;

• la localisation géographique ;

• l'heure de la transaction ;

• le type de carte bancaire.

Ces caractéristiques permettent de mieux représenter le comportement des transactions et d'améliorer la performance globale du système de détection des fraudes.

Tableau 2.3 : Exemple d'importance des variables

Variable	**Importance**
Montant de transaction	**0.35**
Heure de transaction	**0.21**
Localisation	**0.18**
Canal de paiement	**0.14**
Ancienneté du compte	**0.12**

3.2. Conclusion

Ce chapitre a permis de présenter l'évolution des méthodes de détection de fraude bancaire, depuis les approches traditionnelles jusqu'aux techniques modernes basées sur l'intelligence artificielle.

Les systèmes classiques de contrôle restent indispensables dans les établissements financiers, mais leurs limites apparaissent clairement face à l'augmentation du volume des transactions et à la sophistication des attaques frauduleuses.

Les technologies de Machine Learning offrent aujourd'hui des solutions plus rapides et plus performantes pour détecter les comportements suspects. Parmi ces technologies, l'algorithme AdaBoost présente plusieurs avantages importants :

- bonne capacité de classification ;
- gestion efficace des données déséquilibrées ;
- amélioration progressive des performances ;
- interprétation des variables importantes.

Le chapitre suivant sera consacré à l'implémentation pratique du modèle AdaBoost, à la préparation des données, à l'entraînement du modèle ainsi qu'à l'évaluation expérimentale des performances obtenues.

Chapitre 3 :

Implémentation et Évaluation d'un modèle de détection de fraude par AdaBoost

3.1. Introduction

L'objectif principal de ce chapitre Développer et évaluer un modèle de Machine Learning basé sur AdaBoost pour détecter les transactions frauduleuses.

L'implémentation du système a été réalisée en langage Python dans l'environnement Visual Studio Code. Plusieurs bibliothèques spécialisées ont été utilisées, notamment Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn ainsi que Imbalanced-learn.

Le développement du modèle s'est déroulé selon plusieurs étapes essentielles :

- Le prétraitement des données,
- La gestion du déséquilibre des classes,
- L'entraînement du modèle AdaBoost,
- L'optimisation des hyperparamètres,
- Ainsi que l'évaluation détaillée des performances obtenues.

Enfin, une analyse approfondie des résultats expérimentaux a été réalisée afin de mesurer l'efficacité du système proposé dans un contexte réel de détection de fraude bancaire.

3.2. Présentation des données et prétraitement

3.2.1. Présentation du jeu de données

Dans cette expérimentation, nous avons utilisé le fichier **creditcard.csv**, un dataset largement utilisé dans les recherches scientifiques consacrées à la détection de fraude bancaire.

Ce jeu de données contient des transactions effectuées par cartes bancaires européennes et comprend à la fois des opérations normales et des opérations frauduleuses.

Le dataset contient plusieurs variables numériques anonymisées nommées de V1 à V28. Ces variables ont été générées à l'aide de la méthode PCA (*Principal Component Analysis*) afin de protéger la confidentialité des informations bancaires des utilisateurs

En plus des variables anonymisées, le dataset contient également les attributs suivants :

- **Time** : représente le temps écoulé entre les transactions.
- **Amount** : représente le montant financier de la transaction.
- **Class** : variable cible indiquant si la transaction est frauduleuse ou non.

La variable « Class » possède deux valeurs :

- 0 : transaction non frauduleuse.
- 1 : transaction frauduleuse.

L'analyse initiale des données montre un fort déséquilibre entre les transactions normales et les transactions frauduleuses.

❖ **Distribution des classes**

Les statistiques obtenues sont les suivantes :

Tableau 3.1 Résultats de la séparation des données d'apprentissage et de test

Classe	**Nombre de transactions**
0 (Non frauduleuse)	56863
1 (Frauduleuse)	98

Ces résultats montrent que les transactions frauduleuses représentent une proportion extrêmement faible du dataset total.

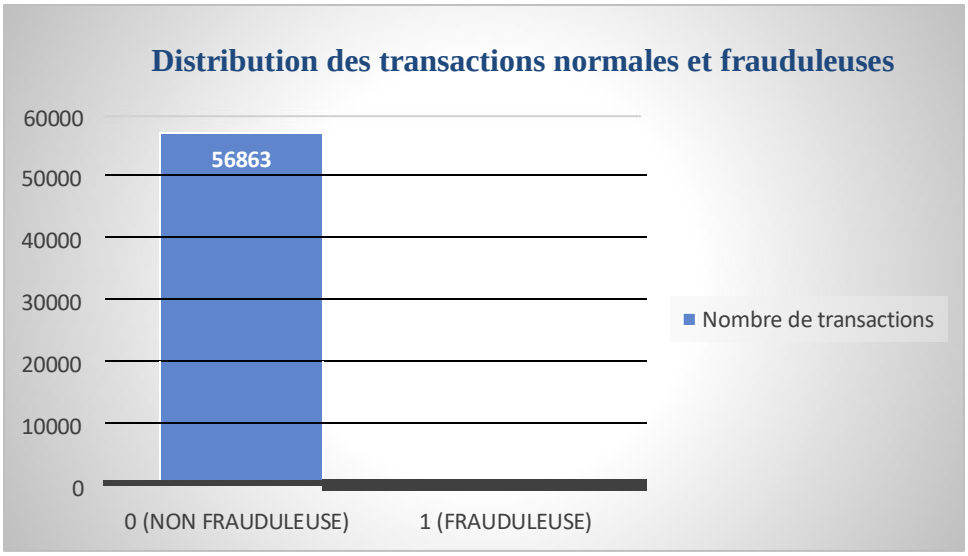


Figure.3.1 Distribution des classes dans le dataset

❖ **Commentaire**

La **Figure.3.1** met clairement en évidence le problème du déséquilibre des classes dans le dataset utilisé. On observe une domination très importante des transactions non frauduleuses par rapport aux transactions frauduleuses. Cette situation constitue un défi majeur dans les systèmes de détection de fraude bancaire, car le modèle risque de privilégier la classe majoritaire au détriment des fraudes réelles. Afin de résoudre ce problème, des techniques de rééquilibrage ont été appliquées avant l'entraînement du modèle AdaBoost.

3.2.2. Prétraitement des données

Avant l'entraînement du modèle, plusieurs étapes de prétraitement ont été réalisées afin d'améliorer la qualité des données et les performances du système de classification.

Dans un premier temps, les données ont été importées à l'aide de la bibliothèque Pandas :

Data = pd.read_csv('creditcard.csv')

Après cela, les variables explicatives et la variable cible ont été séparées :

X = data.drop(['Class', 'Time'], axis=1)

y = data['Class']

La variable « Time » a été supprimée car elle n'apporte pas d'informations pertinentes pour la prédiction dans cette étude.

Par la suite, une étape de normalisation a été appliquée grâce à la méthode StandardScaler :

```
scaler=StandardScaler()
X_train_scaled=scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

Cette étape permet de standardiser les caractéristiques numériques afin d'éviter qu'une variable possédant des valeurs élevées influence excessivement le modèle.

Enfin, les données ont été divisées en deux ensembles :

- 70 % pour l'entraînement,
- 30 % pour les tests.

```
X_train, X_test, y_train, y_test = train_test_split( X_scaled,
    y,
    test_size=0.3,
    random_state=42,
    stratify=y
)
```

L'utilisation du paramètre `stratify=y` permet de conserver la même proportion des classes dans les ensembles d'entraînement et de test.

- Le rééquilibrage des données a permis d'améliorer la représentation des transactions frauduleuses dans l'ensemble d'entraînement. Cette étape est essentielle afin d'éviter que le modèle AdaBoost soit influencé uniquement par la classe majoritaire

3.3. Gestion du déséquilibre des classes

Le déséquilibre des classes représente l'un des principaux défis dans les systèmes de détection de fraude bancaire. En effet, les transactions frauduleuses sont très rares comparées aux transactions normales.

Dans ce type de problème, un modèle peut obtenir une Accuracy très élevée tout en étant incapable de détecter correctement les fraudes. Afin de résoudre cette difficulté, deux techniques complémentaires ont été utilisées :

- SMOTE (*Synthetic Minority Over-sampling Technique*),
- Random Under Sampling.

La méthode SMOTE permet de générer artificiellement de nouvelles instances de la classe minoritaire :

```
smote = SMOTE(sampling_strategy=0.3, random_state=42)
```

Ensuite, une méthode de sous-échantillonnage a été appliquée afin de réduire le nombre de transactions normales :

```
undersample = RandomUnderSampler(
    sampling_strategy=0.5,
    random_state=42
)
```

Les deux techniques ont été intégrées dans un pipeline unique :

```
resampling_pipeline = Pipeline([
    ('smote', smote),
    ('undersample', undersample)
])
```

Après l'application du rééchantillonnage, la distribution des classes devient :

Tableau 3.2 : Distribution des classes après le rééquilibrage du dataset

Classe	Nombre
Non fraude	23880
Fraude	11940

Cette étape a considérablement amélioré la capacité du modèle à apprendre les caractéristiques des transactions frauduleuses.

- L'optimisation des hyperparamètres ainsi que du seuil de classification a permis d'améliorer les performances globales du modèle AdaBoost. Cette étape a contribué à obtenir un meilleur équilibre entre la précision et la capacité de détection des fraudes.

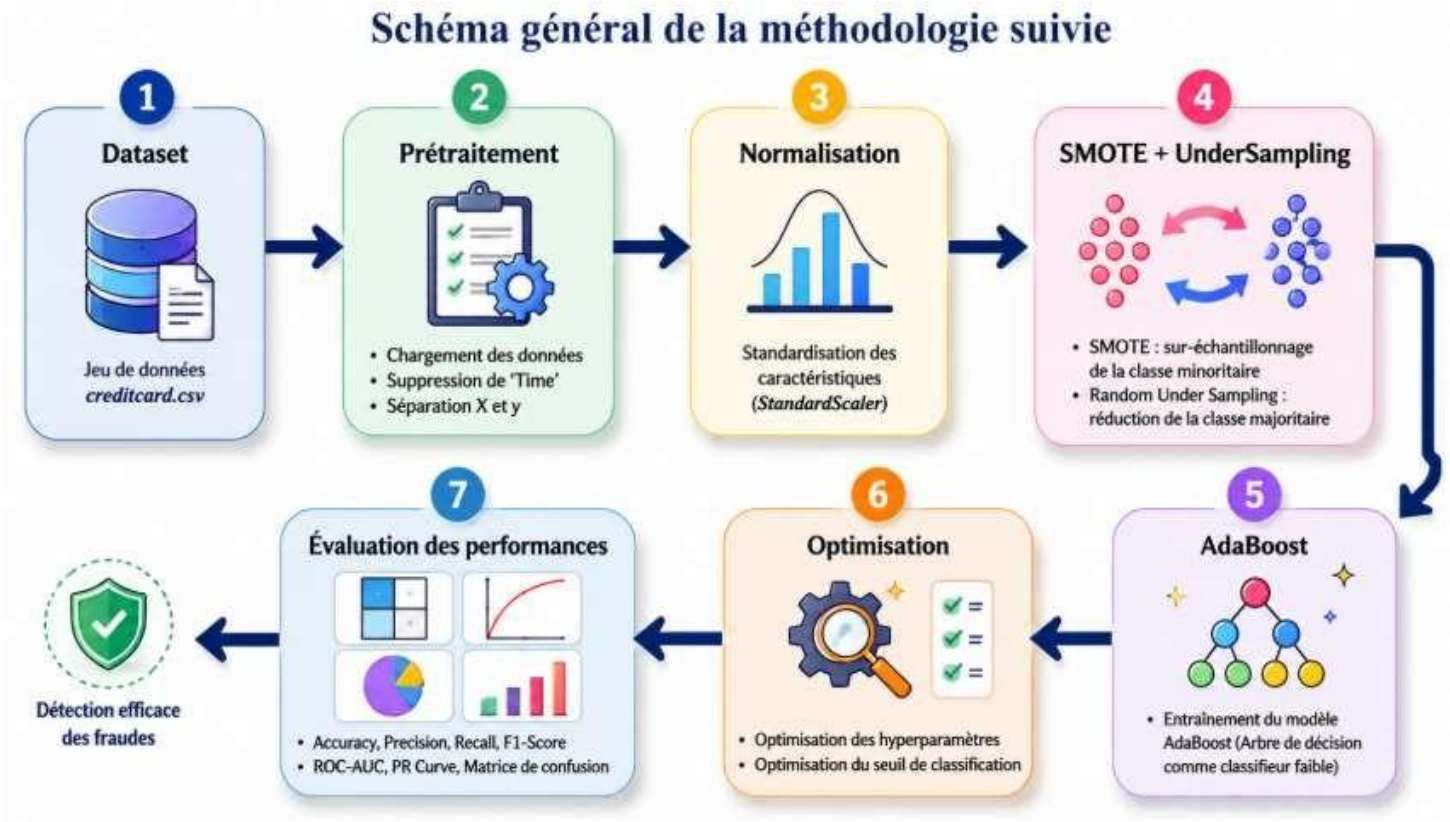


Figure.3.2 Architecture générale du système de détection de fraude bancaire basé sur AdaBoost

✓ La méthodologie suivie dans cette expérimentation est résumée dans la figure suivante

Cette figure présente les principales étapes suivies dans le développement du système intelligent de détection de fraude bancaire. Le processus commence par le prétraitement des données, suivi du rééquilibrage des classes, puis l'entraînement et l'évaluation du modèle AdaBoost.

3.4. Entraînement et optimisation du modèle AdaBoost

3.4.1. Construction du modèle

Après le prétraitement et le rééquilibrage des données, nous avons procédé à l'entraînement du modèle AdaBoost.

Dans cette expérimentation, le modèle AdaBoost a été implémenté à l'aide de la bibliothèque **Scikit-learn** afin de détecter les transactions frauduleuses.

Le modèle a été initialisé avec le paramètre **random_state = 42**, garantissant la reproductibilité des résultats obtenus. Le classificateur de base (*base estimator*) utilisé est un arbre de décision (**DecisionTreeClassifier**) confuguré avec les hyperparamètres suivants :

- **max_depth = 3** : limite la profondeur de l'arbre afin de réduire le risque de surapprentissage (Overfitting).
- **min_samples_split = 15** : impose un minimum de 15 observations pour diviser un nœud.
- **min_samples_leaf = 5** : exige qu'une feuille terminale contienne au moins cinq observations, ce qui améliore la robustesse du modèle.

Le modèle AdaBoost a ensuite été entraîné sur l'ensemble des données d'apprentissage, puis évalué sur les données de test afin de mesurer ses performances ses performances à l'aide des métriques Accuracy,Precision, **Recall** et **F1 - score** et **AUC**

3.4.2. Optimisation des hyperparamètres

Afin d'améliorer les performances du modèle, une phase d'optimisation des hyperparamètres a été réalisée à l'aide de la méthode GridSearchCV avec une validation croisée de type 5-fold.

```
grid_search = GridSearchCV(
    ada,
    param_grid,
    cv=5,
    scoring='f1',
    n_jobs=-1
)
```

Les meilleurs hyperparamètres obtenus sont présentés dans le tableau suivant :

Tableau 3.3 : Optimisation des hyperparamètres par GridSearchCV

Paramètre	Valeur optimale
n_estimators	250
learning_rate	0.2
max_depth	3
min_samples_split	15
min_samples_leaf	5

L'optimisation des hyperparamètres a permis d'améliorer les performances du modèle tout en limitant le risque de surapprentissage.

3.4.3. Optimisation du seuil de classification

Dans les problèmes de détection de fraude, l'utilisation du seuil classique 0.5 n'est pas toujours optimale. Pour cette raison, une phase supplémentaire d'optimisation du seuil de classification a été réalisée.

Le seuil optimal obtenu est :

Tableau 3.4 : Seuil de classification optimal

Paramètre	Valeur
Seuil optimal	0.5745

Les performances associées à ce seuil sont les suivantes :

Tableau 3.5 : Performances obtenues avec le seuil optimal

Métrique	Valeur
Precision	86.21 %
Recall	86.21 %
F1-Score	86.21 %

Cette optimisation a permis d'obtenir un meilleur équilibre entre la précision et la capacité du système à détecter les fraudes réelles.

3.5. Résultats et évaluation des performances

3.5.1. Métriques d'évaluation

Afin d'évaluer les performances du modèle AdaBoost, plusieurs métriques ont été utilisées.

- ✓ Accuracy
- ✓ Precision
- ✓ Recall
- ✓ F1-Score

3.5.2. Résultats expérimentaux

Après l'entraînement du modèle AdaBoost, les résultats suivants ont été obtenus :

Tableau 3.6 : Résultats expérimentaux du modèle AdaBoost

Métrique	Résultat
Accuracy	99.97 %
Precision	96.15 %
Recall	86.21 %
F1-Score	90.91 %
ROC-AUC	99.72 %

L'ensemble de test est composé de 17 089 transactions, dont 17 060 transactions non frauduleuses et 29 transactions frauduleuses.

Le rapport de classification montre les résultats suivants :

- **17 059 vrais négatifs (VN)** : transactions normales correctement classifiées ;
- **25 vrais positifs (VP)** : transactions frauduleuses correctement détectées ;
- **1 faux positif (FP)** : transaction normale classifiée à tort comme frauduleuse ;
- **4 faux négatifs (FN)** : transactions frauduleuses non détectées.

La vérification donne : 17059+25+1+4 =17089 transactions, ce qui confirme la cohérence des résultats.

Ces performances montrent que le modèle AdaBoost possède une excellente capacité de classification avec une **accuracy de 99,97 %**, une **precision de 96,15 %** et un **recall de 86,21 %**. Le modèle détecte ainsi la majorité des transactions frauduleuses tout en maintenant un très faible taux de fausses alertes.

3.5.3. Analyse des figures obtenues

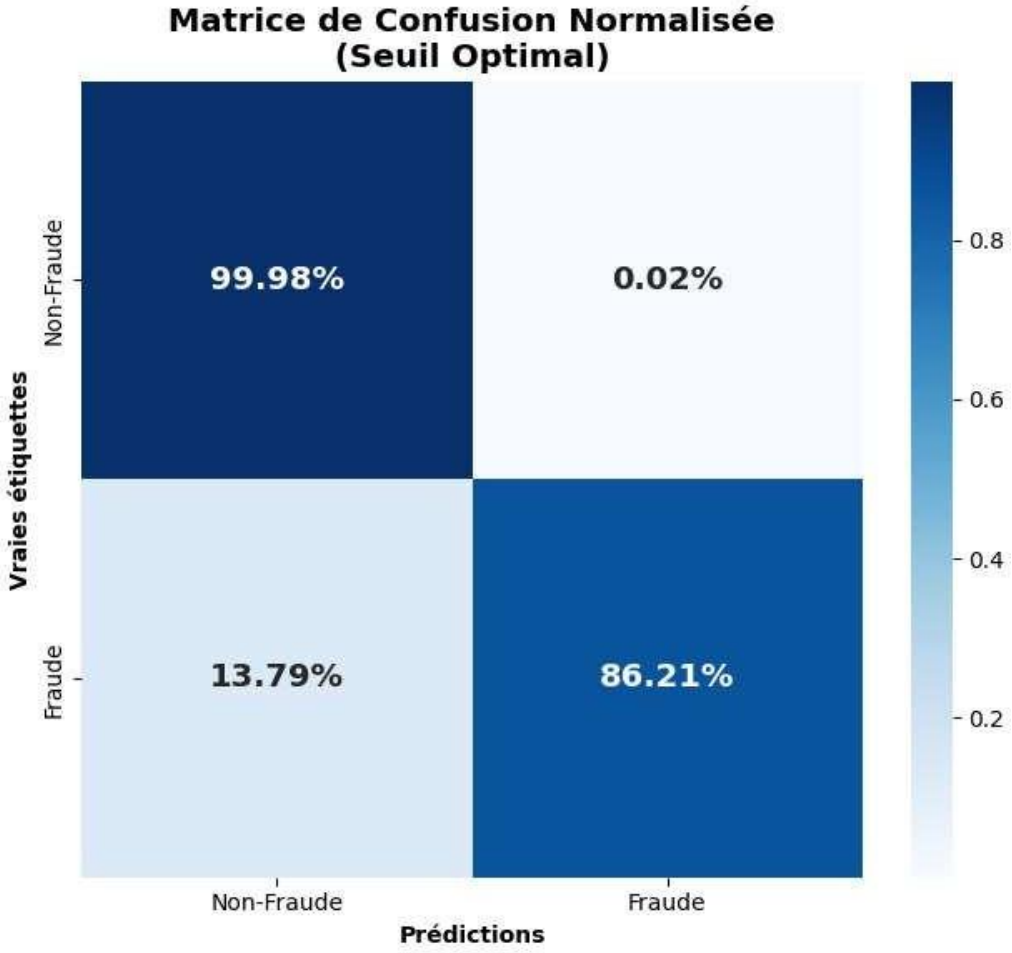


Figure.3.3 Matrice de confusion normalisée du modèle AdaBoost

❖ **Commentaire**

La matrice de confusion montre que le modèle AdaBoost classe correctement la majorité des transactions normales et frauduleuses. Le faible nombre de faux positifs et de faux négatifs confirme l'efficacité du système proposé.

Ces résultats confirment l'efficacité du modèle dans la réduction des erreurs de classification.

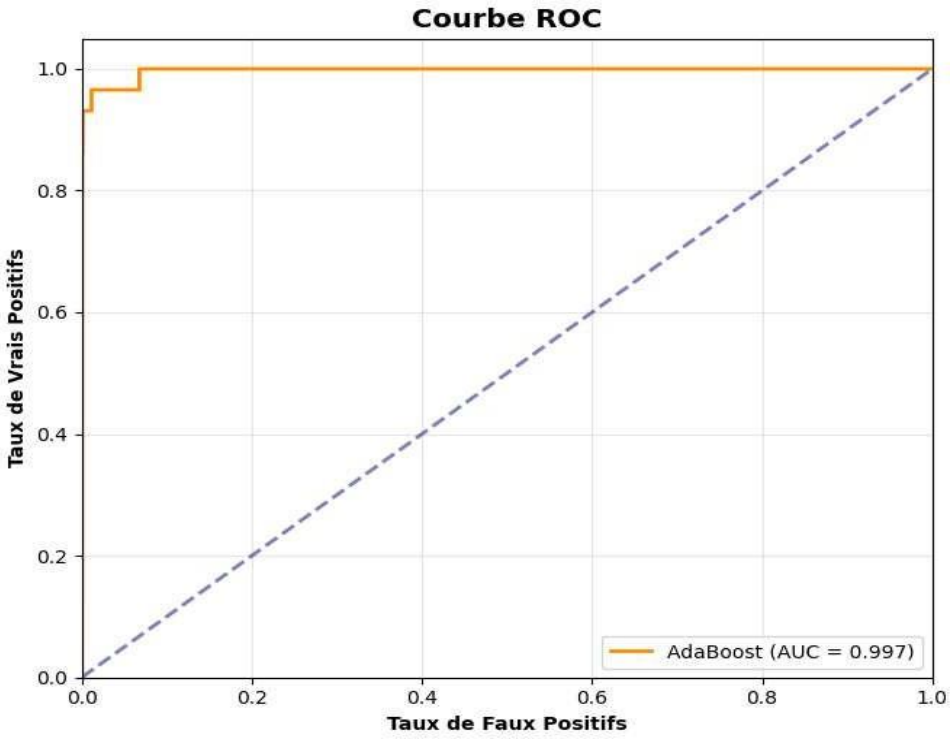


Figure.3.4 Courbe ROC du modèle AdaBoost

❖ **Commentaire**

La courbe ROC présente une valeur AUC proche de 1 (AUC = 0.997), ce qui indique une excellente capacité de discrimination entre les transactions frauduleuses et non frauduleuses.

On peut conclure que le modèle possède une excellente capacité de discrimination entre les deux classes.

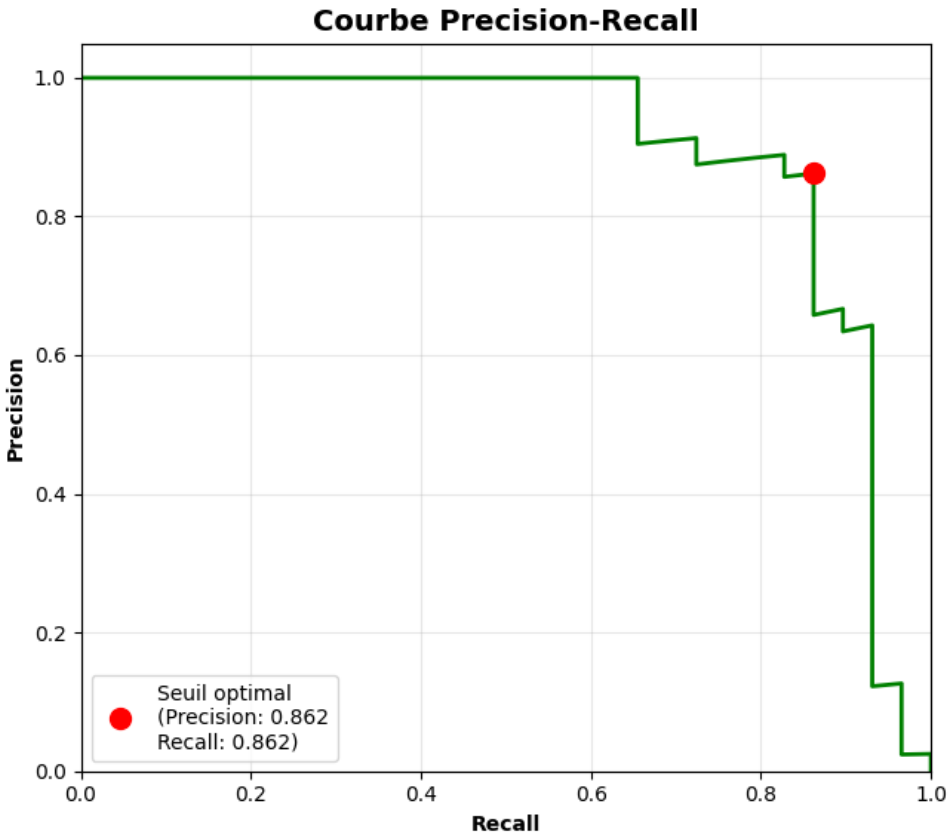


Figure.3.5 Courbe Precision-Recall du modèle AdaBoost

❖ **Commentaire**

La courbe Precision-Recall montre un bon équilibre entre la précision et le rappel. Le point rouge représente le seuil optimal permettant d'obtenir les meilleures performances du modèle.

Le modèle maintient un bon équilibre entre la précision et le rappel malgré le déséquilibre des données.

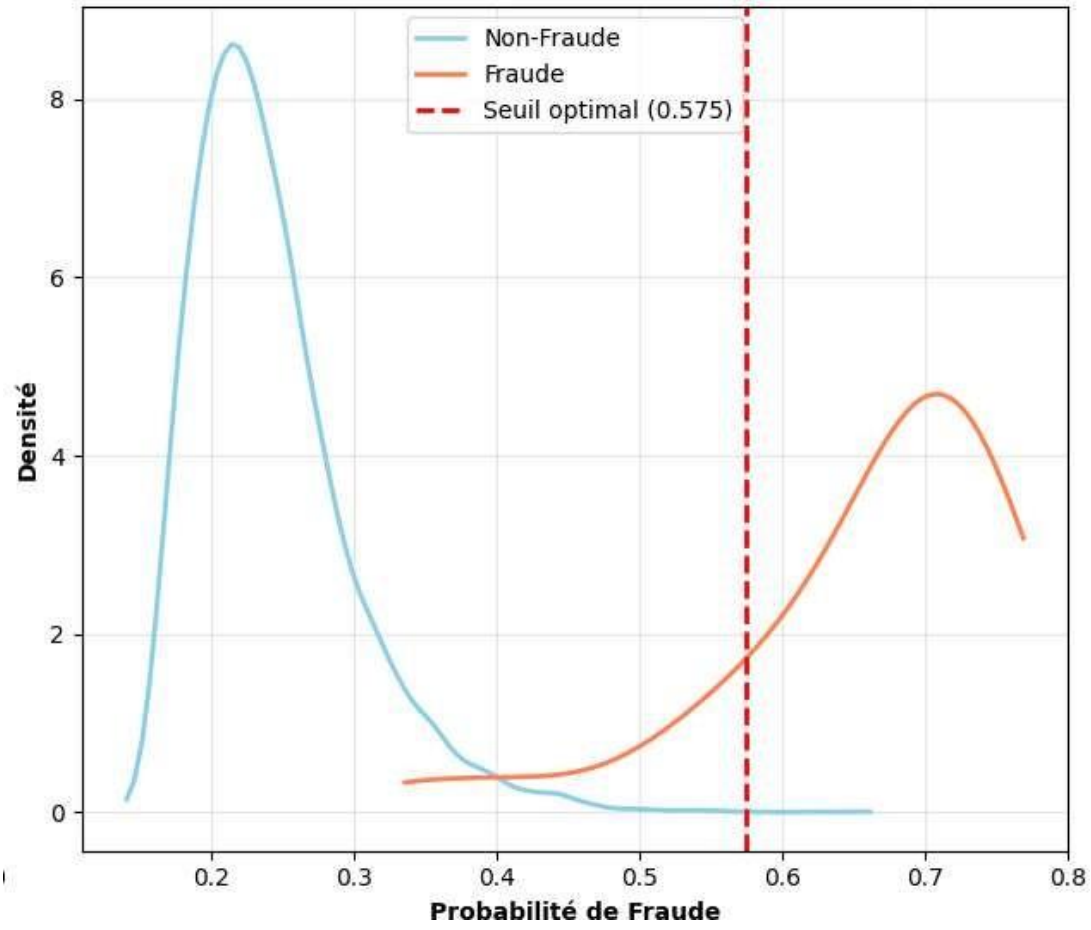


Figure.3.6 Distribution des probabilités par classe

❖ **Commentaire**

Cette figure présente la distribution des probabilités prédites pour les transactions normales et frauduleuses. On observe une séparation claire entre les deux classes ainsi qu'un seuil optimal fixé à 0.575.

La séparation visible entre les deux distributions montre que le modèle distingue efficacement les transactions frauduleuses.

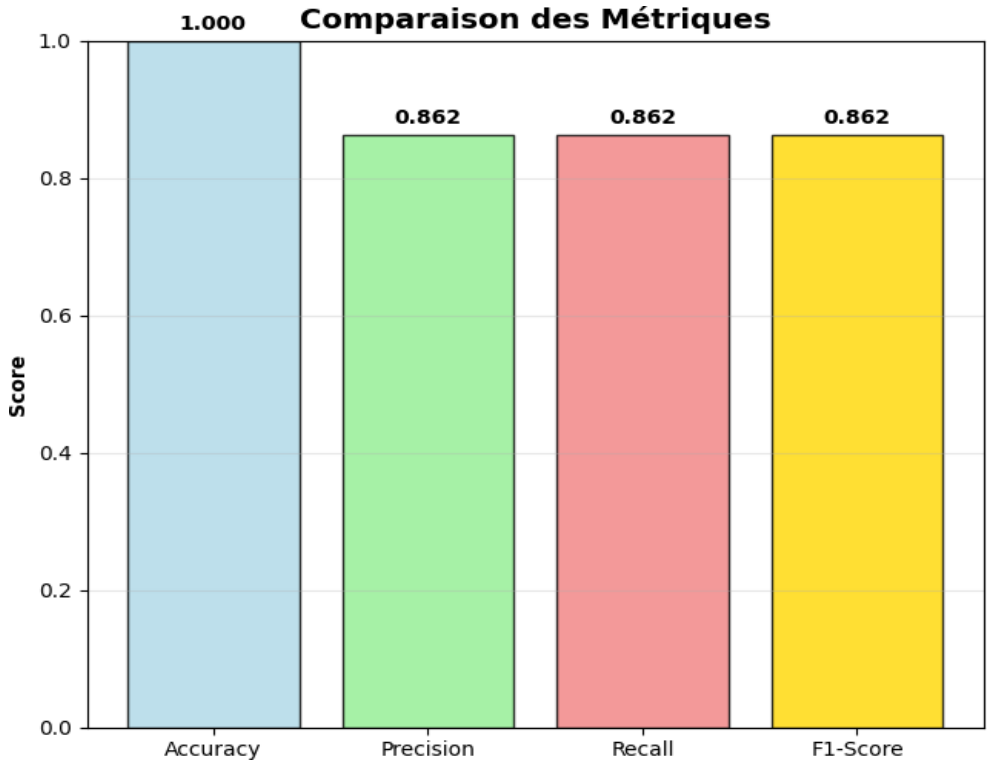


Figure.3.7 Comparaison des métriques d'évaluation

❖ **Commentaire**

Cette figure compare les principales métriques utilisées pour évaluer le modèle AdaBoost. Les résultats montrent une Accuracy très élevée ainsi qu'un bon équilibre entre Precision, Recall et F1-Score.

Les métriques obtenues confirment les bonnes performances globales du système proposé.

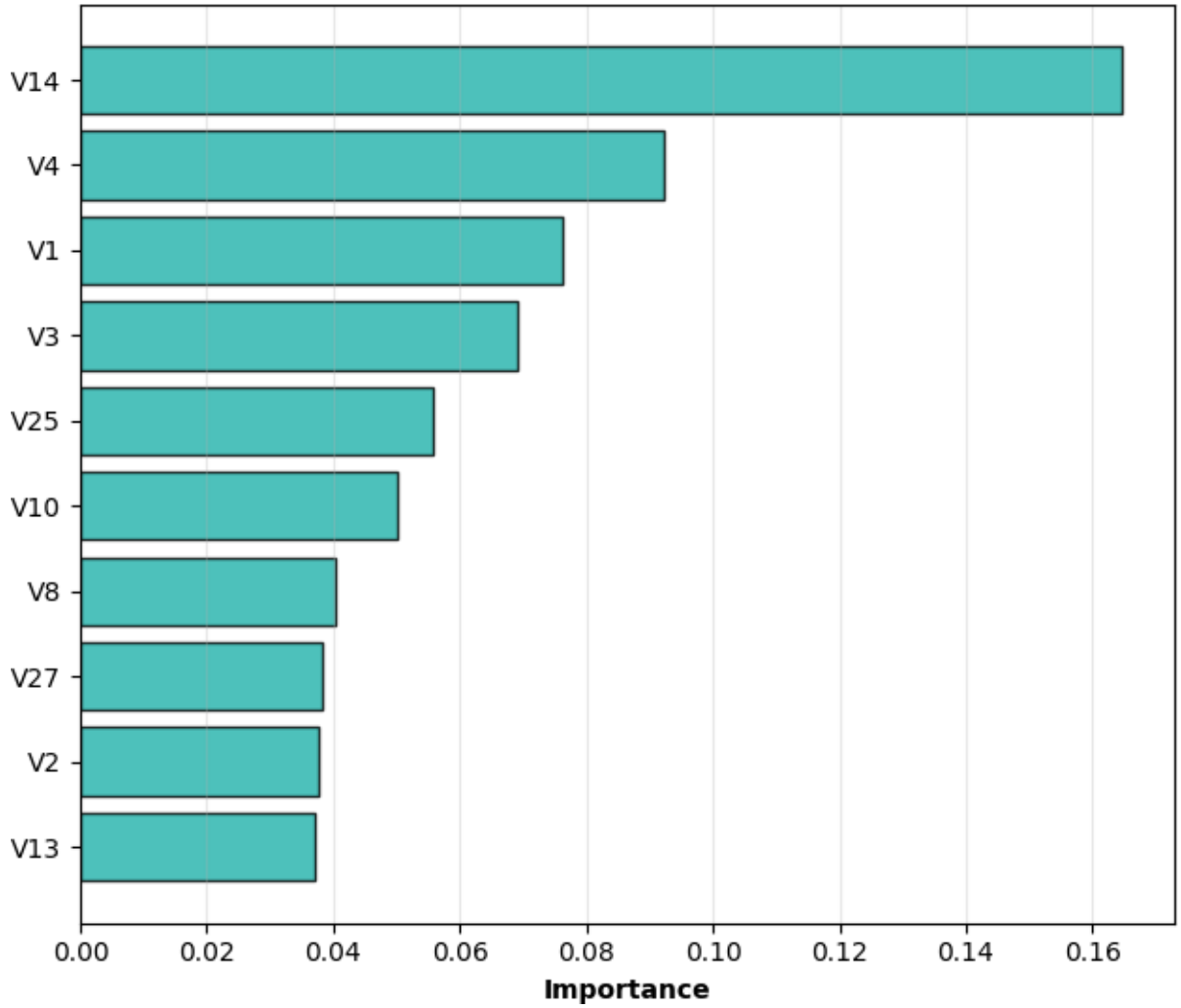


Figure.3.8 Top 10 des caractéristiques les plus importantes

❖ **Commentaire**

Cette figure présente les variables ayant le plus d'influence dans le processus de classification. Les caractéristiques les plus importantes contribuent fortement à la distinction entre les transactions normales et frauduleuses.

Certaines variables jouent un rôle majeur dans la détection des comportements frauduleux.

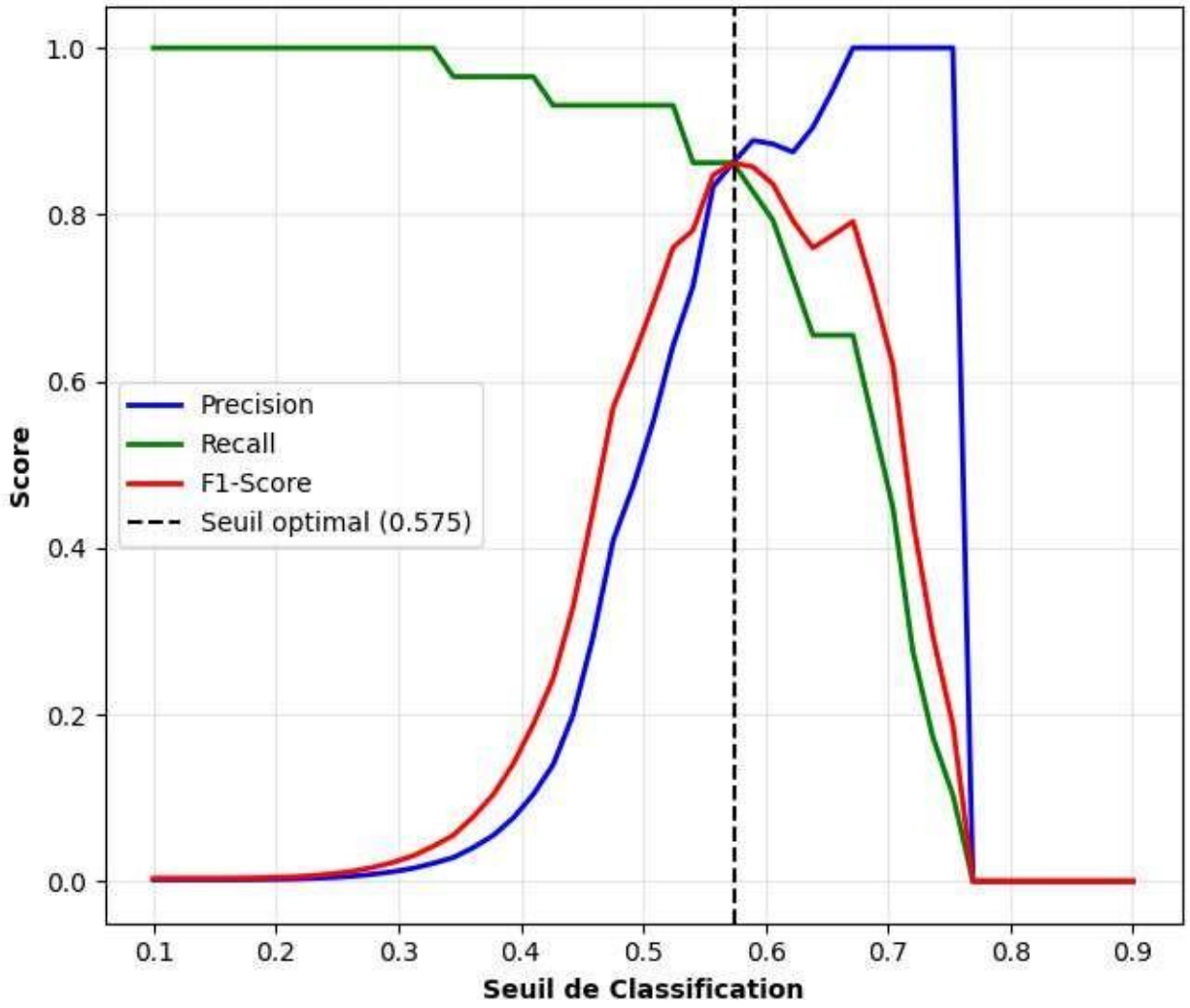


Figure.3.9 Évolution des métriques en fonction du seuil de classification

❖ **Commentaire**

Cette figure illustre l'impact du seuil de classification sur les performances du modèle. Le seuil optimal permet d'obtenir un meilleur compromis entre Precision, Recall et F1-Score. Le seuil optimal améliore le compromis entre Precision et Recall.

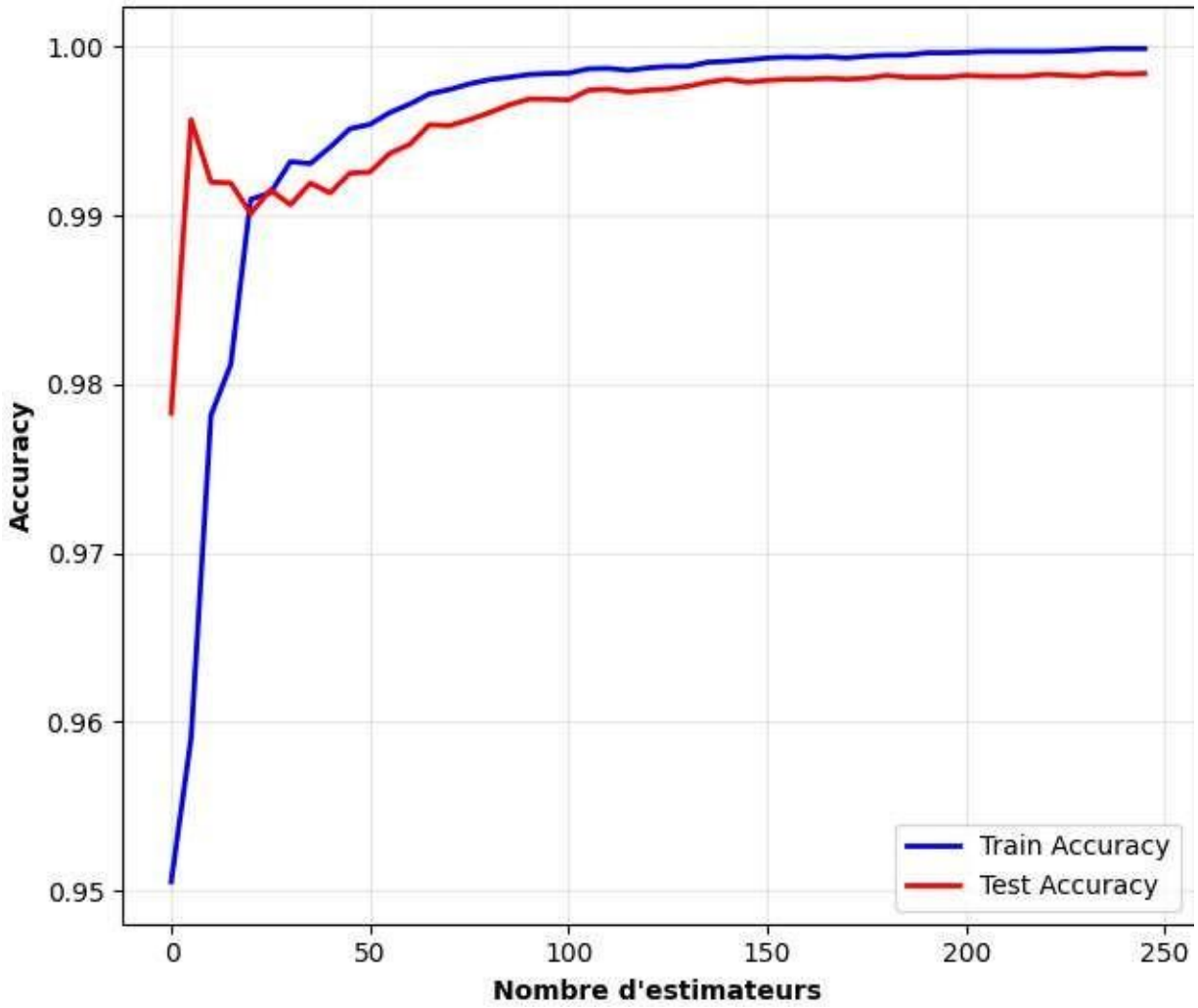


Figure.3.10 Évolution de l'Accuracy pendant l'entraînement et le test

❖ **Commentaire**

Cette figure montre l'évolution de l'Accuracy sur les ensembles d'entraînement et de test pendant l'apprentissage. Les courbes restent proches, ce qui indique une bonne capacité de généralisation du modèle.

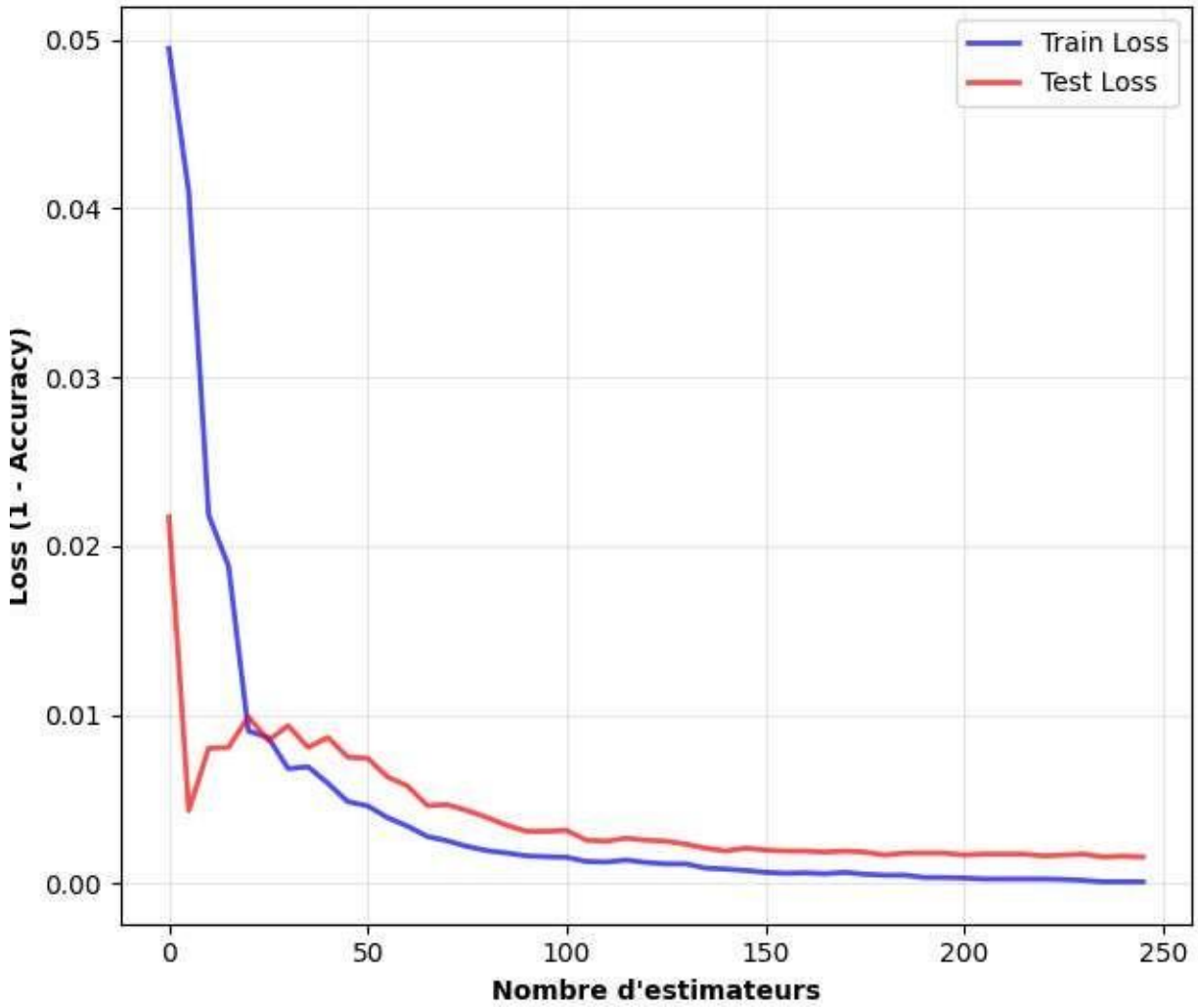


Figure.3.11 Évolution de la Loss pendant l'entraînement et le test

❖ **Commentaire**

Cette figure montre la diminution progressive de la Loss au cours de l'entraînement et le test, ce qui traduit une amélioration continue des performances du modèle AdaBoost.

La diminution progressive de la Loss traduit par un bon apprentissage du modèle. Du coup, l'inexistence du sur-apprentissage.

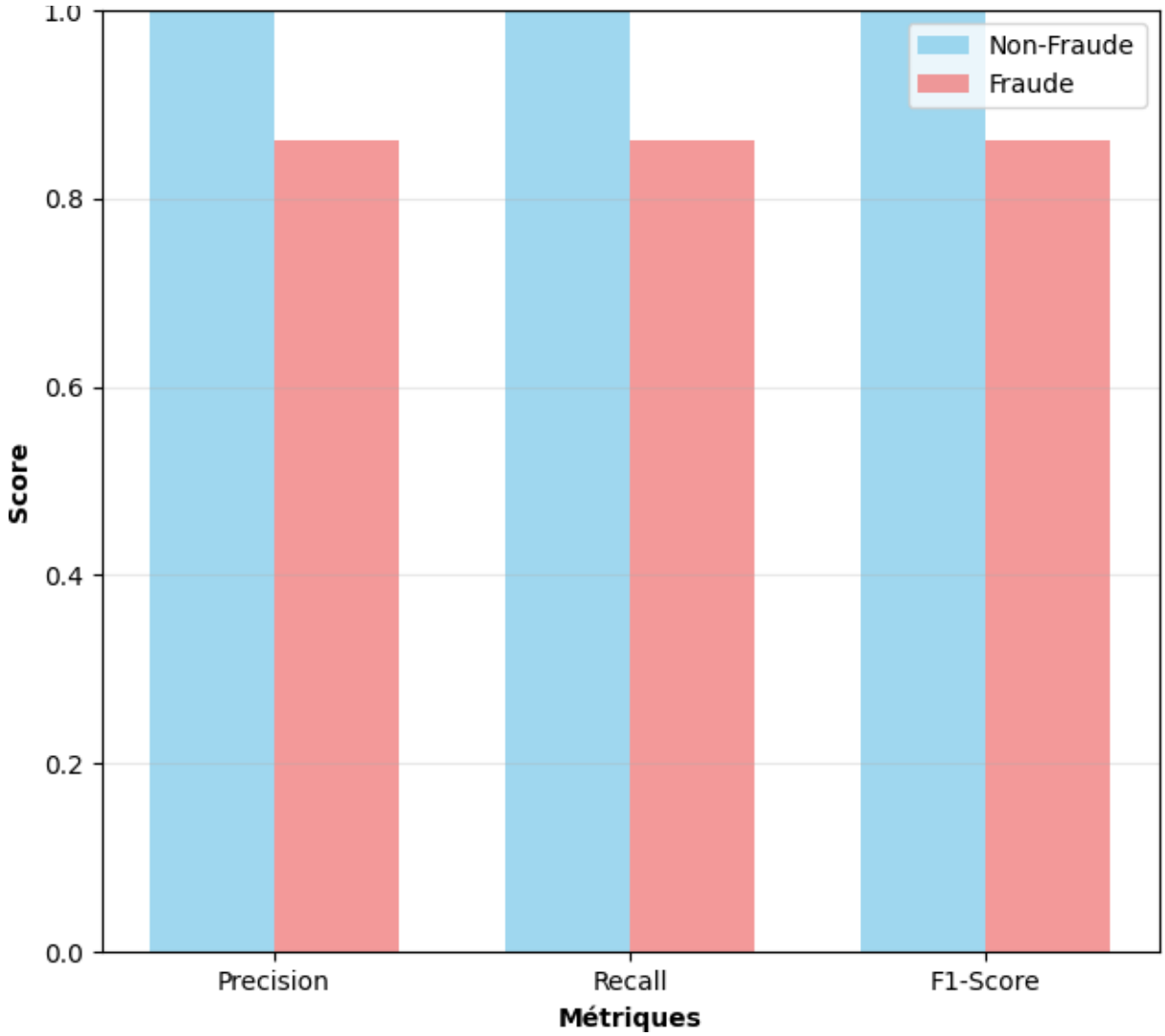


Figure.3.12 Performance du modèle par classe

❖ **Commentaire**

Cette figure compare les performances obtenues pour les classes “Fraude” et “Non-Fraude”. Les résultats montrent que le modèle conserve une excellente performance pour les transactions normales tout en détectant efficacement les fraudes.

Le modèle conserve de bonnes performances pour les transactions normales et frauduleuses.

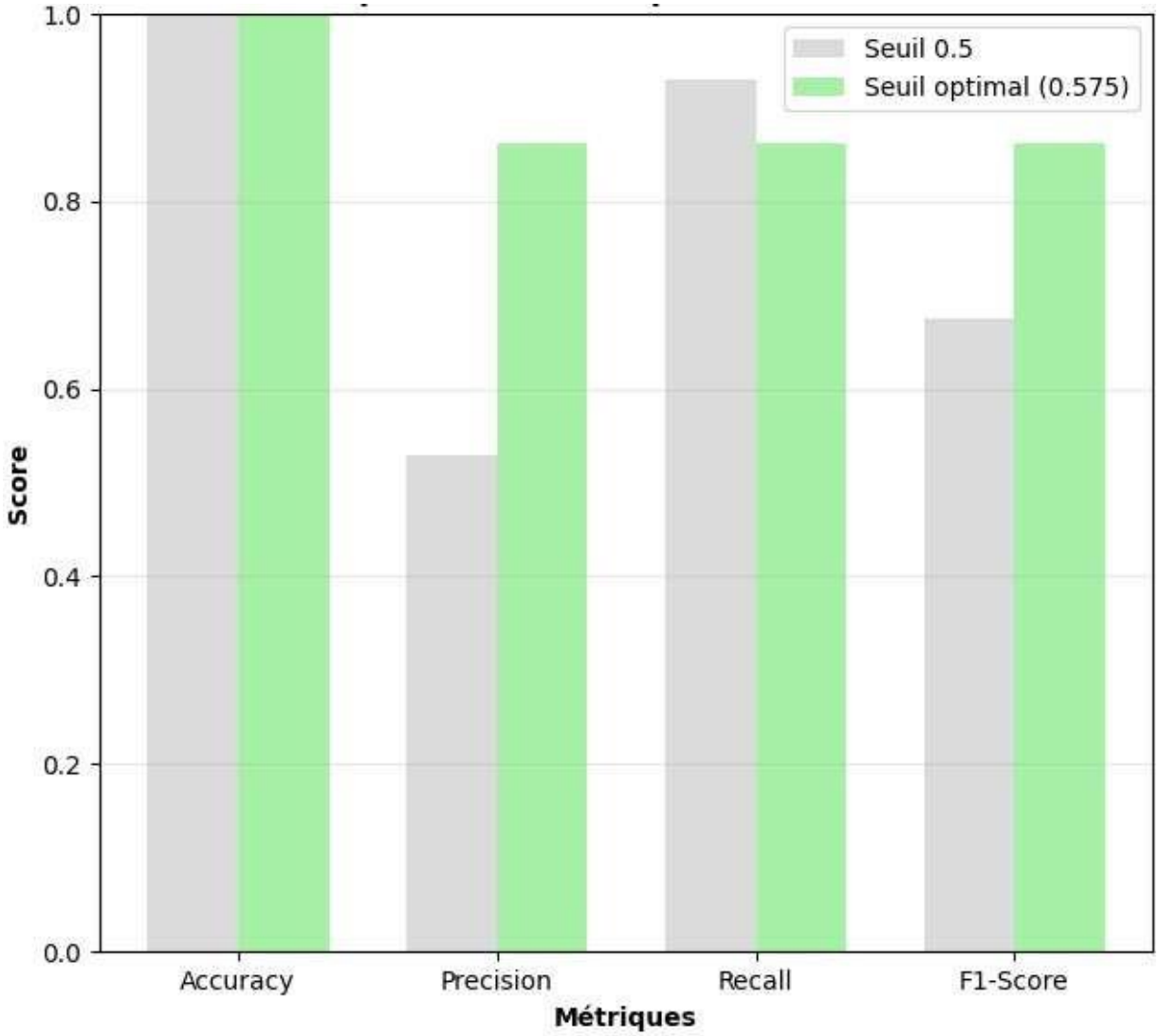


Figure.3.13 Comparaison entre le seuil optimal et le seuil par défaut 0.5

❖ **Commentaire**

Cette figure montre que l'utilisation du seuil optimal améliore la précision et le F1-Score par rapport au seuil classique 0.5, tout en maintenant un bon niveau de Recall.

3.6. Conclusion

Dans ce chapitre, nous avons présenté l'implémentation complète d'un système de détection de fraude bancaire basé sur l'algorithme AdaBoost. L'étude a débuté par la préparation et le prétraitement du jeu de données, suivis par l'application de techniques de rééquilibrage afin de remédier au problème du déséquilibre des classes.

Par la suite, le modèle AdaBoost a été entraîné et optimisé à travers la recherche des meilleurs hyperparamètres ainsi que l'ajustement du seuil de classification. Les résultats obtenus ont montré des performances très satisfaisantes en termes d'Accuracy, de Precision, de Recall, de F1-Score et de ROC-AUC, démontrant la capacité du modèle à identifier efficacement les transactions frauduleuses tout en limitant les erreurs de classification.

L'analyse des différentes représentations graphiques a également permis de confirmer la robustesse du modèle et son aptitude à distinguer les comportements normaux des comportements frauduleux. Ces résultats mettent en évidence l'intérêt des techniques d'apprentissage automatique, et particulièrement d'AdaBoost, dans le développement de systèmes intelligents de lutte contre la fraude bancaire.

Conclusion générale

Conclusion Générale et Perspectives

Ce mémoire avait pour objectif d'étudier l'apport de l'intelligence artificielle dans la détection de la fraude au sein des transactions bancaires. Face à l'augmentation des paiements électroniques et à la sophistication croissante des techniques frauduleuses, les approches traditionnelles de contrôle montrent aujourd'hui certaines limites en matière de rapidité, d'adaptabilité et d'efficacité. L'étude réalisée a permis de mettre en évidence l'intérêt des techniques de Machine Learning pour renforcer les systèmes de détection de fraude et améliorer la sécurité des opérations bancaires.

Sur le plan pratique, l'implémentation d'un modèle basé sur l'algorithme AdaBoost, associée à des techniques de prétraitement et de rééquilibrage des données, a permis d'obtenir des performances satisfaisantes en matière de classification des transactions frauduleuses. Les résultats obtenus confirment la capacité des méthodes d'intelligence artificielle à identifier efficacement les comportements suspects tout en limitant les erreurs de détection. Cette recherche démontre ainsi que l'exploitation des données et des modèles prédictifs constitue un levier stratégique pour les établissements bancaires confrontés à des risques de fraude de plus en plus complexes.

Cependant, cette étude présente certaines limites. Le modèle a été évalué sur un jeu de données spécifique qui ne reflète pas nécessairement l'ensemble des scénarios rencontrés dans les environnements bancaires réels. De plus, l'évolution rapide des techniques de fraude impose une mise à jour régulière des modèles afin de maintenir leur niveau de performance. L'intégration opérationnelle de ces systèmes nécessite également des ressources techniques importantes, une qualité élevée des données ainsi qu'une infrastructure adaptée au traitement en temps réel.

À la lumière des résultats obtenus, il est recommandé aux établissements bancaires d'investir davantage dans les solutions d'intelligence artificielle, de renforcer leurs mécanismes de surveillance continue et de développer des stratégies combinant expertise humaine et systèmes intelligents. L'utilisation de modèles prédictifs avancés doit s'accompagner d'une gouvernance rigoureuse des données et d'une politique de cybersécurité adaptée aux nouvelles menaces.

Les perspectives de recherche futures concernent l'exploration d'algorithmes plus performants tels que XGBoost, LightGBM, les réseaux de neurones profonds et les approches hybrides combinant

plusieurs techniques d'apprentissage. L'intégration de données comportementales, l'analyse en temps réel et l'utilisation de l'intelligence artificielle explicable (Explainable AI) représentent également des axes prometteurs pour améliorer la compréhension des décisions prises par les modèles et renforcer la confiance des utilisateurs.

Enfin, l'efficacité future des systèmes de détection de fraude reposera sur la convergence entre les innovations technologiques, l'évolution des cadres réglementaires et la capacité des institutions financières à s'adapter continuellement aux nouveaux schémas frauduleux. Dans ce contexte, l'intelligence artificielle apparaît comme un outil incontournable pour construire des systèmes de détection plus intelligents, plus transparents et plus résilients face aux défis de la fraude bancaire moderne.

Références bibliographiques

Liste des références

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, MA, USA: MIT Press.
2. Delamaire, L., Abdou, H. A., & Pointon, J. (2009). Credit card fraud and detection techniques: A review. *Banks and Bank Systems*, **4**(2), 57–68.
3. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review. *Decision Support Systems*, **50**(3), 559–569. https://doi.org/10.1016/j.dss.2010.08.006
4. Interpol. (2023). *Global Financial Fraud Assessment Report 2023*. Lyon, France: International Criminal Police Organization (INTERPOL).
5. Banque d'Algérie. (2023). *Rapport annuel sur la stabilité financière 2023*. Alger, Algérie : Banque d'Algérie.
6. European Central Bank. (2022). *Report on Card Fraud*. Frankfurt am Main, Germany: European Central Bank.
7. Association of Certified Fraud Examiners (ACFE). (2022). *Occupational Fraud 2022: A Report to the Nations*. Austin, Texas, USA: ACFE.
8. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, **60**, 19–31. https://doi.org/10.1016/j.jnca.2015.11.016
9. Phua, C., Lee, V., Smith-Miles, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *Artificial Intelligence Review*, **34**(1), 1–14. https://doi.org/10.1007/s10462-009-9119-y
10. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321–357.
11. Ruchay, A., Feldman, E., Cherbadzhi, D., & Sokolov, A. (2023). *The Imbalanced Classification of Fraudulent Bank Transactions Using Machine Learning.* Mathematics, 11(13), 2862, pp. 1–21. MDPI.

12. Alarfaj, F. K., Jindal, H., Rawat, D. B., Goel, A., & Reddy, P. (2022*). Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms*. IEEE Access, 10, pp. 39700–39715. IEEE.

13. Banque d'Algérie. (2024). *Rapport annuel sur les moyens de paiement et la stabilité financière.* Alger : Banque d'Algérie, pp. 40–66.

14. Basel Committee on Banking Supervision. (2011). *Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems*. Basel : Bank for International Settlements (BIS), pp. 1–77.

15. Alexander, C. (2003). *Managing Operational Risks in Financial Markets.* Oxford : Butterworth-Heinemann, pp. 88–140.

16. Financial Action Task Force (FATF). (2023). *Illicit Financial Flows and Fraud Schemes.* Paris : OECD Publishing, pp. 10–52.

17. Verizon. (2023). *Data Breach Investigations Report (DBIR)*. New York : Verizon Enterprise, pp. 18–35.

18. Europol. (2023). *Internet Organised Crime Threat Assessment (IOCTA) 2023.* Luxembourg : Publications Office of the European Union, pp. 55–89.

19. Europol & INTERPOL. (2022). *Financial Crime and Anti-Fraud Centre Report.* Lyon : INTERPOL Publications, pp. 22–49.

20. Ileberi, E., Sun, Y., & Wang, Z. (2021). *Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost.* Heliyon, 7(1), e06349, pp. 1–15. Elsevier.

21. *COSO. (2017). Enterprise Risk Management: Integrating with Strategy and Performance.* New York : COSO Publishing, pp. 30–58.

22. Parlement européen & Conseil de l'Union européenne. (2015). *Directive (UE) 2015/2366 concernant les services de paiement dans le marché intérieur (DSP2)*. Journal officiel de l'Union européenne, L337, pp. 35–127.

23. Parlement européen & Conseil de l'Union européenne. *(2016). Règlement (UE) 2016/679 relatif à la protection des données personnelles (RGPD).* Journal officiel de l'Union européenne, L119, pp. 1–88.

24. International Organization for Standardization (ISO). (2013). *ISO/IEC 27001:2013 Information technology — Security techniques — Information security management systems requirements.* Genève : ISO, pp. 1–30.

25. Parlement européen & Conseil de l'Union européenne. (2022). *Directive (UE) 2022/2555 concernant un niveau élevé commun de cybersécurité dans l'Union (NIS2).* Journal officiel de l'Union européenne, L333, pp. 80–152.

26. Journal Officiel de la République Algérienne. (2018). *Loi n°18-07 du 10 juin 2018 relative à la protection des personnes physiques dans le traitement des données à caractère personnel.* JORA, No 34, pp. 3–18.

27. *CERT-Algeria. (2023). Rapport national sur les incidents de cybersécurité.* Alger : CERT-Algeria, pp. 12–40.

28. PCI Security Standards Council. (2022). *Payment Card Industry Data Security Standard (PCI DSS) Version 4.0.* Wakefield : PCI SSC, pp. 5–97.

29. Banque d'Algérie, GIE Monétique & CERT-Algeria. *(2023). Rapports consolidés sur la cybersécurité bancaire et les paiements électroniques en Algérie.* Alger, pp. 1–55.

30. Raoui, A., & El Gnaoui, L. (2023). *La fraude en agences bancaires : moyens de contrôle et mécanismes de prévention*. International Journal of Strategic Management and Economic Studies (IJSMES), 2(4), 45-58.

31. Tharwat, A. (2018). AdaBoost classifier: A comprehensive survey. *Information Fusion*, 42, 169-178. https://doi.org/10.1016/j.inffus.2017.10.001

32. Bank Al-Maghrib. (2014). *Circulaire relative au contrôle interne des établissements de crédit*. Rabat : Bank Al-Maghrib.

33. Committee of Sponsoring Organizations of the Treadway Commission (COSO). (2013). *Internal Control – Integrated Framework*. New York : COSO.

34. Arens, A. A., Elder, R. J., & Beasley, M. S. (2017). *Auditing and Assurance Services: An Integrated Approach* (16th ed.). Pearson Education.

35. Committee of Sponsoring Organizations of the Treadway Commission (COSO). (2013). *Three Lines of Defense Model*. Adapted from COSO Internal Control Framework.

36. Raoui, A., & El Gnaoui, L. (2023). *La fraude en agences bancaires : moyens de contrôle et mécanismes de prévention*. IJSMES.

37. Siruguet, J.-L., Barthélémy, J., & Kerviler, I. (2006). *Le contrôle interne bancaire et la fraude*. Paris : Dunod.

38. European Central Bank. (2022). *Digital Banking Risks Report*. Frankfurt : ECB Publications.

39. Konde, S. (2024). Supercharging fraud detection: Implementing real-time AI solutions in modern banking. *Medium*. https://medium.com

40. Fernando, C. (2009). What is Solution Architecture? *Medium / Architecture Notes.*

41. Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv preprint*, arXiv:1009.6119. https://arxiv.org/abs/1009.6119

42. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

43. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). Wiley.

44. Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

45. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

46. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139. https://doi.org/10.1006/jcss.1997.1504

47. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM. https://doi.org/10.1145/2939672.2939785

48. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. https://doi.org/10.1007/BF00994018

49. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. https://doi.org/10.1038/nature14539

50. Sadiq, M., & Namasivayam, S. (2024). *Machine learning based fraud detection framework in banking systems*. International Journal of Intelligent Systems and Applications in Engineering,16(2),55–68. https://ejournal.svgacademy.org/index.php/ijiee/article/view/189

51. Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2018). Credit card fraud detection: A realistic modeling and a novel learning strategy. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8), 3784-3797. https://doi.org/10.1109/TNNLS.2017.2736643

52. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139. https://doi.org/10.1006/jcss.1997.1504

53. Soni, B. (2023). Understanding Boosting in Machine Learning: A comprehensive guide. *Medium*. https://medium.com

54. Nguyen, T.-H., & Vu, A.-T. (2021). An efficient differential evolution for truss sizing optimization using AdaBoost classifier. *Computer Modeling in Engineering & Sciences (CMES)*, 128(2), 745-762. https://doi.org/10.32604/cmes.2021.015945

55. Van Otten, N. (2024, September 6). Confusion Matrix: A beginner's guide & how-to tutorial in Python. *Spot Intelligence*. https://spotintelligence.com

56. Kaggle. (2023). *Credit Card Fraud Detection Dataset*. Kaggle Dataset Repository. https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

57. Kavyasrirelangi. (2024). Understanding model evaluation metrics for machine learning. *Medium*. https://medium.com

58. Puthuchery, L. (2023). How to detect and prevent overfitting in machine learning models. *Medium*. https://medium.com

59. Kohavi, R., & Provost, F. (1998). Glossary of terms. *Machine Learning*, 30(2-3), 271-274. https://doi.org/10.1023/A:1006543900057

60. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

61. ResearchGate. (2024). *Schéma de courbe ROC en orange et AUC en gris*. ResearchGate

62. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324