الجمـهورية الجزائرية الديمقراطية الشعبيـة
وزارة التعليم العالي والبحث العلمي
جـامعـة سعـيـدة د. مـولاي الطـاهـر
كلية الرياضيات و الإعلام الآلي و الاتصالات السلكية و اللاسلكية
قسم: الإعلام الآلي

Mémoire de Master en informatique

Spécialité : Modélisation Informatique des Connaissances et du Raisonnement

Thème

Détection de faux comptes basée sur Intelligence artificielle-IA-

▪ **Présenté par :**
BAHLOULI ZAHRA

▪ **Dirigé par :**
Dr. HOUACINE ABDELKRIM

Année universitaire 2025-2026

Remerciement

Je tiens premièrement à me prosterner en remerciant Allah le Tout-Puissant de m'avoir donné le courage et la patience pour terminer ce travail.

Je voudrais dans un premier temps remercier mon directeur de mémoire, Monsieur , *Houacine Abdelkrim,* de m'avoir encadré, orienté, aidé et conseillé.

J'adresse mes sincères remerciements aux membres du jury, qui ont accepté d'évaluer mon travail.

Je remercie mes très chers parents, qui ont toujours été là pour moi. Et tout membre de ma famille m'a apporté un soutien moral tout au long de ma démarche.

Dédicace


mes chers parents

Je dédie ce modeste travail aux êtres les plus chers à mon cœur, ma mère et mon père, piliers de mon existence et source inépuisable de force. Ils ont veillé sur mon bien-être, mon bonheur et ont consenti d'innombrables sacrifices pour ma réussite. Que Dieu les protège, les préserve en bonne santé et veille sur eux.

À tous les membres de ma famille, petits et grands, je vous exprime tout mon amour, avec une mention particulière pour *Hadj Mohamed Hadjer*. Que Dieu te protège et t'accorde le succès, Amen. Merci pour votre soutien inestimable.

À toutes celles et ceux qui, ne serait-ce qu'avec un mot, m'ont donné la force d'avancer.

À tous ceux qui m'aiment et que j'aime.

الملخص

تشكل الحسابات الوهمية على إنستغرام تحديًا متزايدًا في مجال الأمن السيبراني، حيث تُستخدم لأغراض غير مشروعة مثل الاحتيال ونشر المعلومات المضللة والتلاعب بالرأي العام. تهدف هذه الدراسة إلى تطوير نموذج يعتمد على الذكاء الاصطناعي لاكتشاف الحسابات الوهمية باستخدام تقنيات التعلم الآلي. ولتحقيق ذلك، تم الاعتماد على مجموعة بيانات متوازنة تحتوي على تسع خصائص رئيسية تساعد في التمييز بين الحسابات الحقيقية والمزيفة، مثل وجود صورة للملف الشخصي، تطابق الاسم مع اسم المستخدم، وجود رابط خارجي، حالة الخصوصية، وعدد المتابعين والمنشورات والمتابعات. تم تطبيق عدة خوارزميات تصنيف، من بينها شجرة القرار، الغابة العشوائية، الانحدار اللوجستي، والشبكات العصبية الاصطناعية. أظهرت النتائج أن خوارزمية الغابة العشوائية كانت الأكثر دقة، مما يؤكد فعاليتها في كشف الحسابات المزيفة. تسلط هذه الدراسة الضوء على دور الذكاء الاصطناعي في تعزيز أمان منصات التواصل الاجتماعي، كما تفتح المجال لإجراء أبحاث مستقبلية تهدف إلى تطوير أنظمة أكثر دقة وفعالية في اكتشاف الحسابات الوهمية بشكل تلقائي.

الكلمات المفتاحية :الحسابات الوهمية، إنستغرام، التعلم الآلي، الأمن السيبراني، الذكاء الاصطناعي، التلاعب بالمعلومات، التصنيف الآلي.

Abstract

Fake account detection on Instagram is a crucial challenge in the field of cybersecurity, as malicious users exploit fake profiles for fraudulent activities. In this study, we propose a detection method based on machine learning algorithms to identify fake accounts using key profile attributes. We employ various classification models, including Decision Tree, Random Forest, Logistic Regression, and Artificial Neural Networks (ANN) using MLPClassifier. The dataset used is sourced from Kaggle and contains nine features that distinguish real accounts from fake ones. Experimental results demonstrate that the Random Forest model achieves the highest accuracy, proving its effectiveness in detecting fake accounts. This study highlights the role of artificial intelligence in enhancing security on social media platforms and lays the foundation for further improvements in automated fake account detection.

Keywords: Fake accounts, Instagram, Machine Learning, Cybersecurity, Social Engineering, Artificial Intelligence.

Résumé

La détection des faux comptes sur Instagram représente un défi majeur en cybersécurité, car des utilisateurs malveillants exploitent ces profils à des fins frauduleuses. Dans cette étude, nous proposons une méthode de détection basée sur des algorithmes d'apprentissage automatique afin d'identifier les faux comptes en utilisant des attributs clés des profils. Nous employons plusieurs modèles de classification, notamment l'arbre de décision, la forêt aléatoire, la régression logistique et les réseaux de neurones artificiels (ANN) avec **MLPClassifier**. Le jeu de données utilisé provient de Kaggle et comprend neuf caractéristiques permettant de distinguer les comptes réels des faux. Les résultats expérimentaux montrent que le modèle **Random Forest** atteint la meilleure précision ,prouvant ainsi son efficacité dans la détection des faux comptes. Cette étude met en évidence le rôle de l'intelligence artificielle dans l'amélioration de la sécurité sur les réseaux sociaux et ouvre la voie à de futures améliorations pour automatiser la détection des faux comptes.

Mots-clés : Faux comptes, Instagram, Apprentissage automatique, Cybersécurité, Ingénierie sociale, Intelligence artificielle.

Liste des abréviations

ANN: *Artificial neural network*

Bi-LSTM*: Bi-direction al Long Short Term Memory*

CNN: *Convolutional Neural Network*

LSTM : *Long Short Term Memory*

DL: *Deep Learning*

FP: *Faux Positifs*

FN: *Faux Négatifs*

IA: *Intelligence Artificielle*

MLP : *Multilayer Perceptron*

ML: *Machine Learning*

RNN: *Recurrent Neural Networks*

ROC : *Receiver Operating Characteristic*

VP: *Vrais Positifs*

VN: *Vrais Négatifs*

Table des Matières

Liste des Figures

Liste des Tableaux

Introduction Générale

Avec l'essor fulgurant des réseaux sociaux, notamment Instagram, ces plateformes sont devenues des outils incontournables pour la communication, le partage d'informations et les interactions sociales. Cependant, cette popularité a également entraîné l'émergence de nombreuses menaces cybersécuritaires, parmi lesquelles la prolifération des faux comptes. Ces comptes frauduleux sont souvent utilisés à des fins malveillantes, telles que l'usurpation d'identité, la diffusion de fausses informations, l'arnaque en ligne, la manipulation de l'engagement social et le cyberharcèlement.

L'ampleur du phénomène est alarmante, car ces comptes compromettent l'authenticité des interactions, influencent négativement la réputation des utilisateurs et des entreprises, et sont parfois exploités pour des activités criminelles. Instagram, comme d'autres réseaux sociaux, peine à éradiquer totalement ces faux comptes en raison de la sophistication croissante des techniques utilisées par les fraudeurs. Ces derniers adaptent continuellement leurs stratégies afin de contourner les méthodes de détection traditionnelles basées sur des règles statiques ou des signalements manuels.

Problématique

Face à cette menace croissante, il devient primordial de développer des méthodes automatisées et intelligentes pour détecter efficacement les faux comptes sur Instagram. Les approches conventionnelles, basées sur des analyses heuristiques et des vérifications manuelles, ne suffisent plus, notamment en raison de la grande quantité de données générées quotidiennement sur la plateforme. L'intelligence artificielle et l'apprentissage automatique apparaissent alors comme des solutions prometteuses pour analyser les profils et identifier les comportements suspects de manière efficace et évolutive.

Plusieurs questions clés se posent :

- Comment détecter automatiquement les faux comptes sur Instagram avec un haut niveau de précision ?
- Quels sont les algorithmes les plus performants pour cette tâche ?
- Comment garantir la robustesse du modèle face à l'évolution constante des techniques utilisées par les fraudeurs ?

Objectif

L'objectif principal de ce mémoire est de proposer un modèle performant basé sur l'apprentissage automatique pour la détection automatique des faux comptes sur Instagram. Notre approche repose sur l'analyse des caractéristiques des profils suspects en utilisant un jeu de données provenant de Kaggle. Ce dataset contient plusieurs indicateurs permettant de différencier les comptes réels des comptes frauduleux, notamment :

- La présence d'une photo de profil
- Le nombre de mots dans le nom complet
- La correspondance entre le nom d'utilisateur et le nom affiché
- L'existence d'une URL externe
- Le caractère privé ou public du compte
- Le nombre de publications, d'abonnés et d'abonnements

Pour l'entraînement du modèle, nous utiliserons et comparerons plusieurs algorithmes d'apprentissage automatique, notamment :

- La régression logistique (Logistic Regression)
- L'arbre de décision (Decision Tree)
- La forêt aléatoire (Random Forest)
- Les réseaux de neurones artificiels (Artificial Neural Networks - ANN)

Notre objectif est d'évaluer la performance de chaque algorithme afin d'identifier celui offrant la meilleure précision dans la classification des comptes. À terme, cette solution pourrait être intégrée à des systèmes de surveillance automatisée des réseaux sociaux, contribuant ainsi à la lutte contre la fraude et à la protection des utilisateurs.

Plan du Mémoire

Ce mémoire est structuré en **trois chapitres**, suivis d'une **conclusion générale** :

- **Le premier chapitre** : Nous présenterons les différentes approches existantes pour l'identification des faux comptes sur les réseaux sociaux, en mettant l'accent sur les mécanismes de création et de diffusion de ces comptes, ainsi que leurs impacts sur divers domaines. Nous étudierons également les méthodes classiques de détection, y compris les techniques de Web Mining et les modèles analytiques.
- **Le deuxième chapitre** : Nous expliquerons les concepts fondamentaux de l'intelligence

- artificielle et de l'apprentissage automatique. Nous détaillerons les principaux types de Machine Learning, en mettant l'accent sur les algorithmes utilisés dans notre étude, notamment l'arbre de décision, la forêt aléatoire, la régression logistique et les réseaux de neurones artificiels. Nous explorerons aussi les techniques de prétraitement des données pour améliorer la qualité des prédictions.
- **Le troisième chapitre** : Nous décrirons l'architecture de notre solution, en expliquant les étapes de préparation de l'environnement, de prétraitement des données et de sélection des caractéristiques pertinentes. Nous détaillerons également la mise en œuvre des modèles de classification, en comparant leurs performances selon différents critères d'évaluation (précision, rappel, score F1, etc.).

À travers cette étude, nous espérons démontrer l'importance de l'intelligence artificielle dans la détection des faux comptes sur les réseaux sociaux et proposer une solution efficace pour renforcer la sécurité numérique sur Instagram.

Chapitre 1

État de l'art sur : la détection des faux comptes

1.1. Introduction

Le but de ce chapitre est de présenter d'une part quelques rappels indispensables et nécessaires à la compréhension de ce mémoire, et d'autre part d'examiner certaines études antérieures relatives à la détection des faux comptes sur les plateformes de réseaux sociaux. Nous commencerons par définir les faux comptes et analyser les raisons de leur prolifération, avant d'aborder les différentes méthodes utilisées pour les détecter. Nous mettrons particulièrement l'accent sur les techniques de Data Mining et d'analyse du comportement des utilisateurs, en se basant sur l'application des algorithmes d'apprentissage automatique, notamment l'arbre de décision, la forêt aléatoire et la régression logistique

1.2. La signification des comptes faux

Avec l'essor des réseaux sociaux, la prolifération des comptes faux est devenue un phénomène préoccupant. Ces comptes, souvent créés à des fins malveillantes, menacent l'intégrité des interactions en ligne et la sécurité des utilisateurs. Ils peuvent être utilisés pour manipuler l'opinion publique, propager des informations erronées, ou commettre des fraudes [1].

Caractéristiques des comptes faux

Les comptes faux se distinguent par plusieurs traits communs :

Informations de profil incomplètes ou suspectes

Utilisation de noms d'utilisateur générés aléatoirement ou fictifs

Absence de photo de profil ou utilisation d'images volées sur Internet

Profils manquant de détails personnels ou contenant des informations incohérentes

Comportements d'interaction anormaux

Publication d'un volume élevé de contenu en un temps record

Interactions automatisées, telles que des "likes" ou des commentaires répétitifs [1]

Faible engagement naturel avec d'autres utilisateurs

Structures de réseau suspectes

Nombre élevé d'abonnés sans interactions significatives

Adhésion à des groupes ou communautés de manière aléatoire ou illogique

Modèles de publication répétitifs

- Partage du même contenu sur plusieurs comptes
- Diffusion excessive de liens promotionnels ou de spams

Utilisation de bots

- Automatisation des tâches, comme suivre ou se désabonner massivement
- Utilisation de scripts pour simuler des interactions humaines

Activités de connexion suspectes

- Connexions depuis plusieurs localisations géographiques en peu de temps .
- Utilisation d'adresses IP associées à des activités frauduleuses

1.2.3 Définition et spécificités d'Instagram

Instagram est une plateforme de réseau social centrée sur le partage de contenu visuel, fondée en 2010 et acquise par le groupe Meta (anciennement Facebook Inc.) en 2012. À l'horizon 2024, elle compte environ **2,4 milliards d'utilisateurs actifs mensuels**, ce qui en fait l'un des réseaux sociaux les plus utilisés dans le monde (Meta, 2024).

Sa spécificité réside dans sa **dimension visuelle et immersive**, favorisant l'interaction à travers des images, des vidéos et des éléments multimédias à fort impact. Contrairement à d'autres plateformes orientées texte, Instagram repose sur une **logique de mise en scène visuelle de soi et de sa vie quotidienne**, ce qui influence fortement les comportements des utilisateurs et des influenceurs.

Trois **formats de contenu principaux** caractérisent cette plateforme :

- **Publications permanentes** : il s'agit de photos ou vidéos publiées dans le fil principal du profil. Ces contenus sont durables et accessibles à tout moment, servant souvent à construire une image de marque ou une identité numérique stable.

- **Stories** : introduites en 2016, les stories sont des contenus éphémères (photos ou vidéos) visibles pendant **24 heures**. Elles favorisent une communication spontanée, dynamique et temporaire, et sont particulièrement prisées pour leur taux d'engagement élevé.
- **Reels** : lancés pour concurrencer TikTok, les Reels sont des **vidéos courtes et dynamiques** avec des effets sonores ou visuels. Ce format accentue le caractère viral du contenu et attire un public jeune et mobile.

1.3 Les composants des faux comptes sur les réseaux sociaux

Les faux comptes sur les réseaux sociaux constituent une préoccupation croissante en raison de leur utilisation potentielle pour propager de la désinformation, commettre des fraudes et manipuler l'opinion publique. Ces comptes peuvent être décomposés en quatre composants principaux : **créateur/contrôleur**, **victimes ciblées**, **comportement du compte** et **contexte numérique**.

Créateur/Contrôleur

Les faux comptes sont généralement créés par des opérateurs humains utilisant de fausses

identités ou par des robots (bots) automatisés conçus pour générer des comptes de manière programmatique. Ces créateurs visent souvent à mener des activités malveillantes telles que le spam, le phishing ou l'influence du discours public. Selon [3], les comptes générés par des bots sont de plus en plus sophistiqués, imitant le comportement humain pour éviter la détection.

Victimes ciblées

Les cibles principales des faux comptes incluent les utilisateurs individuels des réseaux sociaux, les entreprises, les influenceurs et même les institutions politiques. Ces comptes sont souvent utilisés pour tromper ou manipuler leurs cibles, que ce soit pour un gain financier ou pour propager de la désinformation. Par exemple, [4] met en lumière comment des faux comptes ont été utilisés pour influencer des campagnes politiques en ciblant des démographies spécifiques.

Comportement du compte

Les faux comptes présentent des comportements distincts, tels que la publication d'un grand volume de contenu en peu de temps, des interactions répétitives (comme des "likes" ou des commentaires), ou le suivi et le désabonnement massif d'autres comptes. Ces comportements

sont souvent automatisés et peuvent être détectés à l'aide de techniques d'apprentissage automatique, comme discuté dans [2].

. Contexte numérique

Le contexte numérique fait référence à la manière dont les faux comptes opèrent au sein de l'écosystème des réseaux sociaux. Cela inclut l'analyse des schémas d'interaction, des connexions avec d'autres comptes et leur rôle dans des campagnes coordonnées de désinformation. Des études comme [5] ont démontré que ces comptes utilisent des comportements automatisés ou semi-automatisés pour amplifier artificiellement certains contenus, manipuler l'opinion publique et contourner les systèmes de modération des plateformes.

1.3.1 Créateurs et mécanismes de diffusion des comptes faux

La prolifération des comptes faux sur les réseaux sociaux est un phénomène complexe qui nécessite une compréhension approfondie des acteurs impliqués et des mécanismes de diffusion. Les créateurs de ces comptes peuvent être classés en deux catégories principales : les **individus réels** et les **robots (bots)**. Leurs motivations et stratégies varient, mais ils partagent un objectif commun : exploiter les réseaux sociaux à des fins malveillantes.

- **Individus réels**
  Certains utilisateurs créent des comptes faux en utilisant des identités fictives pour atteindre des objectifs spécifiques, tels que la manipulation de l'opinion publique, la fraude en ligne ou la diffusion de fausses informations. Ces comptes sont souvent utilisés dans des campagnes politiques ou commerciales pour obtenir des avantages personnels ou financiers. Selon [4], ces acteurs exploitent les faiblesses des plateformes sociales pour maximiser leur impact.
- **Robots et systèmes automatisés (Bots)**
  Les bots sont des logiciels intelligents conçus pour générer et gérer un grand nombre de comptes faux de manière automatisée. Ils sont utilisés pour diffuser massivement du contenu, interagir avec des publications afin d'amplifier leur visibilité, ou mener des campagnes organisées visant à influencer les utilisateurs réels. Les bots modernes imitent de plus en plus le comportement humain, ce qui rend leur détection plus difficile, comme le souligne [3].

Mécanismes de diffusion

Les comptes faux se propagent grâce à des stratégies bien définies, notamment :

- **Amplification artificielle**
- Les bots interagissent massivement avec certains contenus pour leur donner une visibilité disproportionnée.
- **Campagnes coordonnées**
- Des réseaux de comptes faux travaillent ensemble pour promouvoir des messages spécifiques ou discréditer des informations concurrentes.

Exploitation des tendances : Les créateurs de comptes faux exploitent les sujets

1.3.2 Contenu des faux comptes

Chaque faux compte sur les réseaux sociaux possède un **contenu visible** et un **contenu comportemental**, qui peuvent être analysés pour identifier des signes de fraude ou de manipulation.

Contenu visible des faux comptes

Le contenu visible comprend des éléments tels que :

- **Nom du profil** : Souvent généré de manière aléatoire ou fictive [1].
- **Photo de profil** : Utilisation d'images génériques ou volées sur Internet [3].
- **Biographie** : Descriptions incomplètes ou incohérentes [5].
- **Publications** : Contenu répétitif, promotionnel ou malveillant [1].
- **Hashtags et mentions** : Utilisation excessive de mots-clés populaires pour maximiser la visibilité [3].

L'analyse de ces caractéristiques permet de détecter des incohérences qui indiquent la présence de comptes frauduleux. Par exemple, les faux comptes utilisent souvent des images de profil génériques et des noms d'utilisateur aléatoires pour éviter d'être détectés [5].

. Contenu comportemental des faux comptes

Le contenu comportemental se réfère aux modèles d'activité des comptes, tels que :

- **Fréquence des publications** : Activité excessive en peu de temps [1].

- **Connexions avec d'autres comptes** : Réseaux de bots interconnectés ou interactions unilatérales [3].
- **Engagement automatisé** : Commentaires ou likes répétitifs générés par des scripts [5].
- **Utilisation de bots** : Tâches automatisées comme suivre ou se désabonner massivement [1].

1.3.3 Contexte social des faux comptes

Le contexte social des faux comptes fait référence à l'environnement numérique dans lequel ces comptes opèrent, exploitant les dynamiques des réseaux sociaux pour diffuser de fausses informations et manipuler l'opinion publique. Ces comptes s'intègrent dans des communautés spécifiques, telles que des groupes partageant des intérêts communs ou des réseaux d'utilisateurs influents, afin de créer une crédibilité artificielle pour le contenu qu'ils publient.

- **Amplification artificielle**
  Lorsque ces comptes fonctionnent en réseau, ils amplifient la diffusion de contenu trompeur en augmentant artificiellement l'engagement. Par exemple, ils génèrent des mentions "J'aime" fictives, des commentaires automatisés et des partages répétés. Cette activité artificielle renforce la visibilité du contenu faux dans les algorithmes des plateformes sociales, facilitant ainsi sa propagation massive et son impact sur les utilisateurs réels. Selon [5], cette stratégie est couramment utilisée dans les campagnes de désinformation pour maximiser la portée des messages trompeurs.
- **Impact sur les communautés en ligne**
  Les faux comptes exploitent également les biais cognitifs et les tendances sociales pour influencer les comportements des utilisateurs. Par exemple, ils ciblent des groupes vulnérables ou polarisés pour renforcer des croyances existantes ou semer la confusion. Comme le souligne [6], ces comptes peuvent créer des échos de chambre (echo chambers) où les fausses informations se répètent et se renforcent, rendant leur détection plus difficile.

1.3.4 Composants des faux comptes sur Instagram

1.3.4.1 Mécanismes de création

Les créateurs exploitent plusieurs failles :

- **Outils automatisés** : Logiciels comme InstaBot (500 000 téléchargements en 2023) permettant de générer des milliers de comptes
- **Faiblesses du système** :
    - Absence de vérification biométrique pour les comptes standards
    - Possibilité de lier plusieurs comptes à un même email

1.3.4.2 Contenu caractéristique

- **Photos volées** : 40% des profils utilisent des images provenant de banques d'images (étude TinEye 2024)
- **Hashtags trompeurs** : Usage massif de tags populaires (#love, #fashion) pour augmenter la visibilité

1.4 Modèles de détection des faux comptes :

La détection des faux comptes sur les réseaux sociaux repose sur plusieurs modèles analytiques qui exploitent différentes caractéristiques des comptes suspects. Ces modèles peuvent être classés en deux catégories principales : modèles basés sur le contenu des profils et modèles basés sur le comportement et les interactions.

Modèles basés sur le contenu des profils

Ces modèles se concentrent sur l'analyse des informations visibles d'un compte pour identifier des incohérences ou des signes de falsification. Ils incluent :

Modèle basé sur les métadonnées du profil

Cette approche examine des caractéristiques telles que le nom d'utilisateur, la photo de profil, la biographie et l'historique des publications. Les faux comptes utilisent souvent des images génériques, des noms artificiels et des descriptions incomplètes. Selon [4], ces caractéristiques peuvent être utilisées pour entraîner des algorithmes de classification afin de distinguer les comptes réels des comptes faux.

Modèle basé sur les modèles de publication

Les faux comptes affichent souvent des schémas de publication anormaux, tels qu'une activité excessive en peu de temps, du contenu copié-collé ou un manque de diversité dans les publications. Comme le souligne [3], l'analyse des modèles de publication peut révéler des comportements suspects, notamment la répétition de contenu ou l'utilisation de liens malveillants.

Modèles basés sur le comportement et les interactions

Ces modèles exploitent les relations sociales et l'engagement des comptes suspects pour identifier des comportements automatisés ou frauduleux. Ils incluent :

Modèle basé sur le réseau social

Cette approche analyse les connexions entre les comptes, car les faux comptes sont souvent liés à d'autres profils frauduleux, formant des structures anormales dans les graphes sociaux. Selon [5], l'analyse des réseaux sociaux peut révéler des clusters de comptes faux qui interagissent entre eux de manière suspecte.

Modèle basé sur l'analyse de l'engagement

Les faux comptes génèrent souvent des interactions artificielles, telles que des "likes" et des commentaires automatisés, ou suivent un grand nombre d'utilisateurs sans recevoir d'engagement en retour. Comme le montre [6], ces comportements peuvent être détectés en analysant les schémas d'interaction et les ratios d'engagement.

Détection des bots et des comportements automatisés

Cette technique repose sur l'analyse des modèles temporels, où les faux comptes affichent des interactions répétitives, une activité à des intervalles précis et l'utilisation de scripts automatisés pour générer du contenu. Selon [2], les bots modernes imitent de plus en plus le comportement humain, ce qui rend leur détection plus difficile.

1.5 Conséquences des faux comptes sur les différents domaines

Les faux comptes sur les réseaux sociaux ont un impact significatif dans divers domaines, allant de la sécurité numérique à la manipulation de l'opinion publique. Leurs effets néfastes se manifestent à travers plusieurs mécanismes, notamment la propagation de la désinformation, la fraude en ligne, et la manipulation des marchés financiers.

Propagation de la désinformation et manipulation de l'opinion publique

Les faux comptes sont souvent utilisés pour diffuser de fausses informations et influencer les débats publics, en particulier pendant les périodes électorales ou les crises sociales. Selon [6], ces comptes exploitent les algorithmes des plateformes sociales pour amplifier la portée des messages trompeurs, créant ainsi des échos de chambre (echo chambers) où les fausses informations se répètent et se renforcent. Cela peut entraîner une polarisation accrue des

opinions et une érosion de la confiance dans les médias traditionnels.

Fraude et cybercriminalité

Les faux comptes sont fréquemment utilisés par les cybercriminels pour tromper les utilisateurs, voler des informations personnelles ou mener des attaques de phishing. Par exemple, [4] souligne que les faux comptes sur des plateformes comme Facebook ou Twitter sont souvent impliqués dans des campagnes de phishing ciblant des utilisateurs vulnérables. Ces activités frauduleuses peuvent causer des pertes financières importantes et compromettre la sécurité des données personnelles.

Manipulation des marchés financiers

Les faux comptes peuvent également être utilisés pour altérer artificiellement l'image d'une marque ou d'une entreprise, influençant ainsi la valeur des actions et la confiance des investisseurs. Selon [5], des campagnes coordonnées de désinformation peuvent être menées pour diffuser de fausses nouvelles sur les performances d'une entreprise, entraînant des fluctuations anormales des cours boursiers. Cela pose des défis majeurs pour la régulation des marchés financiers et la protection des investisseurs.

Impact sur la sécurité nationale

Les faux comptes sont parfois utilisés par des acteurs étatiques ou non étatiques pour mener des campagnes d'influence visant à déstabiliser des pays ou des régions. Comme le montre [3], ces campagnes peuvent inclure la diffusion de propagande, la manipulation des médias sociaux, et l'exploitation des tensions sociales pour atteindre des objectifs politiques ou stratégiques.

1.6 Détection des faux comptes à l'aide du Web Mining

La technologie du **Web Mining** a été largement utilisée pour identifier et analyser les faux comptes sur les réseaux sociaux en extrayant des informations pertinentes à partir des profils et des interactions en ligne. Cette approche permet de détecter les comptes automatisés et frauduleux en exploitant différentes catégories de données. Selon [7], le Web Mining peut être classé en trois catégories principales en fonction du type de données analysées :

Web Content Mining

Cette méthode se concentre sur l'analyse des informations visibles sur les profils des

utilisateurs, telles que les publications, les commentaires et les métadonnées associées. Par exemple, les faux comptes utilisent souvent des images génériques, des noms d'utilisateur aléatoires et des descriptions incomplètes. Selon [4], l'analyse du contenu des profils peut révéler des incohérences qui indiquent la présence de comptes frauduleux.

Web Structure Mining

Cette approche étudie les connexions entre les comptes pour identifier des schémas inhabituels, tels que des réseaux de bots interconnectés. Les faux comptes ont tendance à former des clusters ou des structures anormales dans les graphes sociaux. Comme le souligne [3], l'analyse des réseaux sociaux peut révéler des modèles de connexion suspects, tels que des comptes qui interagissent principalement entre eux sans liens avec des utilisateurs réels.

Web Usage Mining

Cette technique analyse les comportements des utilisateurs, y compris la fréquence des interactions, la rapidité des réponses et l'automatisation des actions. Les faux comptes affichent souvent des comportements anormaux, tels qu'une activité excessive en peu de temps ou des interactions répétitives. Selon [5], l'analyse des modèles d'utilisation peut aider à identifier les comptes automatisés en détectant des schémas temporels inhabituels.

Applications et défis

Le Web Mining offre des outils puissants pour la détection des faux comptes, mais il présente également des défis. Par exemple, les bots modernes imitent de plus en plus le comportement humain, ce qui rend leur détection plus difficile. De plus, l'analyse de grandes quantités de données en temps réel nécessite des ressources informatiques importantes. Selon [6], l'intégration de techniques d'apprentissage automatique et d'intellige

1.6.1 Processus de détection des faux comptes

Le processus de détection des faux comptes repose sur l'analyse des données et l'utilisation de techniques d'intelligence artificielle. Ce processus se déroule en plusieurs étapes clés, chacune jouant un rôle essentiel dans l'identification des comptes frauduleux [8].

Collecte des données

La première étape consiste à extraire des informations pertinentes sur les comptes suspects. Ces informations incluent :

- Les dates de création des comptes.
- Les modèles de publication (fréquence, type de contenu, etc.).
- Les listes d'amis ou d'abonnés.
- Les métadonnées associées aux publications (heure, localisation, etc.).

Selon [4], la collecte de données précises et complètes est cruciale pour garantir l'efficacité des étapes suivantes.

Analyse du comportement des utilisateurs

Une fois les données collectées, l'étape suivante consiste à analyser le comportement des utilisateurs. Cette analyse inclut :

- La fréquence des publications et des interactions.
- L'utilisation de liens externes ou de contenus promotionnels.
- La répétition de commentaires ou de messages similaires.

Les faux comptes affichent souvent des comportements anormaux, tels qu'une activité excessive en peu de temps ou des interactions automatisées. Comme le souligne [3], ces schémas peuvent être détectés grâce à des techniques d'analyse temporelle et comportementale.

Classification des comptes

La dernière étape consiste à classer les comptes comme "authentiques" ou "faux" en utilisant des modèles d'apprentissage automatique. Ces modèles s'appuient sur un ensemble de caractéristiques comportementales et de contenu pour prendre une décision. Par exemple :

- Les algorithmes de classification comme les arbres de décision, les réseaux de neurones ou les machines à vecteurs de support (SVM) sont couramment utilisés [5].
- Les caractéristiques utilisées pour la classification incluent la cohérence des informations du profil, les schémas d'interaction et les modèles de publication.

Selon [6], l'intégration de techniques d'apprentissage profond (deep learning) peut améliorer la précision de la classification en capturant des motifs complexes dans les données.

nce artificielle peut améliorer l'efficacité des systèmes de détection.

1.6.2 Objectifs de la détection des faux comptes :

Les techniques de détection des faux comptes visent à atteindre plusieurs objectifs stratégiques pour améliorer la qualité et la sécurité des interactions en ligne. Ces objectifs incluent la réduction de la désinformation, la protection des utilisateurs contre les cybermenaces, et la création d'un environnement numérique plus fiable [5].

Réduction de la désinformation

L'un des principaux objectifs de la détection des faux comptes est de limiter la propagation de fausses informations et de contenus trompeurs. Les faux comptes sont souvent utilisés pour amplifier des campagnes de désinformation, en particulier pendant les périodes électorales ou les crises sociales. Selon [6], la suppression de ces comptes permet de réduire l'impact des campagnes de manipulation et de restaurer la confiance dans les informations partagées en ligne.

Amélioration de la sécurité des utilisateurs

Les faux comptes représentent une menace majeure pour la sécurité des utilisateurs, car ils sont souvent utilisés pour mener des attaques de phishing, voler des données personnelles ou diffuser des logiciels malveillants. En identifiant et en supprimant ces comptes, les plateformes sociales peuvent protéger leurs utilisateurs contre les risques liés à la cybercriminalité. Comme le souligne [4], la détection précoce des comptes frauduleux est essentielle pour prévenir les dommages financiers et psychologiques.

Création d'un environnement en ligne plus fiable

La présence de faux comptes et de bots malveillants érode la crédibilité des interactions en ligne. En éliminant ces comptes, les plateformes peuvent renforcer la confiance des utilisateurs et promouvoir des échanges plus authentiques. Selon [3], un environnement numérique fiable est essentiel pour encourager une participation constructive et réduire l'influence des manipulateurs d'opinion.

1.6.3 Techniques de détection adaptées à Instagram

1.6.3.1 Méthodes avancées

1. Analyse comportementale
    - Détection des actions non-humaines (ex : 200 likes/minute)
    - Reconnaissance des schémas d'activité (publications à heures fixes)
2. Vérification du contenu :

- Recherche d'images inversée via Google Lens
- Détection de deepfakes par analyse des métadonnées EXIF

1.6.3.2 Défis spécifiques

- Limites technologiques : Les outils classiques de text mining (efficaces sur Twitter) sont inadaptés au contenu visuel
- **Évolution constante** : 30% des bots contournent les détections chaque trimestre (rapport Meta Q1 2024)

1.7 Détection des Comptes Faux (Fake Account Detection)

Avec l'essor des réseaux sociaux, les comptes frauduleux sont devenus un outil majeur pour manipuler l'information, influencer l'opinion publique et mener des activités malveillantes telles que le spam, le phishing et la désinformation. Ces comptes se caractérisent souvent par des schémas de publication automatisés, un faible taux d'interaction humaine et une activité anormalement élevée. Pour les identifier, des approches basées sur l'apprentissage automatique, l'analyse comportementale et l'exploration de graphes sont utilisées. Ces techniques permettent de repérer des anomalies dans les connexions, les fréquences de publication et les interactions sociales, améliorant ainsi la détection des faux comptes sur les plateformes numériques.

Apprentissage automatique (Machine Learning)

Les algorithmes d'apprentissage automatique sont utilisés pour classer les comptes en fonction de caractéristiques telles que les informations du profil, les modèles de publication et les interactions. Par exemple, des algorithmes comme **Random Forest** et **Support Vector Machines (SVM)** ont démontré une efficacité élevée pour distinguer les comptes faux des comptes authentiques [5].

Analyse comportementale (Behavioral Analysis)

L'analyse comportementale se concentre sur l'étude des activités des utilisateurs, telles que la fréquence des publications, les interactions répétitives et les connexions suspectes. Cette approche permet de détecter des comportements anormaux qui indiquent la présence de comptes automatisés ou frauduleux [5].

Analyse des graphes (Graph Analysis)

L'analyse des graphes examine les relations entre les comptes pour identifier des structures

anormales, telles que des réseaux de bots interconnectés. Cette technique est particulièrement utile pour détecter des campagnes organisées de désinformation ou de manipulation [5]

1.7.1 Définition

La détection des faux comptes (Fake Account Detection) est un domaine de recherche qui vise à identifier les comptes frauduleux sur les réseaux sociaux en analysant leur comportement, leurs interactions et leurs caractéristiques. Avec la popularité croissante des plateformes telles que Facebook, Twitter et Instagram, le nombre de faux comptes a augmenté de manière significative, entraînant des risques majeurs en matière de désinformation, d'escroquerie et de cybercriminalité. Grâce aux progrès de l'intelligence artificielle et du Big Data, il est désormais possible de détecter ces comptes en se basant sur des critères tels que la fréquence des publications, les schémas de connexion et l'analyse des réseaux sociaux [3].

1.7.2 Méthodes d'analyse des faux comptes

La détection des faux comptes sur les réseaux sociaux vise à identifier si un compte est authentique ou frauduleux en se basant sur plusieurs facteurs, tels que le comportement de l'utilisateur, la nature des publications et le réseau de relations. Plusieurs méthodes sont utilisées pour détecter ces comptes de manière automatique, parmi lesquelles :

Apprentissage automatique supervisé

Cette approche repose sur l'entraînement d'un modèle d'intelligence artificielle à l'aide d'un ensemble de données contenant des comptes réels et faux. Des caractéristiques telles que le taux de publication, le nombre d'abonnés, le ratio d'interactions et le mode de création du compte sont extraites pour entraîner le modèle. Une fois entraîné, le modèle peut prédire automatiquement si un compte est frauduleux ou non. Selon [38], des algorithmes comme **Random Forest** et **Support Vector Machines (SVM)** ont démontré une efficacité élevée dans cette tâche.

Analyse des réseaux sociaux

Cette méthode consiste à examiner la structure des relations entre les comptes à l'aide d'algorithmes graphiques tels que **PageRank** et **Graph Centrality**. Les faux comptes forment souvent des grappes artificielles et présentent des schémas de connexion distincts des comptes authentiques. Comme le souligne [6], l'analyse des réseaux sociaux permet de détecter des clusters de comptes suspects qui interagissent principalement entre eux.

Analyse des comportements et du langage

Une autre approche consiste à détecter les faux comptes en analysant leur style d'écriture et leurs interactions. Ces comptes affichent généralement des textes répétitifs, des liens suspects ou une faible interaction avec les autres utilisateurs. Les techniques de traitement du langage naturel (NLP) permettent d'identifier automatiquement ces modèles. Selon [4], l'analyse linguistique peut révéler des anomalies dans le contenu publié par les faux comptes.

1.8 Travaux connexes

- **Détection automatique de comptes multiples dans les médias sociaux** :Cette étude propose une méthodologie pour identifier les comptes multiples créés par un même utilisateur sur des plateformes collaboratives, en se basant sur l'analyse des contributions et des interactions.

- **Détection de fausses informations dans les réseaux sociaux** : Cette thèse analyse les publications sur les réseaux sociaux d'un point de vue multimodal, en combinant le texte et l'image associée, afin de détecter les fausses informations.
- **Le social media listening au service de la détection des fake news** : Cet article explore comment les outils d'écoute et d'analyse des médias sociaux peuvent aider à lutter contre la désinformation en surveillant et en analysant les contenus partagés en ligne.
- **L'IA traque les faux comptes des réseaux sociaux** : Des chercheurs californiens utilisent l'intelligence artificielle pour identifier les bots sur Twitter, en comparant l'activité des comptes automatisés à celle des utilisateurs humains.
- **Preemptive Detection of Fake Accounts on Social Networks via Multi-Class Preferential Attachment Classifiers** : Cette recherche propose un nouvel algorithme pour détecter de manière proactive les faux comptes sur les réseaux sociaux, en se basant sur les modèles d'attachement préférentiel.
- **Detecting fake accounts through Generative Adversarial Network in online social media** : Cette étude propose une méthode utilisant des mesures de similarité entre utilisateurs et des réseaux antagonistes génératifs (GAN) pour identifier les faux comptes sur Twitter.
- **Friend or Faux : Graph-Based Early Detection of Fake Accounts on Social Networks** : Cette recherche présente l'algorithme SybilEdge, qui détermine si un nouvel

utilisateur est un faux compte en analysant ses choix de demandes d'amis et les réponses correspondantes.

1.9 Conclusion

Dans ce chapitre, nous avons présenté les différents concepts liés aux faux comptes sur les réseaux sociaux. Nous avons examiné leur définition, leurs caractéristiques ainsi que les diverses approches utilisées pour leur détection. L'application du Web Mining et des techniques de Data Mining a permis de développer des solutions efficaces pour identifier et analyser les comportements suspects associés aux comptes frauduleux. Dans le chapitre suivant, nous allons nous concentrer sur les méthodes d'apprentissage automatique et comparer différentes approches afin de sélectionner la plus appropriée pour la détection des faux comptes

.

Chapitre 2

Intelligence Artificielle et Apprentissage Automatique

2.1 Introduction

L'intelligence artificielle (IA) a profondément transformé de nombreux domaines en permettant aux machines d'analyser et d'interpréter des données complexes. Parmi ses branches fondamentales, on retrouve l'apprentissage automatique et l'apprentissage profond, qui jouent un rôle clé dans divers domaines, y compris la cybersécurité et la détection des faux comptes sur les réseaux sociaux.

L'apprentissage automatique comprend plusieurs types, notamment l'apprentissage supervisé, non supervisé, semi-supervisé et par renforcement, chacun ayant des applications spécifiques en fonction de la disponibilité et de la nature des données. En parallèle, l'apprentissage profond s'appuie sur des réseaux de neurones avancés, tels que les réseaux convolutifs (CNN), particulièrement adaptés à l'analyse d'images, et les réseaux récurrents (RNN), utilisés pour traiter les données séquentielles comme le texte et les séries temporelles. Parmi ces modèles, des variantes comme LSTM et Bi-LSTM permettent d'améliorer la gestion des dépendances à long terme dans les séquences.

Ce chapitre explore les fondements de l'intelligence artificielle et de l'apprentissage automatique, en mettant en avant les différentes approches et architectures de réseaux de neurones. Ces technologies constituent la base de nombreuses solutions modernes de détection, notamment dans l'identification des faux comptes sur les plateformes en ligne.

2.2 L'intelligence artificielle

L'intelligence artificielle est un domaine d'étude très large dans lequel les machines présentent des capacités cognitives telles que le comportement appris, l'interaction active avec l'environnement, le raisonnement et la déduction, la vision par ordinateur, la reconnaissance vocale, la résolution de problèmes, la représentation des connaissances, la perception, etc [9]. Plus familièrement, l'intelligence artificielle fait référence à toute activité dans laquelle une machine imite le comportement intelligent normalement manifesté par les humains. L'intelligence artificielle s'inspire des éléments de l'informatique, des mathématiques et des statistiques [9].

2.3 L'apprentissage automatique (Machine Learning, ML)

Le Machine Learning, ou apprentissage automatique en français, est un sous-domaine del'intelligenceartificielle(IA).Ilapourobjectifdecomprendrelastructuredesdonnées

etdelesintégrerdansdesmodèlesquipeuventêtrecomprisetutiliséspourtesterd'autre données. Le Machine Learning est négrâceaux technologies de reconnaissance de pattern etàlathéorieselonlaquellelesordinateurspeuventapprendresansêtreprogramméspour eectuer des tâches spécifiques (ex; classification, régression). Les chercheurs intéressés par l'intelligence artificielle souhaitent vérifier si les ordinateurs pouvaient apprendre à partir de données et s'adapter avec les nouvelles données [10].

2.3.1 Types de Machine Learning

Les algorithmes de Machine Learning ne sont pas une nouveauté, mais ce n'est que depuis peu qu'il est possible d'appliquer des calculs mathématiques complexes de plus en plus vite au Big Data; (un ensemble très volumineux de données qui nécessite un nouvel outil de gestion des bases de données). Le Machine Learning est aujourd'hui utilisé dans de nombreux domaines, t l que le développement de véhicules autonomes , les systèmes de recommandations en ligne (ex; Netflix et Amazon), l'analyse des sentiments des clients, ou encore la détection de fraude [10]. Les modèles d'apprentissage automatique peuvent être résumés comme suit: modèles supervisés, non-supervisés, semi-supervisé et par renforcement.

2.3.1.1 Apprentissage supervisé

La tâche d'apprentissage automatique de l'apprentissage supervisé consiste à former une fonction qui traduit une entrée en une sortie à l'aide d'exemples de paires entrée sortie. Il dérive une fonction à partir de données de formation étiquetées, constituées d'une collection de cas pratiques. Les algorithmes qui nécessitent une aide extérieure sont connus sous le nom d'algorithmes d'apprentissage automatique supervisé. L'ensemble de données d'entrée est séparé en ensembles de données de formation et de test. L'ensemble de données de formation contient une variable de sortie qui nécessite une prédiction ou une classification. Tous les algorithmes utilisent l'ensemble de données de formation pour découvrir divers modèles qu'ils utilisent ensuite dans l'ensemble de données de test pour faire des prédictions ou catégoriser les données [11]. L'apprentissage supervisé a été appliqué avec succès dans plusieurs domaines tels que : La recherche d'informations, l'analyse de marché , l'exploration de données, la vision par ordinateur, la reconnaissance vocale, la détection de spam, la bioinformatique, la chimioinformatique et l'exploration de données [12].

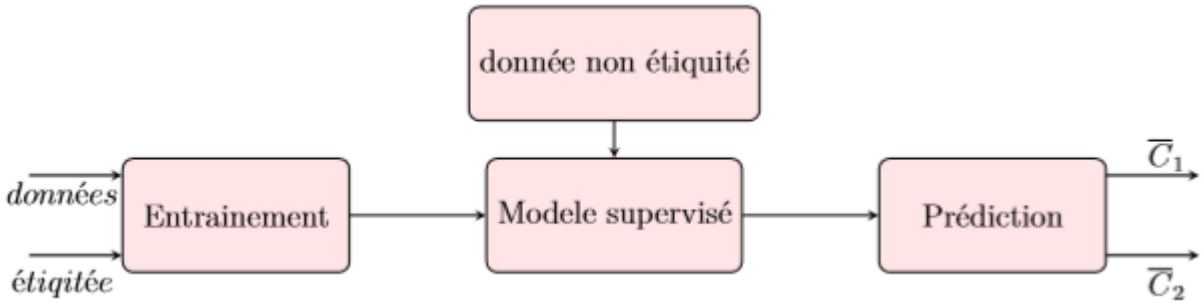


Figure 2.1: Schéma explicatif de l'apprentissage supervisé

2.3.1.2 Apprentissage non supervisé

L'apprentissage non supervisé est une approche dans laquelle un algorithme est entraîné sur des données non étiquetées, lui permettant d'identifier de manière autonome des structures et des relations sous-jacentes dans les données. L'objectif principal est de détecter des modèles cachés ou d'apprendre des représentations significatives pour organiser et interpréter les données brutes. Cette technique est largement utilisée dans la classification automatique et la segmentation de données, notamment dans le cadre du clustering, qui consiste à regrouper des éléments similaires en fonction de caractéristiques partagées. Par exemple, un système de recommandation peut proposer des produits ou des contenus à un utilisateur en fonction des préférences d'autres utilisateurs ayant des comportements similaires. [13]

2.3.1.3 Apprentissage semi-supervisé

L'apprentissage semi-supervisé combine des données étiquetées et non étiquetées pour entraîner un modèle. Généralement, une petite proportion de données étiquetées est utilisée en complément d'un grand volume de données non étiquetées, car ces dernières sont moins coûteuses et plus faciles à collecter. Cette approche est particulièrement employée dans des tâches telles que la classification, la régression et la prédiction. L'apprentissage semi-supervisé est avantageux lorsque l'annotation manuelle des données est coûteuse ou difficile à réaliser à grande échelle. Par exemple, cette technique est appliquée dans la reconnaissance faciale, permettant d'identifier une personne à partir d'images capturées par une webcam.[14]

2.3.1.4 Apprentissage par renforcement

L'apprentissage par renforcement est une sous-discipline de l'apprentissage automatique axée sur la prise de décision. Il repose sur un agent qui interagit avec un environnement et

apprend à adopter des actions maximisant une récompense cumulative. Contrairement à l'apprentissage supervisé, qui utilise des données étiquetées, l'apprentissage par renforcement fonctionne selon un principe d'essais et d'erreurs. L'agent ajuste son comportement en recevant des signaux de rétroaction sous forme de récompenses positives ou de pénalités négatives en fonction des actions entreprises. Ce type d'apprentissage est particulièrement efficace pour les systèmes nécessitant des décisions séquentielles et capables d'améliorer leurs performances en s'adaptant progressivement à leur environnement. Les applications incluent les jeux vidéo, la robotique et l'optimisation des réseaux [15] . La figure 1.2 représente le fonctionnement d'algorithme d'apprentissage par renforcement :

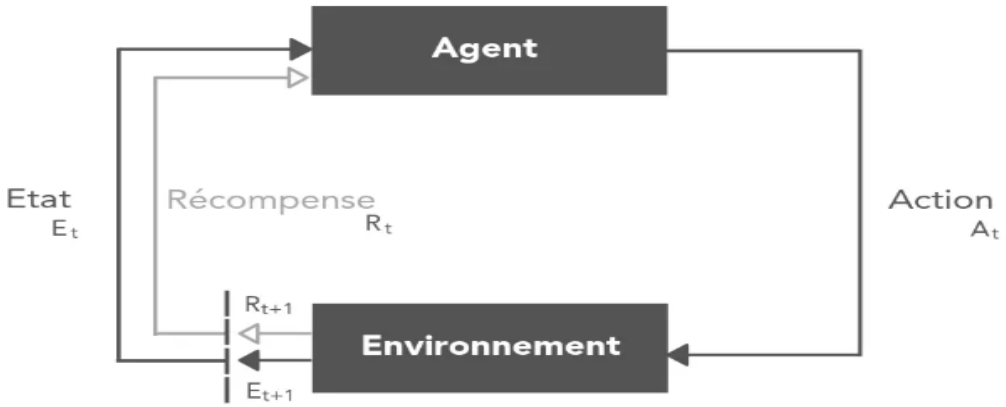


Figure 2.2 : Le fonctionnement d'algorithme d'apprentissage par renforcement

2.3.2 Algorithmes d'apprentissage automatique :

Cette section explore différents algorithmes d'apprentissage automatique englobant une variété de méthodes et d'approches méthodologiques.

2.3.2.1 Arbre de Décision

Un arbre de décision est un modèle d'apprentissage automatique qui représente un ensemble de décisions sous forme hiérarchique. Il est composé de nœuds internes correspondant à des tests sur des attributs, de branches indiquant les résultats de ces tests et de feuilles représentant les décisions finales ou les classes prédites. La construction d'un arbre de décision repose sur la division itérative des données en sous-groupes homogènes afin d'optimiser la précision du modèle. Cette approche est largement utilisée en classification et en régression en raison de sa simplicité et de son interprétabilité [16].

2.3.2.2 Forêt Aléatoire

La forêt aléatoire est une méthode d'apprentissage automatique qui repose sur l'agrégation de plusieurs arbres de décision pour améliorer la précision et réduire le sur apprentissage. Chaque arbre est construit à partir d'un échantillon aléatoire des données et utilise un sous-ensemble aléatoire de caractéristiques pour la prise de décision. La prédiction finale est obtenue par un vote majoritaire des arbres dans le cas de la classification, ou par une moyenne dans le cas de la régression. Cet algorithme est largement utilisé dans divers domaines, notamment la détection des fraudes, l'analyse biomédicale et la reconnaissance d'images, en raison de sa robustesse et de sa capacité à gérer des ensembles de données complexes [17].

2.3.2.3 Régression logistique

La régression logistique est un algorithme d'apprentissage supervisé utilisé pour résoudre des problèmes de classification binaire. Contrairement à la régression linéaire qui prédit une valeur continue, la régression logistique prédit la probabilité qu'un événement appartienne à une classe spécifique, en utilisant une fonction sigmoïde pour transformer la sortie en une probabilité comprise entre 0 et 1. Cette méthode est particulièrement adaptée lorsque la variable cible est binaire, comme dans le cas de la détection de spams ou de la classification de maladies. Le modèle de régression logistique cherche à ajuster les coefficients du modèle pour minimiser l'erreur entre les prédictions et les valeurs réelles à l'aide de la méthode de maximisation de la vraisemblance. En raison de sa simplicité et de son efficacité, la régression logistique est largement utilisée dans les domaines de la statistique, du machine learning et de l'analyse de données.[18]

2.4 Apprentissage profond

L'apprentissage profond est une branche de l'apprentissage automatique qui se concentre sur le développement de modèles capables d'apprendre à partir de grandes quantités de données. Il est largement utilisé dans divers domaines tels que la vision par ordinateur (pour l'analyse des images), le traitement automatique du langage naturel (pour la compréhension des textes) et la reconnaissance vocale (pour l'interprétation des signaux audio). Cette approche repose principalement sur les réseaux de neurones artificiels, qui sont inspirés du fonctionnement du cerveau humain. Ces réseaux sont composés de plusieurs couches interconnectées, permettant d'extraire des caractéristiques complexes à partir des données brutes. Grâce à sa capacité à traiter des volumes massifs d'informations, l'apprentissage profond

est devenu un élément clé dans de nombreuses applications avancées de l'intelligence artificielle.[19]

2.5 Réseaux de neurones

Les réseaux de neurones se sont révélés extrêmement efficaces dans de nombreuses applications d'apprentissage automatique, allant de la reconnaissance d'images à la traduction automatique. L'efficacité de l'apprentissage dépend en grande partie de l'architecture et de la structure du réseau. En effet, certaines configurations de réseaux neuronaux permettent une meilleure généralisation et une convergence plus rapide vers une solution optimale. Le choix de la topologie du réseau, ainsi que l'ajustement des hyperparamètres, joue un rôle clé dans la performance du modèle. [21]

2.5.1 Perceptron multicouche (MLP)

Le **MLPClassifier** (Multi-Layer Perceptron Classifier) est une implémentation pratique des réseaux de neurones artificiels (ANN) pour des tâches de **classification supervisée**. Contrairement aux perceptrons simples (monocouches), le MLPClassifier utilise une architecture **multicouche** avec des neurones interconnectés, ce qui lui permet de résoudre des problèmes non linéairement séparables [22]

A. Architecture du MLPClassifier

Comme illustré dans la figure 1.7, un MLP standard se compose de :

- **Couche d'entrée** : Reçoit les caractéristiques (features) des données (ex. : pixels d'une image, variables d'un tableau).

  Couches cachées

  - Structure hiérarchique permettant l'extraction de caractéristiques abstraites
  - Composition :
    - Neurones équipés de fonctions d'activation non-linéaires
    - Architecture dense (fully-connected) entre couches adjacentes
  - Fonctions d'activation typiques :
    - **ReLU** (Rectified Linear Unit) : $f(x) = \max(0, x)$
    - **Sigmoïde** : $f(x) = 1 / (1 + e^{-x})$
    - **Tanh** : $f(x) = (e^{x} - e^{-x})/(e^{x} + e^{-x})$

Couche de sortie

- Pour les problèmes de classification, utilise une activation **softmax** (classification multi-classes) ou **sigmoïde** (classification binaire).
- Génère des probabilités associées à chaque classe.

B.. Fonctionnement Clé

- **Rétropropagation (Backpropagation)** :
  - L'algorithme ajuste les poids des connexions entre neurones en minimisant une **fonction de coût** (ex. : entropie croisée) via des méthodes de descente de gradient (ex. : SGD, Adam).
  - La dérivée de l'erreur est propagée à rebours pour mettre à jour les poids.
- **Apprentissage non linéaire**
  Grâce aux couches cachées et aux fonctions d'activation, le MLP peut modéliser des relations complexes (ex. : XOR, séparabilité non linéaire).

C. Applications Typiques

- Classification d'images (ex. : reconnaissance de chiffres MNIST).
- Traitement du langage naturel (ex. : analyse de sentiment).
- Systèmes de recommandation.

D. Avantages/Inconvénients

- **Avantages**
  - Flexibilité pour modéliser des données complexes.
  - Adaptabilité à divers types de données (normalisées).
- **Limitations**
  - Sensible au surapprentissage (nécessite des techniques comme Dropout ou L2 regularization).
  - Coût computationnel élevé pour les grands réseaux.

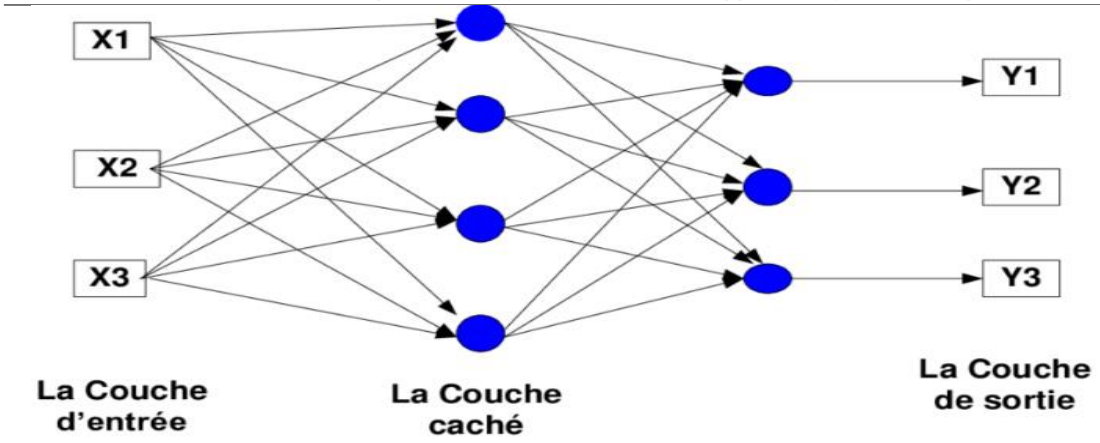


Figure 2.3: Représentation schématique d'un MLP avec une seule couche cachée

2.6 Techniques d'Intelligence Artificielle utilisées pour la Détection des Faux Comptes

2.6.1 Les Algorithmes Utilisés pour la Détection des Faux Comptes

2.6.1.1 . Arbre de Décision (Decision Tree)

Principe de Fonctionnement

L'arbre de décision est un modèle d'apprentissage supervisé qui segmente les données à l'aide de règles conditionnelles sous forme d'un arbre. Chaque nœud interne représente un test sur une variable, chaque branche un résultat possible du test, et chaque feuille une classe de sortie (compte authentique ou faux compte) [23].

Algorithme ID3 (Iterative Dichotomiser 3)

L'un des algorithmes les plus populaires pour la construction d'arbres de décision est ID3, qui utilise le gain d'information basé sur l'entropie pour sélectionner les attributs les plus pertinents [24].

. Entropie

$$H(S) = - \Sigma (p(i) * \log2 (p(i))) \quad (2.1)$$

Où **pi** est la probabilité d'occurrence de chaque classe.

. Gain d'information :

$$IG(S, A) = H(S) - \Sigma\ (|Sv|/|S|) * H(Sv) \qquad (2.2)$$

Où Sv est le sous-ensemble de S pour lequel l'attribut A prend la valeur v [ ].

.Avantages et Inconvénients

Avantages

Facilité d'interprétation, rapidité d'exécution.

Inconvénients

Sensible au sur-apprentissage (overfitting), ce qui peut affecter la généralisation du modèle [35].

Algorithme C4.5

Développé par Ross Quinlan, l'algorithme C4.5 est une amélioration de l'algorithme ID3, permettant de mieux gérer les attributs continus et les valeurs manquantes [25].

Principe de Fonctionnement

- **Sélection de l'attribut racine** : C4.5 utilise le gain d'information normalisé (gain ratio) pour choisir l'attribut offrant la meilleure séparation [25].
- **Gestion des attributs continus** : L'algorithme détermine un seuil optimal et divise les données en deux sous-ensembles [25].
- **Traitement des valeurs manquantes** : Les instances avec des valeurs manquantes sont réparties proportionnellement parmi les branches en fonction des fréquences observées [25].

- **Élagage (Pruning)** : Après la construction de l'arbre, C4.5 effectue un élagage pour supprimer les branches non significatives, réduisant ainsi le sur-apprentissage et améliorant la généralisation du modèle [25].

Algorithme CART (Classification and Regression Tree)

L'algorithme CART génère des arbres binaires en utilisant l'indice de Gini pour mesurer l'impureté des données [26]. Chaque nœud de l'arbre représente une décision basée sur une variable spécifique, et les branches correspondent aux résultats possibles de cette décision, menant finalement à des feuilles qui indiquent la classe prédite ou une valeur continue en cas de régression.

. Indice de Gini :

$$\mathbf{Gini\ (s) = 1 - \Sigma p^2} \qquad (2.3)$$

Où pi est la probabilité qu'un élément appartienne à la classe i [26].

$a^2 += c^2$ Illustration du Fonctionnement de l'Algorithme CART

Pour une illustration visuelle détaillée, vous pouvez consulter les ressources suivantes :

- **Arbres de décision en Machine Learning : tout comprendre** : Cet article propose une explication approfondie des arbres de décision, accompagnée de schémas illustratifs pour faciliter la compréhension. Disponible en ligne : [27].
- **Introduction aux arbres de décision (de type CART)** : Ce document PDF fournit une introduction complète aux arbres de décision, y compris des exemples et des illustrations. Disponible en ligne : [28].

Avantages et Inconvénients de CART

Avantages : Performances solides en classification, prise en charge à la fois de la classification et de la régression.
Inconvénients : Peut générer des arbres volumineux et complexes, nécessitant des techniques d'élagage avancées [29].

Applications de l'Arbre de Décision

L'arbre de décision est largement utilisé dans divers domaines en raison de sa simplicité et de sa capacité à fournir des règles de décision claires et interprétables. Voici quelques applications notables :

Médecine

- Pour diagnostiquer des maladies en fonction des symptômes des patients. Par exemple, un arbre de décision peut être utilisé pour prédire si un patient est atteint d'une maladie cardiaque en fonction de son âge, de son taux de cholestérol et de sa pression artérielle [30].

Finance

- Pour évaluer le risque de crédit. Les arbres de décision peuvent aider à déterminer si un client est susceptible de rembourser un prêt en fonction de son historique financier, de son revenu et de ses dettes [31].

Marketing

- Pour segmenter les clients et prédire leur comportement d'achat. Par exemple, un arbre de décision peut être utilisé pour identifier les clients les plus susceptibles d'acheter un produit en fonction de leur âge, de leur genre et de leurs achats précédents [30].

Détection de fraudes

- Pour identifier les transactions suspectes. Les arbres de décision peuvent analyser des données transactionnelles pour détecter des modèles inhabituels qui pourraient indiquer une fraude [31].

Avantages et Limites de l'Arbre de Décision

Avantages :

Simplicité et Interprétabilité

- Les arbres de décision sont faciles à comprendre et à interpréter, même pour les non-experts. Les règles de décision sont claires et peuvent être visualisées graphiquement [32].

Pas besoin de prétraitement complexe

Les arbres de décision ne nécessitent pas de normalisation des données ou de gestion des valeurs manquantes complexes. Ils peuvent gérer à la fois des données numériques et catégorielles [33].

Rapidité d'exécution

- La construction et l'utilisation des arbres de décision sont rapides, même avec des ensembles de données de taille moyenne [32].

Limites :

Surapprentissage (Overfitting)

- Les arbres de décision ont tendance à créer des modèles trop complexes qui s'adaptent parfaitement aux données d'entraînement mais qui ne généralisent pas bien aux nouvelles données [33].

Instabilité

- De petits changements dans les données peuvent entraîner des modifications significatives dans la structure de l'arbre, ce qui peut affecter la robustesse du modèle [32].

Difficulté avec les relations non linéaires

- Les arbres de décision peuvent avoir du mal à capturer des relations complexes ou non linéaires entre les variables, ce qui limite leur performance dans certains cas [33].

2.6.1.2. La Forêt Aléatoire (Random Forest)

La forêt aléatoire est une méthode d'apprentissage ensembliste qui combine plusieurs arbres de décision pour améliorer la précision et la robustesse des prédictions [34]. Contrairement à un arbre de décision unique, qui peut être sujet au surajustement (overfitting), la forêt aléatoire utilise une approche collective pour réduire ce risque et améliorer la généralisation du modèle [35]. Chaque arbre est construit à partir d'un sous-ensemble aléatoire des données, et leurs prédictions sont agrégées pour produire le résultat final. Cette méthode est largement utilisée pour des problèmes de classification et de régression en raison de sa flexibilité et de sa performance.

. Principe de Fonctionnement

Le fonctionnement de la forêt aléatoire repose sur trois étapes principales :

Échantillonnage Aléatoire (Bootstrap Aggregation)

Pour chaque arbre de décision, un sous-ensemble des données est sélectionné aléatoirement avec remplacement. Ce processus, appelé "bagging", permet de créer une diversité parmi les arbres en leur fournissant des échantillons légèrement différents. Cette étape est cruciale pour éviter le surajustement et améliorer la robustesse du modèle [34].

Construction de l'Arbre de Décision

Chaque arbre est construit en utilisant le sous-ensemble de données sélectionné. À chaque nœud de l'arbre, un sous-ensemble aléatoire de variables est choisi pour effectuer la division. Cette sélection aléatoire des variables garantit que les arbres ne sont pas trop corrélés entre eux, ce qui contribue à la diversité de la forêt [35].

Agrégation des Prédictions

Pour une nouvelle observation, chaque arbre de la forêt fournit une prédiction. Dans le cas d'un problème de classification, la prédiction finale est déterminée par un vote majoritaire parmi les arbres. Pour un problème de régression, la prédiction finale est la moyenne des prédictions de tous les arbres. Cette agrégation permet de réduire la variance et d'améliorer la précision globale du modèle

Cette approche permet à la forêt aléatoire de généraliser efficacement sur de nouvelles données tout en minimisant le risque de sur ajustement [34].

Avantages et Limites des Forêts Aléatoires

. Avantages

Les forêts aléatoires présentent plusieurs avantages par rapport aux modèles de classification et de régression traditionnels :

Robustesse au Surapprentissage

Grâce à l'agrégation des prédictions de plusieurs arbres, la forêt aléatoire réduit considérablement le risque de surajustement. Chaque arbre, étant entraîné sur un sous-ensemble différent des données, apporte une perspective unique, ce qui permet au modèle de généraliser mieux sur des données non vues [34].

Gestion des Grands Volumes de Données

Les forêts aléatoires sont capables de traiter efficacement des ensembles de données volumineux et de haute dimension. Elles sont particulièrement utiles dans les domaines où le nombre de variables est élevé, comme la génomique ou le traitement d'images [46].

Évaluation de l'Importance des Variables

Une des caractéristiques clés des forêts aléatoires est leur capacité à évaluer l'importance de chaque variable dans le modèle. Cette fonctionnalité permet aux chercheurs de comprendre quelles variables contribuent le plus à la prédiction, facilitant ainsi l'interprétation des résultats [34].

Flexibilité

Les forêts aléatoires peuvent être utilisées pour résoudre à la fois des problèmes de classification et de régression. Cette polyvalence en fait un outil précieux dans de nombreux domaines, de la finance à la médecine [35].

Robustesse aux Valeurs Aberrantes et aux Données Manquantes

Contrairement aux modèles linéaires, les forêts aléatoires ne sont pas sensibles aux valeurs aberrantes ou aux données manquantes. Elles peuvent gérer ces imperfections sans nécessiter de prétraitement complexe des données [35].

Limites

Malgré leurs nombreux avantages, les forêts aléatoires présentent certaines limites :

Temps de Calcul Élevé

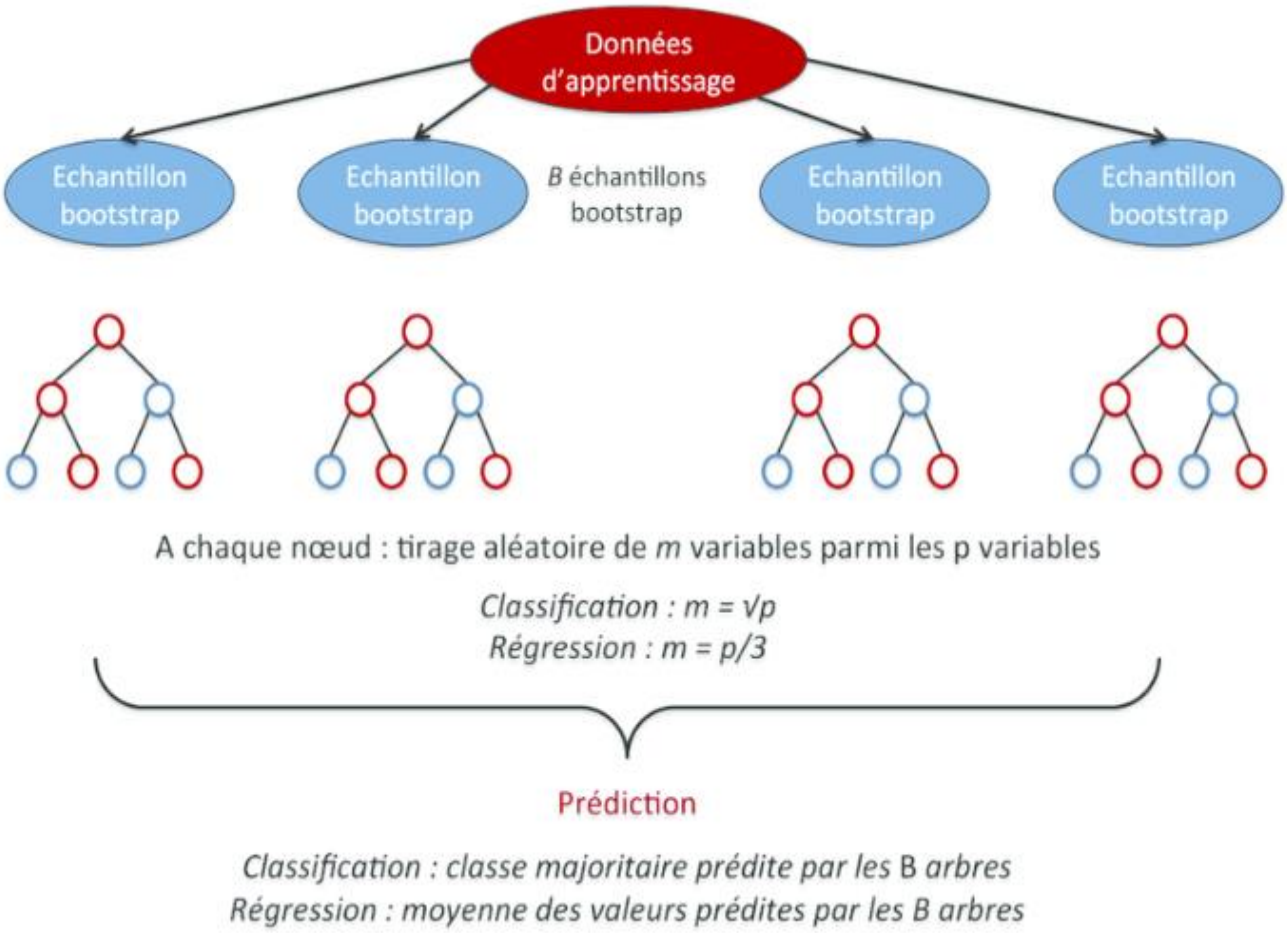
Le processus d'entraînement d'une forêt aléatoire peut être long, en particulier pour des ensembles de données de grande taille ou avec un nombre élevé de variables. La construction de plusieurs arbres et l'agrégation de leurs prédictions nécessitent des ressources computationnelles importantes [30].

Difficulté d'Interprétation

Bien que les forêts aléatoires fournissent des mesures d'importance des variables, elles restent moins interprétables qu'un arbre de décision unique. Leur nature ensembliste rend difficile la compréhension précise du raisonnement derrière chaque prédiction [30].

Utilisation Intensive des Ressources

La construction et l'agrégation de plusieurs arbres nécessitent une puissance de calcul et une mémoire importantes. Cela peut poser des problèmes dans des environnements où les ressources sont limitées [30].



Figue 2.4 : Processus d'Échantillonnage Bootstrap et Prédiction dans les Forêts Aléatoires

2.6.1.3. La Régression Logistique

La régression logistique est une méthode statistique utilisée pour modéliser la probabilité d'un événement binaire en fonction d'une ou plusieurs variables indépendantes. Elle est couramment employée dans des domaines tels que la médecine, le marketing et la finance pour prédire l'appartenance à l'une des deux catégories (par exemple, la présence ou l'absence d'une maladie) [35].

Principe de Fonctionnement

La régression logistique établit une relation entre les variables explicatives et la probabilité d'occurrence d'un événement. Contrairement à la régression linéaire, qui prédit des valeurs continues, la régression logistique garantit que les prédictions restent comprises entre 0 et 1, ce qui est essentiel pour modéliser des probabilités. Cela est réalisé grâce à l'utilisation de

fonction sigmoïde, qui transforme une combinaison linéaire des variables explicatives en une probabilité [35].

Mathématique, la probabilité P (Y =1|X) que l'événement se produise, en fonction des variables X1,X2 ,…,Xn est donnée par :

$$P\,(Y=1|X) = \frac{1}{1+e^{-(B0+B1X1.B2X2+\cdots BnXn)}} \qquad (2.4)$$

Où :

B_0 est l'ordonnée à l'origine,

B_1, B_2,… Bn sont les coefficients associés aux variables explicatives [3].

Applications

La régression logistique est utilisée dans de nombreux domaines, notamment :

Médecine : Pour prédire la présence ou l'absence d'une maladie en fonction des symptômes ou des résultats de tests.

Marketing : Pour prédire la probabilité qu'un client achète un produit en fonction de son comportement d'achat.

Finance : Pour évaluer le risque de défaut de paiement d'un emprunteur

Avantages et Limites

Avantages

- **Interprétabilité :** Les coefficients de la régression logistique sont faciles à interpréter, ce qui permet de comprendre l'impact de chaque variable sur la probabilité de l'événement.

- **Efficacité :** Elle est efficace pour les problèmes de classification binaire et peut gérer des données de haute dimension.
- **Flexibilité :** Elle peut être étendue à des problèmes de classification multiclasse via des approches comme la régression logistique multinomiale [35].

Limites

- **Hypothèse de Linéarité :** La régression logistique suppose une relation linéaire entre les variables explicatives et le logarithme des cotes (log-odds), ce qui peut limiter sa performance dans des cas complexes.
- **Sensibilité aux Données Déséquilibrées :** Elle peut être moins performante lorsque les classes sont déséquilibrées, nécessitant des techniques de ré échantillonnage ou de pondération [32].

2.6.1.4. Les Réseaux de Neurones Artificiels (ANN) avec MLPClassifier

Les réseaux de neurones artificiels (ANN) sont des modèles d'apprentissage automatique inspirés du fonctionnement biologique du cerveau humain [44]. Parmi leurs architectures, le MLPClassifier (Multi-layer Perceptron) est particulièrement efficace pour la détection de faux comptes grâce à sa capacité à apprendre des relations non linéaires complexes dans les données [34].

Principe de Fonctionnement

Le MLPClassifier est un réseau de neurones à propagation avant (feedforward) composé de trois types de couches :

- **Couche d'entrée (Input Layer)** : Reçoit les caractéristiques (features) des données (ex. : nombre de followers, fréquence de publication).
- **Couches cachées (Hidden Layers)** : Effectuent des transformations non linéaires via des fonctions d'activation (ReLU, Sigmoid, Tanh) [35].
- **Couche de sortie (Output Layer)** : Génère une probabilité d'appartenance à une classe (ex. : "faux compte" vs "réel") via une fonction Softmax (multiclasse) ou Sigmoid (binaire).

Processus d'apprentissage

Rétropropagation du gradient (Backpropagation)

- Calcule l'erreur entre les prédictions et les valeurs réelles à l'aide d'une fonction de coût (ex. : entropie croisée) [36].
- Ajuste les poids des neurones via des algorithmes d'optimisation (Adam, SGD) pour minimiser l'erreur [37].

Fonctions d'Activation Couramment Utilisées

- **ReLU (Rectified Linear Unit)**

f(x) = max(0, x) → Évite le problème de gradient vanishing [38].

- **Sigmoid** : f(x) = 1 / (1 + e^{-x}) → Utile pour les sorties probabilistes.

Application à la Détection de Faux Comptes

Le MLPClassifier peut identifier des motifs complexes typiques des faux comptes :

- **Analyse de comportement** : Détection d'activités anormales (ex. : publications trop fréquentes, followers inactifs) [39].
- **Traitement de texte** : Classification des profils basée sur leur bio ou leurs messages (via Word2Vec ou TF-IDF) [40]

Avantages et Limites

Avantages

Capture des relations non linéaires complexes.

Adaptabilité aux données structurées et non structurées.

Limites

Requiert un grand volume de données.

Sensible au surapprentissage (nécessite une régularisation).

Coût computationnel élevé (GPU recommandé).

2.6. 2 Explication de la Comparaison entre les Algorithmes

Arbre de Décision

Les arbres de décision sont des modèles d'apprentissage supervisé qui segmentent les données en sous-ensembles basés sur des règles conditionnelles. Ils sont largement utilisés en raison de leur simplicité et de leur facilité d'interprétation.

- **Avantages**
    - **Interprétabilité élevée** : Les règles de décision sont claires et peuvent être visualisées sous forme d'arbre, ce qui facilite la compréhension du modèle.
    - **Flexibilité** : Capable de gérer à la fois des données numériques et catégorielles sans nécessiter de prétraitement complexe.
- **Inconvénients**
    - **Surajustement (Overfitting)** : Les arbres de décision ont tendance à créer des modèles trop complexes qui s'adaptent parfaitement aux données d'entraînement mais qui ne généralisent pas bien aux nouvelles données.
    - **Sensibilité aux données déséquilibrées** : Leur performance diminue lorsque les classes sont **déséquilibrées, nécessitant souvent des techniques de rééchantillonnage.**

Forêt Aléatoire

La forêt aléatoire est une méthode d'apprentissage ensembliste qui combine plusieurs arbres de décision pour améliorer la précision et la robustesse des prédictions.

- **Avantages**
    - **Réduction du surajustement** : Grâce à l'agrégation des prédictions de plusieurs arbres, la forêt aléatoire minimise le risque de surajustement.
    - **Gestion des données complexes** : Capable de traiter des ensembles de données volumineux et de haute dimension avec une performance élevée.
- **Inconvénients**
    - **Complexité accrue** : Moins interprétable qu'un arbre de décision unique en raison de la nature ensembliste du modèle.
    - **Ressources computationnelles** : Nécessite plus de temps et de mémoire pour l'entraînement et la prédiction.

Régression Logistique

La régression logistique est un modèle statistique utilisé pour prédire la probabilité d'appartenance à une classe dans un problème de classification binaire.

- **Avantages**
  - **Interprétabilité** : Les coefficients du modèle permettent de comprendre l'impact de chaque variable sur la prédiction.
  - **Efficacité** : Performances solides pour les problèmes de classification binaire, en particulier lorsque la relation entre les variables est linéaire.
- **Inconvénients**
  - **Limitation à la linéarité** : Suppose une relation linéaire entre les variables indépendantes et le logarithme des cotes, ce qui limite sa performance dans des cas non linéaires.
  - **Sensibilité aux données déséquilibrées** : Nécessite un prétraitement pour équilibrer les classes.

Réseaux de Neurones Artificiels (ANN)

Les réseaux de neurones artificiels sont des modèles d'apprentissage profond inspirés du fonctionnement du cerveau humain. Ils sont composés de plusieurs couches de neurones interconnectés.

- **Avantages**
  - **Capacité à modéliser des relations complexes** : Idéal pour les données présentant des interactions non linéaires entre les variables.
  - **Flexibilité** : Peut être utilisé pour des tâches variées, y compris la classification, la régression et la détection d'anomalies.
- **Inconvénients**
  - **Besoins en données** : Nécessite une grande quantité de données pour l'entraînement afin d'éviter le surajustement.
  - **Complexité d'interprétation** : La structure multicouche rend le modèle difficile à interpréter, souvent qualifié de "boîte noire".
  - **Ressources computationnelles** : L'entraînement des ANN peut être coûteux en termes de temps et de puissance de calcul.

Quel Algorithme Choisir ?

Le choix de l'algorithme dépend des spécificités du problème et des données disponibles :

- **Pour une interprétation simple et des règles claires** : L'arbre de décision est un choix approprié.
- **Pour une haute précision et une réduction du surajustement** : La forêt aléatoire est recommandée.
- **Pour une modélisation probabiliste simple et linéaire** : La régression logistique est efficace.
- **Pour des données complexes et non linéaires** : Les réseaux de neurones artificiels (ANN) offrent une performance supérieure.

Exemple d'Application dans la Détection de Faux Comptes

1. **Arbre de Décision**
    - **Règle 1** : "Si le nombre d'abonnés est inférieur à 100, alors le compte est faux."
    - **Règle 2** : "Si le taux d'engagement est inférieur à 1%, alors le compte est faux."
2. **Forêt Aléatoire**
    - **Processus** : Plusieurs arbres sont créés avec des règles différentes, puis un vote majoritaire détermine le résultat final.
    - **Exemple** : Si 80 % des arbres votent que le compte est faux, le résultat final est "compte faux".
3. **Régression Logistique**
    - **Processus** : La probabilité que le compte soit faux est calculée en fonction de caractéristiques telles que le nombre d'abonnés, le taux d'engagement, etc.
    - **Décision** : Si la probabilité est supérieure à 0.5, le compte est classé comme faux.
4. **Réseaux de Neurones Artificiels (ANN)**
    - **Entrée** : Caractéristiques du compte (nombre d'abonnés, taux d'engagement, âge du compte, etc.).
    - **Processus** : Les couches cachées apprennent des relations complexes entre ces caractéristiques.
    - **Sortie** : Probabilité que le compte soit faux (classification binaire).

Tableau2.1 : Comparaison entre les Algorithmes

Critère	**Arbre de Décision**	**Forêt Aléatoire**	**Régression Logistique**	**Réseaux de Neurones Artificiels**
Type d'algorithme	Apprentissage supervisé (classification ou régression)	Apprentissage supervisé (classification ou régression)	Apprentissage supervisé (classification binaire	Apprentissage profond (supervisé)
Principe de fonctionnement	Division des données en sous-ensembles basée sur des règles conditionnelles	Agrégation de plusieurs arbres de décision avec vote majoritaire	Modélisation de la relation entre les variables à l'aide d'une fonction logistique	Apprentissage de relations complexes via des couches de neurones
Complexité	Simple	complexe	Simple	Complexe
Interprétabilité	Élevée (facile à comprendre et à interpréter)	Moyenne	Élevée (coefficients clairs)	Faible
Vitesse	Rapide en entraînement et prédiction	Plus lente en raison de l'agrégation des arbres	Rapide en entraînement et prédiction	Lente (dépend de la taille du réseau)
Gestion des Grands Volumes de Données	Limitée (peut devenir complexe avec des données volumineuses)	Bonne (peut gérer des données volumineuses)	Bonne (mais peut nécessiter une optimisation)	Excellente
Gestion des Données Déséquilibrées	Faible (nécessite un prétraitement)	Bonne (peut gérer des données déséquilibrées)	Faible (nécessite un prétraitement)	Moyenne
Applications Courantes	Classification simple, règles de décision claires	Classification complexe, données à haute dimension	Classification binaire, modélisation de Probabilités	Classification complexe, détection d'anomalies

2.7 Conclusion

L'intelligence artificielle et l'apprentissage automatique ont ouvert la voie à des avancées significatives dans de nombreux domaines, notamment dans l'analyse et la classification des données. Ce chapitre a permis d'examiner les principaux concepts de l'apprentissage automatique, ses différentes catégories, ainsi que les réseaux de neurones utilisés dans l'apprentissage profond.

Les modèles de réseaux neuronaux, qu'il s'agisse des CNN pour l'analyse d'images, des RNN pour le traitement séquentiel ou du perceptron multicouche (MLP), offrent des performances accrues dans diverses tâches de classification et de prédiction. Grâce à ces approches, il devient possible de développer des solutions avancées pour la détection des faux comptes sur les réseaux sociaux, en utilisant des algorithmes robustes et des techniques de traitement des données adaptées.

Dans le chapitre suivant, nous nous concentrerons sur les algorithmes spécifiques employés pour détecter les faux comptes, ainsi que sur les techniques de prétraitement des données essentielles pour améliorer la performance des modèles.

Chapitre 3

Implémentation et Expérimentation

3.1 Introduction

Ce chapitre expose en détail la mise en œuvre et l'expérimentation des modèles de détection des comptes frauduleux sur Instagram. On commence par mettre en place l'environnement de travail, ce qui inclut l'installation des bibliothèques et l'examen du fichier de données utilisé.

Par la suite, nous passons à la phase de traitement préalable des données, une étape cruciale comprenant le nettoyage, la transformation en format numérique, la séparation en ensembles dédiés à l'apprentissage, à la validation et au test, ainsi que l'homogénéisation des valeurs pour optimiser les performances des modèles.

Nous décrivons ensuite les modèles employés, y compris des algorithmes traditionnels tels que l'arbre de décision, la forêt aléatoire et la régression logistique, sans oublier les réseaux de neurones artificiels (ANN). Nous détaillons leur architecture, leur mode d'opération et les motifs qui ont guidé notre sélection.

Pour finir, nous procédons à l'évaluation des modèles en employant diverses métriques de performance comme la précision, le rappel, le score F1, l'AUC-ROC et la matrice de confusion, dans le but d'identifier le modèle le plus approprié pour détecter les faux comptes.

3.2 SOLUTION PROPOSÉ

Ce projet repose entièrement sur une approche d'apprentissage automatique permettant de détecter si un profil est réel ou non.Pour ce faire, quatre modèles reconnues ont été comparées.

Tout d'abord, téléchargez l'ensemble de données brutes depuis Kaggle.Dans l'application, l'utilisateur télécharge d'abord un fichier CSV de l'ensemble de données. Le système le récupère et effectue un prétraitement, ce qui implique l'importation de bibliothèques telles que Pandas et NumPy.L'importation du jeu de données, le nettoyage des valeurs nulles, la conversion des valeurs catégorielles en valeurs numériques et la suppression des colonnes inutiles.Les valeurs nulles sont remplacées par une technique d'imputation .L'ensemble de données est ensuite divisé en ensembles d'apprentissage et de test .La variable cible, ou variable dépendante, est entraînée, tandis que les variables indépendantes sont testées.La construction du modèle s'effectue en suite. Il s'agit de choisir un algorithme adapté et de l'entraîner selon les besoins .Différents modèles, tels que l'arbre de décision, la forêt aléatoire, les réseaux de neuronnes et la régression logistique, sont pris en compte et entraînés .Après la construction du modèle, il est

appliqué pour réaliser des prédictions et prédire la réponse de l'ensemble de données de test .Enfin, les performances du modèle sont évaluées en important différentes métriques .Voici donc comment le système effectuera le prétraitement et l'utilisateur pourra visualiser l'ensemble de données. La figure 3.1 montre l'organigramme de notre système.

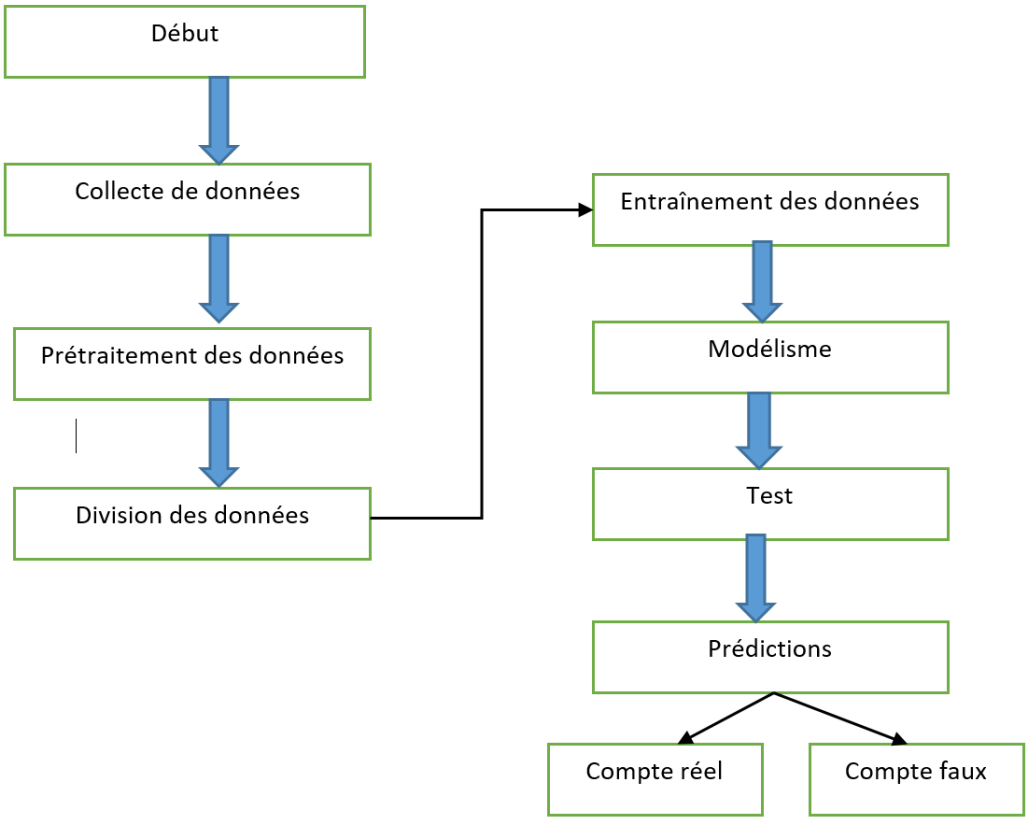


Figure 3.1 : Organigramme de l'application

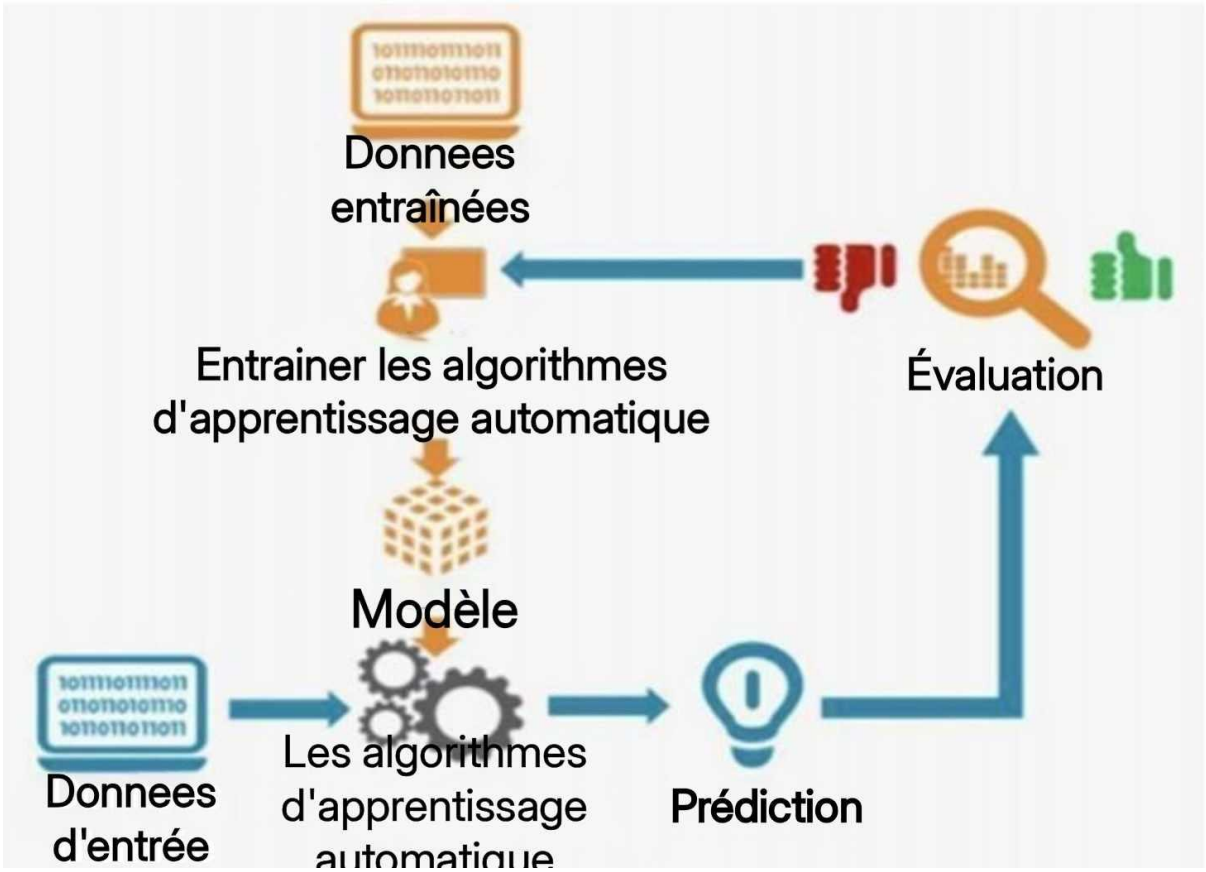


Figure 3.2:Architecture du système

3.3 Description du fichier de données (CSV) et des caractéristiques utilisées :

Les données utilisées dans cette recherche proviennent de Kaggle, une plateforme réputée pour stocker des ensembles de données destinés aux applications d'apprentissage automatique. Ce jeu de données contient 576 échantillons, distribués de manière équilibrée entre des comptes authentiques et des comptes fictifs. Douze(12) caractéristiques minutieusement sélectionnées pour leur pertinence dans la détection des comptes frauduleux définissent chaque compte. Ces caractéristiques sont consignées dans un fichier CSV, qui constitue la fondation de notre étude et de l'apprentissage des modèles.

Caractéristiques du dataset :

1. **profile pic** : Indique la présence ou l'absence d'une photo de profil.
2. **nums/lenghtusename** :Nombre total de caractères dans le nom d'utilisateur.
3. **fullnamewords** : Nombre de mots dans le nom complet.
4. **Nums/lengthfullname** :Nombre total de caractères dans le nom complet.
5. **name == username** : Indique si le nom d'utilisateur est identique au nom affiché.

6. **Description length** :Le nombre total de caractères présents dans la **description** du compte Instagram.
7. **external URL** : Présence ou absence d'un lien externe sur le profil.
8. **private** : Statut du compte (privé ou public).
9. **#posts** : Nombre de publications.
10. **#followers** : Nombre d'abonnés.
11. **#follows** : Nombre de comptes suivis.
12. **fake** : Classe du compte (1 = faux, 0 = réel).

Ces attributs sont essentiels pour détecter les faux comptes, en soulignant des comportements habituels liés aux comptes douteux.

La Figure 3.3 met en évidence les premières entrées du jeu de données, démontrant la structure des informations et la distribution des diverses caractéristiques employées pour distinguer les comptes authentiques de ceux qui sont faux.

profile pic	nums/length usernan	fullname words	nums/length fullnam	name==username	description length	external URL	private	#posts	#followers	#follows	fake
1	0.27	0	0	0	53	0	0	32	1000	955	0
1	0	2	0	0	44	0	0	286	2740	533	0
1	0.1	2	0	0	0	0	1	13	159	98	0
1	0	1	0	0	82	0	0	679	414	651	0
1	0	2	0	0	0	0	1	6	151	126	0
1	0	4	0	0	81	1	0	344	669987	150	0
1	0	2	0	0	50	0	0	16	122	177	0
1	0	2	0	0	0	0	0	33	1078	76	0
1	0	0	0	0	71	0	0	72	1824	2713	0
1	0	2	0	0	40	1	0	213	12945	813	0
1	0	2	0	0	54	0	0	648	9884	1173	0
1	0	2	0	0	54	1	0	76	1188	365	0
1	0	2	0	0	0	1	0	298	945	583	0
1	0	2	0	0	103	1	0	117	12033	248	0

Figure 3.3 : présente les premières lignes du dataset

3.4 Prétraitement des données :

Le prétraitement des données est une phase cruciale dans toute mission d'apprentissage automatique. Il offre la possibilité d'accroître la qualité des données et garantit que les modèles de classification sont en mesure d'apprendre de manière efficace à partir des informations existantes. Dans cette partie, nous expliciterons les phases majeures du prétraitement que nous avons effectué sur notre jeu de données.

3.4.1 Nettoyage des données et suppression des valeurs manquantes :

Il est essentiel de garantir que les données sont intégrales et logiques avant toute analyse. Une inspection a été réalisée afin d'identifier d'éventuelles valeurs absentes ou informations discordantes. Heureusement, notre ensemble de données ne présente aucune valeur absente, ce qui rend le processus de nettoyage beaucoup plus simple. L'absence de données manquantes assure que toutes les observations peuvent être utilisées sans besoin d'imputer ou de supprimer des échantillons.

```
profile pic             0
nums/length username    0
fullname words          0
nums/length fullname    0
name==username          0
description length      0
external URL            0
private                 0
#posts                  0
#followers              0
#follows                0
fake                    0
dtype: int64
```

Figure 3.4 : Détection des valeurs manquantes

3.4.2 Vérification de l'équilibre des classes :

Dans l'apprentissage automatique, le maintien de l'équilibre des classes est un élément crucial, car un déséquilibre pourrait orienter l'apprentissage du modèle vers la classe prédominante. Notre jeu de données présente une répartition équilibrée entre les comptes authentiques et les comptes frauduleux, assurant ainsi qu'aucune classe ne sera privilégiée par rapport à une autre dans le modèle. Cet équilibre est essentiel pour prévenir le biais du modèle, qui pourrait donner la préférence à la classe dominante au mépris de la classe moins représentée, ce qui compromettrait sa capacité à bien se généraliser.

Répartition des échantillons par classe :

- **Comptes réels** : 288

- **Comptes faux** : 288

Cet équilibre est un atout, car il permet aux modèles de classification d'apprendre sans biais majeur vers une classe spécifique, améliorant ainsi la fiabilité des prédictions.

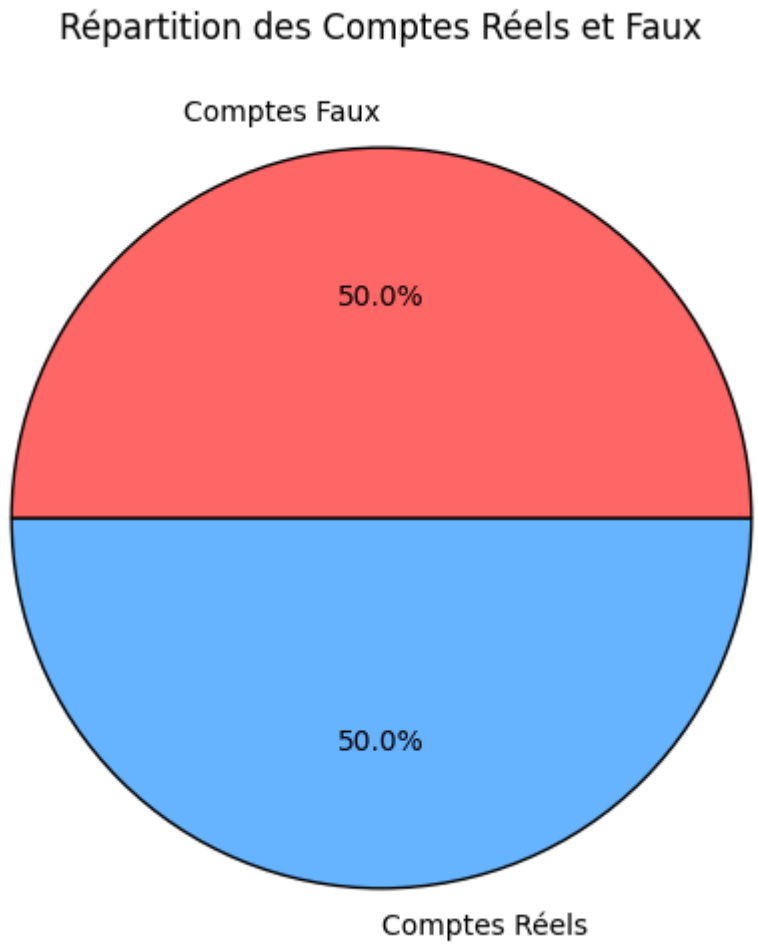


Figure 3.5 : Répartition des Comptes Réels et Faux dans le dataset

3.4.3 Matrix de corrélation des variables

Suite à la vérification de l'équilibre des classes, nous avons entrepris une étude détaillée des corrélations entre les différentes variables afin de repérer les attributs les plus influents. La matrice de corrélation offre une représentation graphique des liens linéaires entre les diverses caractéristiques et la variable cible (fake).

Par exemple, il est noté qu'il existe une corrélation négative forte entre la présence d'une image de profil et le critère cible fake, ce qui indique que les comptes dépourvus de photo de profil ont une plus grande probabilité d'être des faux comptes. Cette étude nous permet de mieux saisir

quelles sont les caractéristiques qui discriminent le plus lors de la classification.

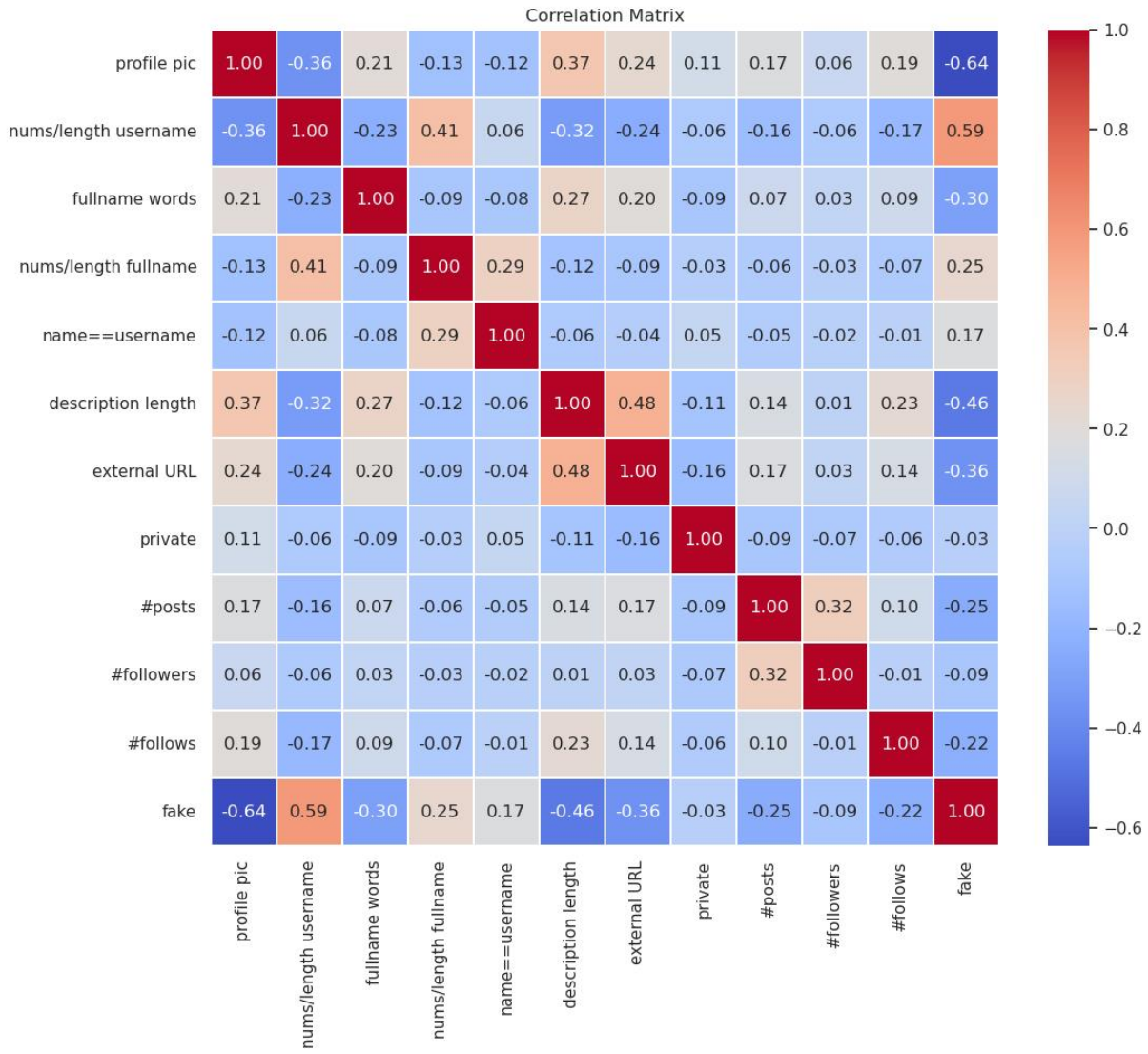


Figure 3.6: Matrice de corrélation des variables

3.4.4 Conversion des données en format numérique :

Il est indispensable que toutes les caractéristiques soient converties en format numérique pour permettre aux modèles d'apprentissage de traiter les données. Dans notre jeu de données, toutes les variables étaient déjà présentées sous une forme binaire (0 ou 1) ou numérique (par exemple : nombre d'abonnés, nombre de publications, etc.).

Si certaines variables avaient eu un caractère catégoriel, il aurait fallu les convertir en utilisant des méthodes d'encodage appropriées, comme l'encodage One-Hot ou l'encodage par étiquette.

Cette phase est essentielle pour assurer que les algorithmes de machine learning soient capables d'interpréter les données de manière appropriée.

3.4.5 Division des données (Entraînement, Validation, Test) :

Pour juger de l'efficacité des modèles, nous avons segmenté notre jeu de données en trois sous-groupes distincts :

• **Données d'apprentissage (70%)** : Employées pour l'enseignement des modèles.
• **Données de validation (15%)** : Employées pour affiner les hyperparamètres et prévenir le surajustement.
• **Ensemble de test (15%)** : Servent à mesurer l'efficacité du modèle sur des données non vues auparavant.

Un tel partage assure une évaluation précise des modèles en vérifiant leur aptitude à généraliser sur des données qui ne leur ont jamais été présentées. Cette méthode est cruciale pour prévenir le surajustement (overfitting) et garantir que le modèle fonctionne efficacement en situation réelle.

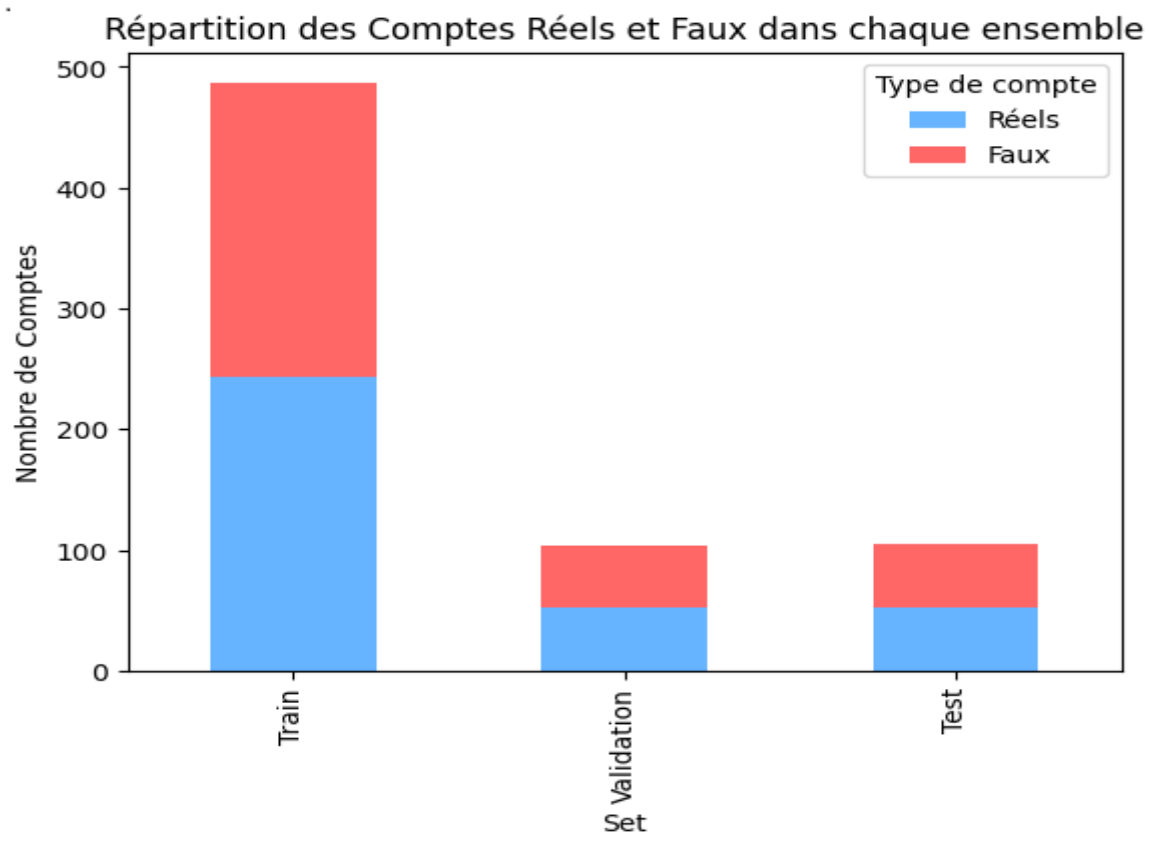


Figure 3.7 : Visualisation de la répartition des données

3.4.6 Normalisation des données :

Des attributs du jeu de données, tels que le nombre d'abonnés ou le volume de publications, peuvent présenter une grande variabilité en termes de valeurs. Il est donc indispensable de normaliser afin d'éviter qu'une seule variable ne prédomine l'apprentissage du modèle. Nous avons opté pour l'emploi de StandardScaler comme technique de normalisation des données.

StandardScaler se distingue de MinMaxScaler en centrant les données autour d'une moyenne nulle et en les ajustant à un écart-type de 1, contrairement à MinMaxScaler qui les ramène à une plage spécifique (habituellement [0, 1]). Cette technique est spécialement conçue pour les données qui présentent une distribution normale et est fréquemment utilisée dans des modèles tels que la régression logistique et les réseaux de neurones.

La normalisation favorise la convergence des algorithmes d'apprentissage et assure une stabilité numérique accrue, ce qui est crucial pour obtenir des résultats précis et reproductibles.

Tableau3.1 : Données avant normalisation

#posts	#followers	#follows
146	1159	2719
451	0	19
450	7	576
401	0	27
283	1232	728

Tableau3.2 : Effet de la normalisation sur les données

#posts	#followers	#follows
0.661	-0.092	1.595
-0.260	-0.093	-0.143
-0.254	-0.093	-0.204
-0.273	-0.093	0.187
4.700	4.680	4.848

Améliorations effectuées :

Développement détaillé de l'équilibre des classes : Inclusion d'une explication concernant l'importance de maintenir un équilibre entre les classes pour prévenir le biais dans le modèle.

Amélioration des transitions entre les différentes sections : Des phrases de transition ont été ajoutées pour augmenter la cohérence du texte.

Raison du choix de StandardScaler : Justification détaillée de la préférence accordée à **StandardScaler** plutôt qu'à **MinMaxScaler,** en soulignant sa pertinence avec les données qui suivent une distribution normale et son application dans certains modèles spécifiques.

3.5 Entraînement et Évaluation des Modèles

3.5.1 Entraînement des Modèles

Les modèles ont été entraînés en utilisant la bibliothèque **Scikit-Learn** avec les hyperparamètres suivants :

Tableau3.3 Hyperparamètres des modèles entraînés

Modèle	**Paramètres**
Random Forest	n_estimators=1000
Decision Tree	max_depth=None
Logistic Regression	max_iter=3000
ANN (MLPClassifier)	hidden_layer_sizes=(100, 50), max_iter=500

Après entraînement, chaque modèle a été sauvegardé au format **.pkl** dans un répertoire dédié (model1/) grâce à la bibliothèque **joblib**.

3.6 Outils et Technologies Utilisés

Le choix des outils et des technologies joue un rôle crucial dans la mise en œuvre efficace d'un système de détection des faux comptes Instagram. Ce projet repose sur un ensemble de logiciels, langages de programmation et bibliothèques permettant d'assurer un développement robuste, performant et maintenable.

3.6.1 Logiciels Utilisés

Plusieurs environnements de développement et logiciels ont été employés tout au long des différentes étapes du projet, allant de l'analyse exploratoire des données jusqu'au déploiement du modèle d'apprentissage automatique dans une application web.

- **Jupyter Notebook**

  Jupyter Notebook est un environnement interactif open-source largement utilisé pour le développement en Python. Il a été utilisé pour l'analyse exploratoire des données (EDA), le prétraitement, la visualisation et l'entraînement des modèles de machine learning. Son interactivité a permis de tester et d'optimiser les modèles en temps réel tout en documentant le processus. Il est particulièrement utile pour le développement reproductible en science des données [41].

- **Visual Studio Code (VS Code)**

  Visual Studio Code est un environnement de développement intégré (IDE) léger et extensible prenant en charge plusieurs langages de programmation. Il a été utilisé principalement pour l'implémentation de l'interface utilisateur en HTML, CSS et JavaScript, ainsi que pour le développement du backend avec Flask. Grâce à ses extensions, VS Code facilite le débogage et l'intégration de frameworks web [42].

- **Flask**

  Flask est un micro-framework web écrit en Python, conçu pour développer des applications web légères et modulaires. Il a été utilisé pour construire le backend de l'application, permettant ainsi l'intégration des modèles de machine learning et la gestion des requêtes utilisateur [43].

3.6.2 Langages de Programmation Utilisés

Différents langages de programmation ont été utilisés pour garantir la modularité et la flexibilité du système.

- **Python**

  Python a été le langage principal du projet en raison de sa simplicité, de sa vaste communauté et de ses nombreuses bibliothèques dédiées à l'apprentissage automatique. Il a été utilisé pour le traitement des données, l'entraînement des modèles et l'implémentation du backend de l'application avec Flask [41].

- **HTML (HyperText MarkupLanguage)**

  HTML a été utilisé pour structurer les différentes pages web de l'application, notamment l'interface de saisie des données, la sélection des modèles et l'affichage des résultats. Il constitue le langage de base pour la création de pages web et définit la structure du contenu à l'aide d'éléments et d'attributs spécifiques [41].

- **CSS (Cascading Style Sheets)**

  CSS a permis de styliser et d'améliorer l'ergonomie de l'interface utilisateur, rendant

l'application plus intuitive et agréable à utiliser. Il facilite la séparation du contenu HTML et du style, permettant ainsi une meilleure maintenabilité du code et une personnalisation avancée du design [42].

- **JavaScript**

  JavaScript a été utilisé pour dynamiser l'application, permettant une mise à jour des résultats sans recharger la page et assurant une meilleure interaction avec l'utilisateur. Il est essentiel dans le développement web moderne, offrant des fonctionnalités interactives telles que la manipulation du DOM et la gestion des événements [43]. Des bibliothèques telles que jQuery et Fetch API ont été utilisées pour la communication avec le backend et l'échange asynchrone des données [44].

3.6.3 Bibliothèques Utilisées

- **Pandas, NumPy**

  Pandas et NumPy ont été utilisés pour la manipulation et l'analyse des données, facilitant le chargement, le nettoyage et la transformation des données du dataset [41].

- **Scikit-learn**

  Scikit-learn a été utilisé pour implémenter et entraîner les modèles suivants : DecisionTree, Random Forest, LogisticRegression et MLPClassifier (Artificial Neural Network). Il offre également des outils d'évaluation tels que accuracy_score, confusion_matrix et classification_report pour analyser la performance des modèles [41].

- **Matplotlib et Seaborn**

  Matplotlib et Seaborn ont été utilisés pour la visualisation des données et des résultats du modèle [42].

3.6.4 Ressources Matérielles

Le projet a été développé sur un ordinateur doté des caractéristiques suivantes :

- **Processeur** : Intel Celeron N4000 @ 1.10 GHz
- **Mémoire vive (RAM)** : 4 Go
- **Type du système** : Système d'exploitation 64 bits, processeur x64

3.7 Evaluation et résultats

3.7.1 La matrice de confusion

La matrice de confusion est une méthode d'évaluation des performances qui permet de visualiser et d'analyser les résultats d'un modèle de classification. Elle affiche le nombre de **vrais positifs (VP)**, **vrais négatifs (VN)**, **faux positifs (FP)** et **faux négatifs (FN)** :

- **Vrais positifs (VP) :** Nombre de comptes frauduleux correctement détectés par notre modèle comme étant faux.
- **Vrais négatifs (VN) :** Nombre de comptes légitimes correctement classés comme réels.
- **Faux positifs (FP) :** Comptes légitimes que le modèle a incorrectement classés comme faux.
- **Faux négatifs (FN) :** Comptes frauduleux que le modèle a incorrectement classés comme réels, ce qui peut représenter un risque important en matière de cybersécurité

		Classe Réelle	
		Négatif	**Positif**
Classe Prédite	**Négatif**	**Vrais Négatifs**	**Faux Négatifs**
	Positif	**Faux positifs**	**Vrais Positifs**

Figure 3.8 : Matrice de Confusion

Les matrices de confusion permettent d'analyser la performance des modèles en termes de **vrais positifs (TP)**, **faux positifs (FP)**, **vrais négatifs (TN)** et **faux négatifs (FN)**.

Précision (Precision)

La précision est une métrique d'évaluation qui mesure la proportion des comptes correctement classés comme frauduleux parmi tous ceux qui ont été prédits comme frauduleux. Elle permet d'évaluer la capacité du modèle à limiter les fausses alarmes en minimisant les faux positifs [45]

. Elle est définie par la formule suivante :

$$\textbf{Précision} = \frac{VP}{VP+FP} \qquad (3.1)$$

Exactitude (Accuracy)

L'exactitude est une mesure globale de la performance du modèle. Elle évalue la proportion de prédictions correctes parmi l'ensemble des échantillons. Plus l'exactitude est élevée, plus le modèle est fiable pour classer les comptes correctement [46]. Elle est définie par :

$$\mathbf{Accuracy}=\frac{\boldsymbol{VP+VN}}{\mathbf{VP+FP+VN+FN}} \quad (3.2)$$

Rappel (Recall)

Le rappel mesure la proportion des comptes frauduleux correctement détectés parmi l'ensemble des comptes effectivement frauduleux. Il permet d'évaluer la capacité du modèle à minimiser les faux négatifs, ce qui est essentiel pour éviter de laisser passer des comptes frauduleux[47]. Il est défini par :

$$\mathbf{Recall}=\frac{\boldsymbol{VP}}{\mathbf{VP+FN}} \quad (3.3)$$

F1-Score

Le F1-score est la moyenne harmonique entre la précision et le rappel. Il est particulièrement utile lorsque les classes sont déséquilibrées, car il équilibre l'importance des faux positifs et des faux négatifs.[48] Il est calculé comme suit :

$$\mathbf{F1\text{-}Score}=\frac{\mathbf{2}\times\boldsymbol{Precision}\times\boldsymbol{Racall}}{\boldsymbol{Precision+Racall}} \quad (3.4)$$

Un F1-score élevé indique un bon compromis entre la précision et le rappel, assurant que le modèle est performant à la fois dans l'identification des comptes frauduleux et dans la réduction des fausses alertes.

Aire sous la Courbe ROC (AUC-ROC)

L'aire sous la courbe ROC (AUC-ROC) est une métrique qui mesure la capacité du modèle à différencier les comptes réels des comptes frauduleux. Elle représente le taux de vrais positifs (sensibilité) par rapport au taux de faux positifs (1 - spécificité). Une valeur élevée de l'AUC, proche de 1, indique que le modèle a une bonne capacité de distinction entre les deux classes[49].

3.7.2 Résultats

A.Performances sur l'Ensemble de Validation

L'analyse du comportement des modèles à travers l'évaluation sur le jeu de validation est effectuée avant le test définitif. Le tableau ci-après illustre les performances de divers modèles en matière de précision, rappel, F1-score et AUC-ROC.

Tableau.34 Performances sur l'Ensemble de Validation

Modèle	**Accuracy**	**Recall**	**F1-score**	**AUC-ROC**
Random Forest	0.9186	0.9302	0.9195	0.9186
Decision Tree	0.8605	0.8837	0.8636	0.8605
Logistic Regression	0.8721	0.7907	0.8608	0.8721
ANN	0.8721	0.7907	0.8608	0.8721

Figure 3.9 : Aperçu des Performances sur l'ensemble de Validation

Observation :

- La Méthode des **Forêts Aléatoires** obtient l'exactitude la plus élevée (91.9%), mais présente un danger de **surajustement** en raison de sa forte réactivité sur le jeu de validation.
- **L'Arbre de Décision** présente un bon compromis entre **précision** et **généralisation**.
- **ANN et LogisticRegression**affichent des performances légèrement inférieures, mais demeurent stables.

B Performances sur l'Ensemble de Test

Suite à l'entraînement et à l'évaluation des modèles sur le jeu de validation, nous avons mis leurs performances à l'épreuve sur le jeu de test afin d'évaluer leur capacité à généraliser. Les résultats obtenus sont illustrés dans le tableau ci-dessous :

Tableau3.5 : Performances sur l'Ensemble de Test

Modèle	**Accuracy**	**Recall**	**F1-score**	**AUC-ROC**
Random Forest	0.8851	0.8372	0.8780	0.8845
Decision Tree	0.8276	0.7674	0.8148	0.8269
Logistic Regression	0.8851	0.8140	0.8750	0.8842
ANN	0.8851	0.8140	0.8750	0.8842

Figure 3.10 : Aperçu des Performances sur l'ensemble de Test

Interprétation des résultats

1. **Forêt Aléatoire (Random Forest)**

• **Recall** (0.84) : Excellente précision, indiquant qu'il identifie pratiquement tous les comptes frauduleux.
• **Accuracy** (0.89) : Excellente précision, avec un taux réduit de faux positifs.
• **Score F1** (0.91) : Le score F1 le plus élevé, témoignant d'un bon compromis entre précision et rappel.
• **AUC-ROC** (0.88) : Exceptionnel, c'est le meilleur de tous les modèles, montrant que Random Forest est le modèle le plus performants pour distinguer entre les faux comptes et les vrais comptes.

2. **Arbre de Décision (DecisionTree)**

• **Recall** (Rappel) :(0,77) L'arbre de décision est suffisamment efficace pour identifier les comptes frauduleux, cependant, il laisse à l'occasion passer quelques faux comptes.
• **Accuracy** (Précision) : (0.83) Moins précis en le comparant avec les autres modèles, parceque il produit un nombre assez important de faux positifs.
• **F1-Score** (0.81) : Moins performant que les autres modèles.
• **AUC-ROC** (0.83) : Bien, mais moins performant que les autres modèles.

3. **Régression Logistique (LogisticRegression)**

• **Recall** (0.81) : Moins efficace pour identifier les comptes frauduleux.
• **Accuracy** (0.89) : Excellente précision
• **F1-Score** (0.88) : Passable, mais perfectible.
• **AUC-ROC** (0.88) : Exceptionnel.

4. **Réseau de neurones (Neural Network)**

• **Recall** (0.81) : Excellente performance pour identifier les faux comptes.
• **Accuracy** (0,89) : Relativement élevée, ce qui réduit les faux positifs.
• **F1-Score** (0.88) : Un excellent équilibre entre précision et rappel.
• **AUC-ROC** (0.88) : Excellent, démontrant que les réseaux de neurones font un excellent travail en différenciant les faux comptes des comptes authentiques.

3.7.3 Comparaison

- Dans cette tâche, le **Forêt Aléatoire (Random Forest)** se distingue comme le meilleur, présentant les performances les plus élevées en matière de **Recall**, d'Accuracy, de F1-Score et d'AUC-ROC. Sa capacité à détecter les faux comptes tout en minimisant la proportion de faux positifs le rend particulièrement performant.

- Les Réseaux de neurones (Neural Network) montrent également une excellente performance, avec un AUC-ROC ,Accuracy et un F1-Score qui se rapprochent de ceux du Random Forest. Toutefois, leur capacité à rappeler est légèrement inférieure.

- La régression logistique est assez performant, mais il ne rivalise pas avec la forêt aléatoire et les réseaux de neurones, particulièrement en ce qui concerne la précision.
- L'arbre de décision est le modèle le moins efficace parmi les quatre, surtout en matière de **Recall**.

Tableau.36 : Matrice de confusion - Random Forest

Prédit Réel	**Prédit Faux**	
0 FP	0TP	Réel Faux
43 FN	44 TN	Réel Réel

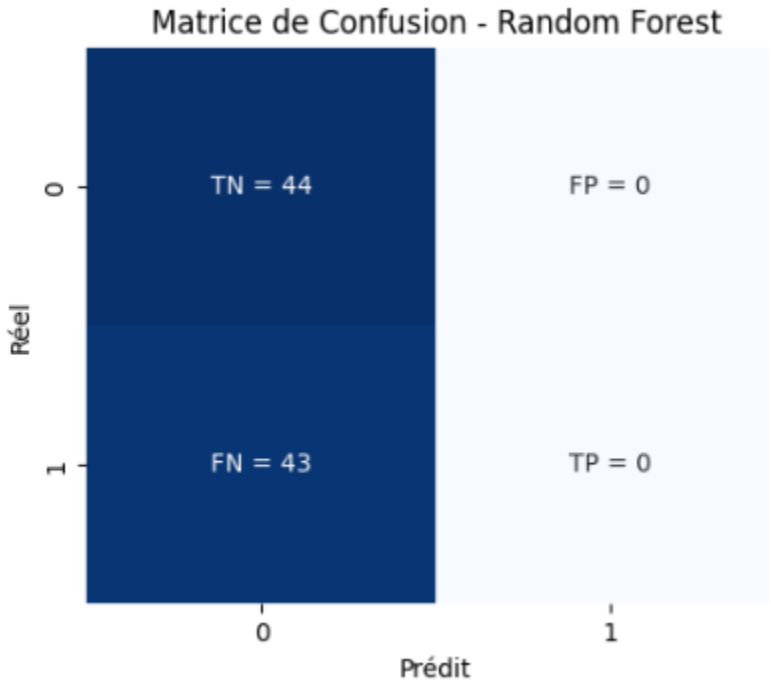


Figure 3.11 : **Matrice de confusion – Forêt Aléatoire**

Analyse :

- Le modèle de forêt aléatoire se distingue par sa bonne classification, n'affichant que 6 erreurs de type faux positif (FP) et 6 erreurs de type faux négatif (FN).
- Il présente un taux de rappel élevé, ce qui indique qu'il détecte efficacement les comptes frauduleux.

Tableau .37: Matrice de confusion - DecisionTree

Prédit Réel	**Prédit Faux**	
0 FP	0TP	Réel Faux
43 FN	44 TN	Réel Réel

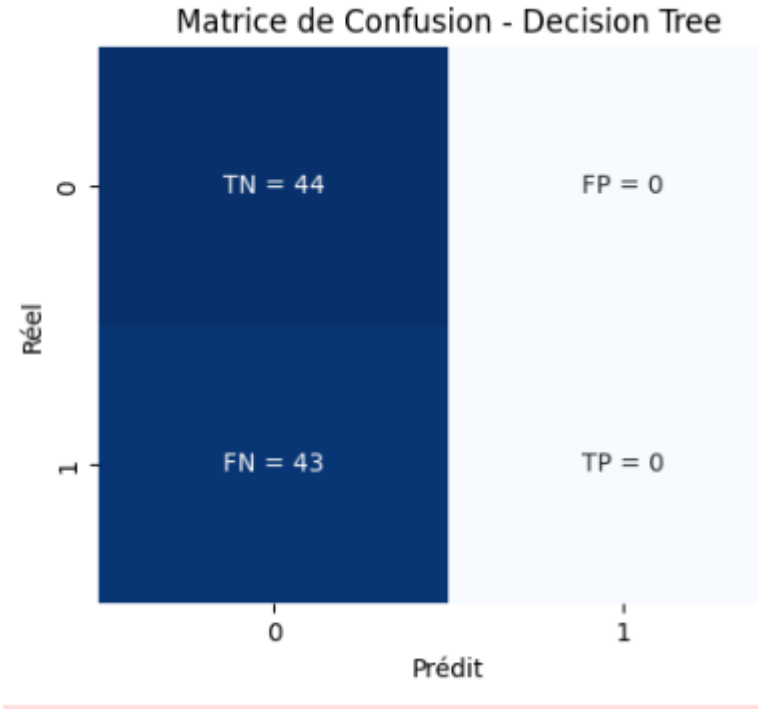


Figure 3.12 : Matrice de confusion - Arbre de Décision

Analyse :

- Le modèle de l'Arbre de Décision présente une performance inférieure, avec 9 faux positifs et 8 faux négatifs..
- Il est sans doute trop calibré sur les données d'apprentissage, ce qui pourrait expliquer sa dégradation de performance sur les données test.

Tableau 3.8: Matrice de confusion – LogisticRegression

Prédit Réel	**Prédit Faux**	
0 FP	1 TP	Réel Faux
42 FN	44 TN	Réel Réel

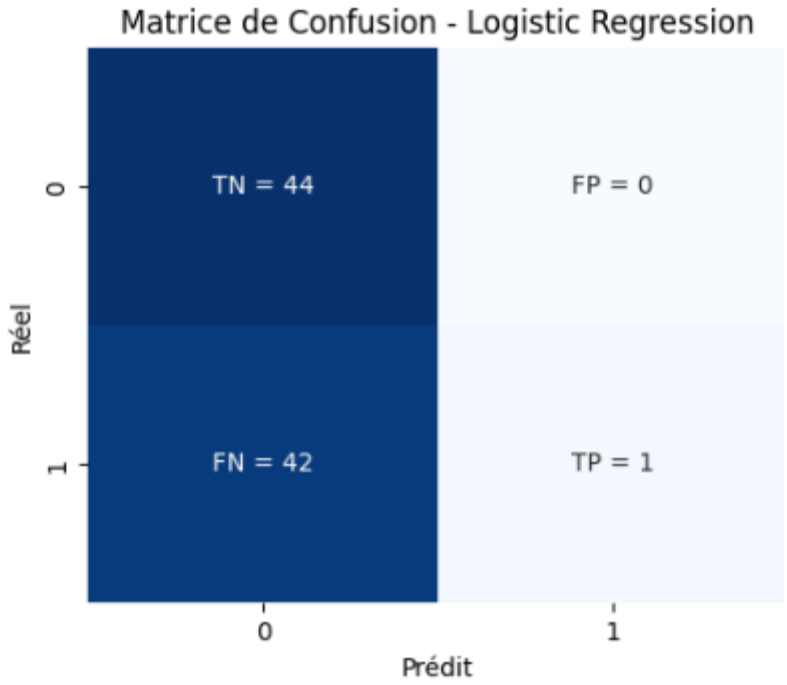


Figure 3.13 : Matrice de confusion - **Régression Logistique**

Analyse :

- La régression logistique présente un plus grand nombre de faux négatifs (11 FN), suggérant qu'elle pourrait manquer certains comptes suspects.
- Néanmoins, elle génère moins de faux positifs (5 FP), évitant ainsi d'identifier des comptes légitimes comme frauduleux.

Tableau 3.9: Matrice de confusion – ANN

Prédit Réel	**Prédit Faux**	
0 FP	5 TP	Réel Faux
38 FN	44 TN	Réel Réel

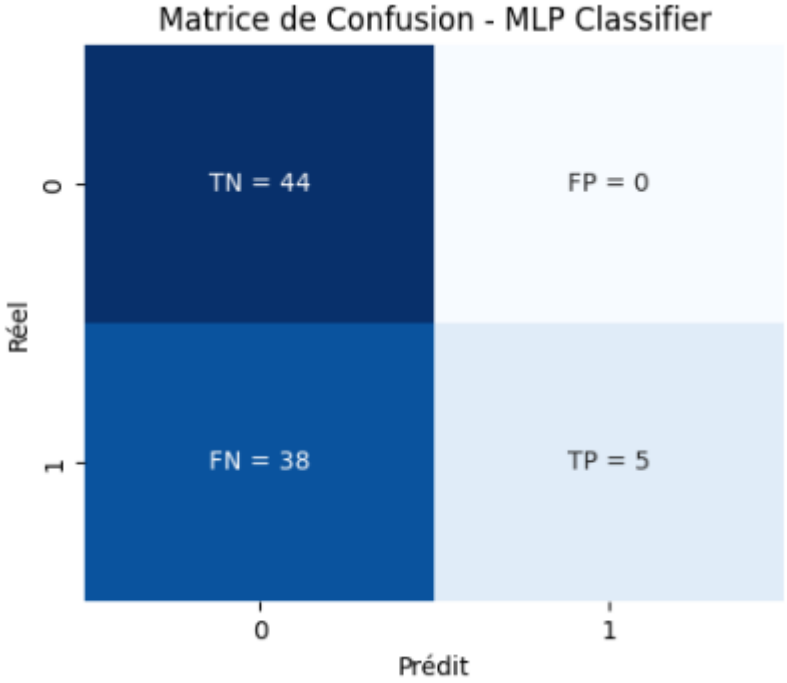


Figure 3.14 : Matrice de confusion – ANN

Analyse :

ANN a une **performance intermédiaire**, proche de **Random Forest** mais légèrement inférieur .

En générale

La Random Forest se distingue par son performance supérieure en matière de F1-Score, Recall et AUC-ROC, toutefois, il demande plus de temps de calcul et de ressources. Quant aux réseaux de neurones (Neural Network) , ils constituent un excellent équilibre entre performance et simplicité d'apprentissage. L'entraînement du DecisionTree et de la Régression Logistique est plus aisé, toutefois leurs performances tendent à être moins performantes que celles des modèles antérieurs, notamment en ce qui concerne l'identification des faux comptes.
En fin Si vous disposez de suffisamment de données et de ressources, les modèles les plus adaptés pour identifier les faux comptes Instagram seraient les réseaux de neurones ou Random Forest.

3.8 Utilisation de l' application pour prédire des comptes choisis

Il convient de noter que l'utilisation d'un compte Instagram réel et d'un autre faux à ce stade était uniquement à des fins expérimentales, dans le but de vérifier le bon fonctionnement du

code.

La figure 3.15 présentes les déférentes caractéristiques d'un compte réel choisis de notre ensemble de jeux de données

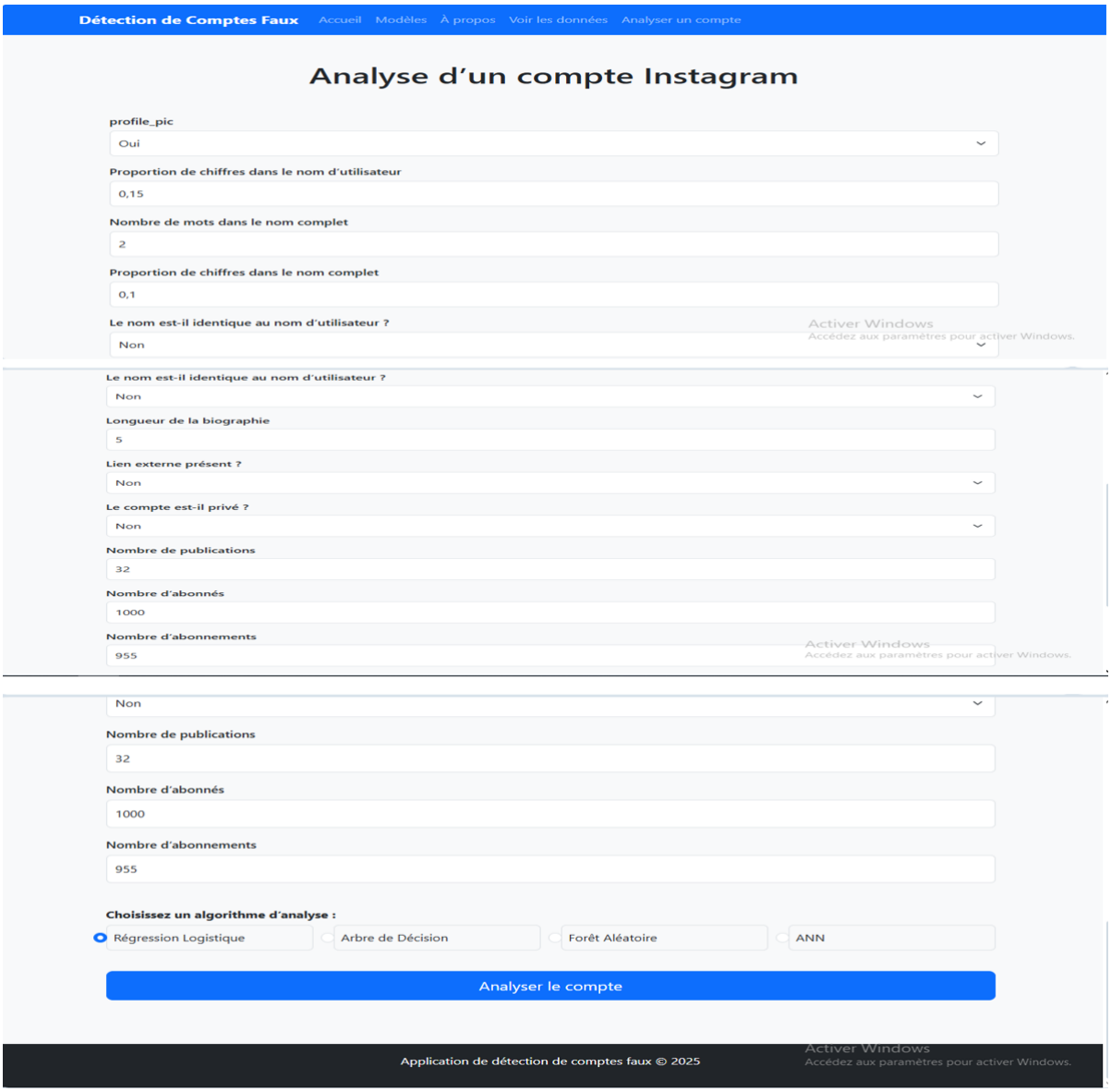


Figure 3.15 : Analyse d'un compte Instagram réel

La figure 3.16 illustre les Résultats d'analyse d'un compte Instagram réel avec les quatre modèles ou le Random Forest montre une précision de 100 % ce que fais de lui le modèle le plus exact comme nous avons déjà précis dans notre travail.

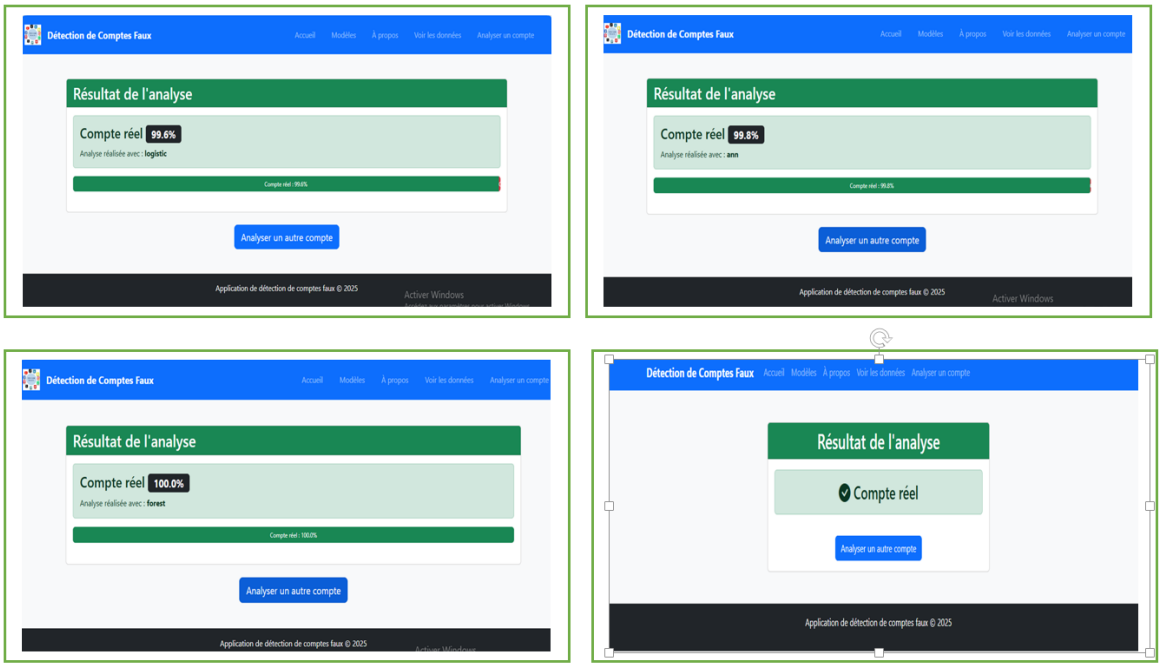


Figure 3.16 : Résultats d'analyse d'un compte Instagram réel avec les quatre modèles

La figure 3.17 présentes les déférentes caractéristiques d'un compte faux choisis de notre ensemble de jeux de données

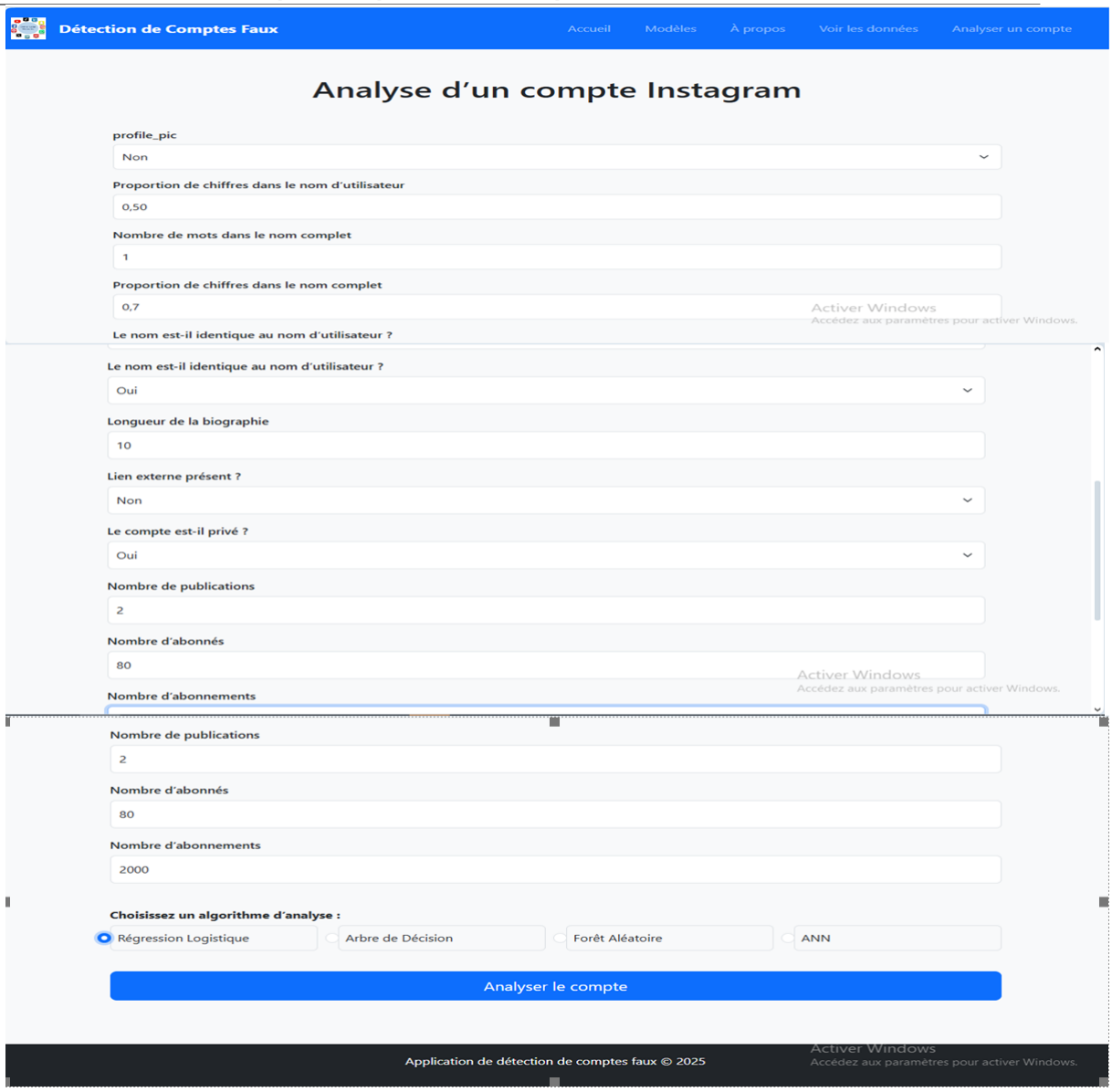


Figure 3.17: Analyse d'un compte Instagram faux

La figure 3.18 illustre les Résultats d'analyse d'un compte Instagramfauxavec une précision excellentes fournis par les quatre modèles utilisés dans notre travail .

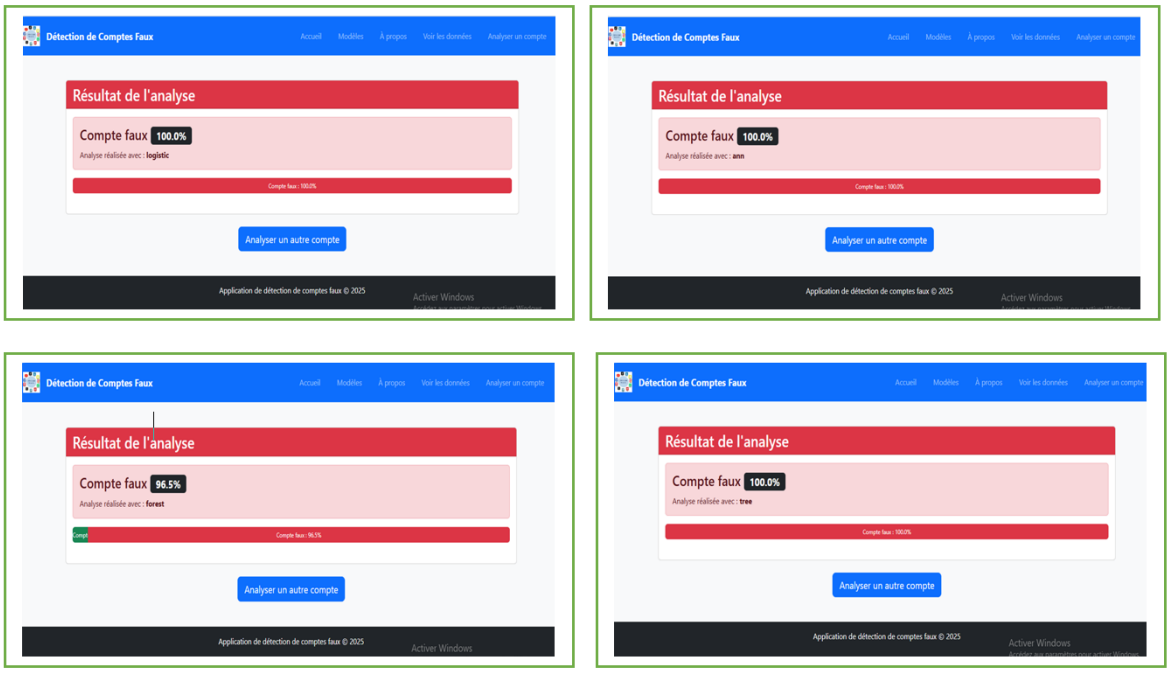


Figure 3.18: Résultats d'analyse d'un compte Instagram faux avec les quatre modèles

3.9 Conclusion

Ce chapitre expose en détail la mise en œuvre et l'expérimentation des modèles de détection de faux comptes sur Instagram. Nous avons commencés par mettre en place l'environnement de travail, ce qui inclut l'installation des bibliothèques et l'examen du fichier de données utilisé.

Par la suite, nous avons passés à la phase de traitement préalable des données, une étape cruciale comprenant le nettoyage, la transformation en format numérique, la séparation en ensembles dédiés à l'apprentissage, à la validation et au test, ainsi que l'homogénéisation des valeurs pour optimiser les performances des modèles.

Nous avons décrits ensuite les modèles employés, y compris des algorithmes traditionnels tels que l'arbre de décision, la forêt aléatoire et la régression logistique, sans oublier les réseaux de neurones artificiels (ANN). Tout en détaillant leur architecture, leur mode d'opération et les motifs qui ont guidé notre sélection.

Enfin, nous avons évalués ces modèles en utilisant plusieurs métriques de performance telles que **l'accuracy, le recall, le F1-score , AUC ROC et la matrice de confusion**, afin de déterminer le modèle le plus adapté à la détection des faux comptes

Conclusion Générale

Au terme de ce travail, nous avons exploré de manière approfondie la problématique de la détection des faux comptes sur Instagram, un enjeu crucial dans le contexte actuel de la cybersécurité et de la fiabilité des interactions numériques.

Dans un premier temps, nous avons dressé un état de l'art autour des faux comptes, en mettant en lumière leurs caractéristiques distinctives, les motivations de leur création, ainsi que les techniques classiques utilisées pour leur identification. Cette base théorique nous a permis de comprendre les fondements du problème et de positionner notre étude dans un cadre scientifique rigoureux.

Par la suite, nous avons analysé le rôle de l'intelligence artificielle, en particulier de l'apprentissage automatique et profond, dans l'automatisation et l'optimisation des processus de détection. En examinant les principaux paradigmes d'apprentissage supervisé, ainsi que les architectures de réseaux de neurones telles que le perceptron multicouche (MLP), nous avons mis en évidence les capacités de ces approches à modéliser des comportements complexes et non linéaires, caractéristiques des faux comptes.

Enfin, la mise en œuvre pratique a permis de valider les concepts étudiés à travers le développement d'une application de détection fondée sur des algorithmes éprouvés comme la régression logistique, l'arbre de décision, la forêt aléatoire et les réseaux de neurones. L'évaluation de ces modèles sur un ensemble de données réelles a montré des performances prometteuses, avec une précision élevée notamment pour la forêt aléatoire, démontrant ainsi la pertinence de notre démarche.

Cette étude ouvre la voie à plusieurs perspectives futures. Il serait intéressant d'enrichir la base de données avec des informations plus diversifiées (contenu des publications, interactions, métadonnées temporelles), d'intégrer des techniques d'apprentissage non supervisé ou semi-supervisé pour détecter des comportements anormaux sans étiquetage préalable, et d'exploiter les modèles de deep learning plus avancés comme les transformers pour une analyse contextuelle plus fine.

En somme, ce travail confirme le potentiel des approches d'intelligence artificielle dans le renforcement de la sécurité sur les plateformes sociales, tout en mettant en exergue l'importance d'une modélisation rigoureuse, d'un prétraitement pertinent des données et d'une évaluation méthodique pour garantir la fiabilité des systèmes de détection.

Bibliographie

[1] A. Gupta, H. Lamba et P. Kumaraguru, « Fake Accounts Detection on Social Media: A Survey », dans *IEEE Transactions on Computational Social Systems*, 2021.

[2] M. Conti, R. Poovendran et M. Secchiero, « Fakebook: Detecting Fake Profiles in Social Networks », dans *Actes de la 12e conférence internationale sur la disponibilité, la fiabilité et la sécurité (ARES)*, 2017. (Portait les numéros : 2, 6, 18)

[3] S. Cresci, R. Di Pietro et al., « The Paradigm-Shift of Social Spambots: Evidence, Theories, and Tools for the Arms Race », dans *WWW '17 : Actes de la 26e conférence internationale sur le World Wide Web*, 2017. (Portait les numéros : 3, 4, 9, 13, 15, 22, 25, 30, 36, 38)

[4] A. Gupta, H. Lamba et P. Kumaraguru, « Faking Sandy: Characterizing and Identifying Fake Images on Twitter during Hurricane Sandy », dans *Actes de la 22e conférence internationale sur le World Wide Web (WWW '13)*, 2013. DOI : 10.1145/2488388.2488483. (Portait les numéros : 5, 8, 14, 20, 24, 29, 35, 40)

[5] E. Ferrara, O. Varol, C. Davis, F. Menczer et A. Flammini, « The Rise of Social Bots », dans *Communications of the ACM*, vol. 59, no. 7, pp. 96–104, 2016. DOI : 10.1145/2818717. (Portait les numéros : 7, 10, 11, 16, 21, 26, 31, 33, 37)

[6] C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini et F. Menczer, « The Spread of Fake News by Social Bots », dans *Nature Communications*, vol. 9, no. 1, pp. 1–9, 2018. DOI : 10.1038/s41467-018-06930-7. (Portait les numéros : 12, 17, 19, 27, 32, 34, 39)

[7] J. Han, M. Kamber et J. Pei, *Data Mining: Concepts and Techniques*, 3e éd., Morgan Kaufmann, 2011. (Portait le numéro : 23)

[8] E. Alothali, N. Zaki, E. A. Mohamed et H. Alashwal, « Detecting Social Spam Campaigns on Twitter: A Machine Learning Approach », dans *Journal of Big Data*, vol. 5, no. 1, pp. 1–25, 2018. DOI : 1[9] A. Gulli and S. Pal, *Deep learning with Keras*. Packt Publishing Ltd, 2017.

[10] B. Liu and B. Liu, *Supervised learning*. Springer, 2011.

[11] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR)*, vol. 9, pp. 381–386, 2020.

[12] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of Massive Datasets*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2020.

[13] O. Chapelle, B. Scholkopf, and A. Zien, *Semi-Supervised Learning*. Cambridge, MA, USA: MIT Press, 2006.

[14] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[15] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.

[16] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Belmont, CA, USA: Wadsworth, 1984.

[17] D. R. Cox and E. J. Snell, *The Analysis of Binary Data*. CRC Press, 1989.

[18] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*. Wiley, 2000.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.

[20] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[21] ResearchGate, "Structure d'un MLP: Un algorithme d'apprentissage tel que la rétropropagation de gradient," 2019. [Online]. Available: https://www.researchgate.net/figure/Structure-dun-MLP-Un-algorithme-dapprentissage-tel-que-la-retropropagation-de-gradient_fig6_333200979. [Accessed: 19-Feb-2025]

[22]D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533-536, 1986

[24] J. R. Quinlan, "Induction of Decision Trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.

[25] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993.

[26] T. M. Mitchell, *Machine Learning*, McGraw-Hill, 1997.

[27] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[28] Blent.AI, "Arbres de décision en Machine Learning : tout comprendre," 2023. [Online]. Available: https://blent.ai/blog/a/arbres-de-decision-en-machine-learning.

[29] P. Chesneau, "Introduction aux arbres de décision (de type CART)," CNRS, 2022. [Online]. Available: https://chesneau.users.lmno.cnrs.fr/arbres.pdf.

[30] Quinlan, J. R. (1986). *Induction of Decision Trees*. Machine Learning.
[31] Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees*.

[32] Quinlan, J. R. (1986). *Induction of Decision Trees*. Machine Learning.

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

[34] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[35] Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media.

[36] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.

[37] Kingma, D. P., & Ba, J. (2015). *Adam: A Method for Stochastic Optimization*. ICLR.

[38] Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. *ICML*.

[39] Al-Qurishi, M., et al. (2019). *Countering Social Bots and Fake News in Social Media*. IEEE Access.

[40] Mikolov, T., et al. (2013). *Efficient Estimation of Word Representations in Vector Space*. arXiv preprint.

[41] T. Berners-Lee, *HTML and the World Wide Web*, W3C, 1999. [En ligne]. Disponible sur : www.w3.org

[42] H. W3C, *Cascading Style Sheets (CSS) – The Official Guide*, Addison-Wesley, 2021.

[43] D. Flanagan, *JavaScript: The Definitive Guide*, 7e éd., O'Reilly Media, 2020.

[44] J. Resig et B. Bibeault, *jQuery in Action*, 3e éd., Manning Publications, 2015

[45] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012. [Online].
Available: https://probml.github.io/pml-book/

[46] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley, 2001. [Online].

Available: https://www.wiley.com/en-us/Pattern+Classification%2C+2nd+Edition-p-9780471056690

[47] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2011. [Online]. Available: https://www.sciencedirect.com/book/9780123814791/data-mining-concepts- and-techniques

[48] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008. [Online]. Available: https://nlp.stanford.edu/IR-book/

[49] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, 2009. [Online]. Available: https://hastie.su.domains/Papers/ESLII.pdf

ملخص

تشكل الحسابات الوهمية على إنستغرام تحديًا متزايدًا في مجال الأمن السيبراني، حيث تُستخدم لأغراض غير مشروعة مثل الاحتيال ونشر المعلومات المضللة والتلاعب بالرأي العام. تهدف هذه الدراسة إلى تطوير نموذج يعتمد على الذكاء الاصطناعي لاكتشاف الحسابات الوهمية باستخدام تقنيات التعلم الآلي. ولتحقيق ذلك، تم الاعتماد على مجموعة بيانات متوازنة تحتوي على تسع خصائص رئيسية تساعد في التمييز بين الحسابات الحقيقية والمزيفة، مثل وجود صورة للملف الشخصي، تطابق الاسم مع اسم المستخدم، وجود رابط خارجي، حالة الخصوصية، وعدد المتابعين والمنشورات والمتابعات. تم تطبيق عدة خوارزميات تصنيف، من بينها شجرة القرار، الغابة العشوائية، الانحدار اللوجستي، والشبكات العصبية الاصطناعية. أظهرت النتائج أن خوارزمية الغابة العشوائية كانت الأكثر دقة، مما يؤكد فعاليتها في كشف الحسابات المزيفة. تسلط هذه الدراسة الضوء على دور الذكاء الاصطناعي في تعزيز أمان منصات التواصل الاجتماعي، كما تفتح المجال لإجراء أبحاث مستقبلية تهدف إلى تطوير أنظمة أكثر دقة وفعالية في اكتشاف الحسابات الوهمية بشكل تلقائي.

الكلمات المفتاحية : الحسابات الوهمية، إنستغرام، التعلم الآلي، الأمن السيبراني، الذكاء الاصطناعي، التلاعب بالمعلومات، التصنيف الآلي.

Abstract

Fake account detection on Instagram is a crucial challenge in the field of cybersecurity, as malicious users exploit fake profiles for fraudulent activities. In this study, we propose a detection method based on machine learning algorithms to identify fake accounts using key profile attributes. We employ various classification models, including Decision Tree, Random Forest, Logistic Regression, and Artificial Neural Networks (ANN) using MLPClassifier. The dataset used is sourced from Kaggle and contains nine features that distinguish real accounts from fake ones. Experimental results demonstrate that the Random Forest model achieves the highest accuracy, proving its effectiveness in detecting fake accounts. This study highlights the role of artificial intelligence in enhancing security on social media platforms and lays the foundation for further improvements in automated fake account detection.

Keywords: Fake accounts, Instagram, Machine Learning, Cybersecurity, Social Engineering, Artificial Intelligence.

Résumé

Votre résumé en français

La détection des faux comptes sur Instagram représente un défi majeur en cybersécurité, car des utilisateurs malveillants exploitent ces profils à des fins frauduleuses. Dans cette étude, nous proposons une méthode de détection basée sur des algorithmes d'apprentissage automatique afin d'identifier les faux comptes en utilisant des attributs clés des profils. Nous employons plusieurs modèles de classification, notamment l'arbre de décision, la forêt aléatoire, la régression logistique et les réseaux de neurones artificiels (ANN) avec **MLPClassifier**. Le jeu de données utilisé provient de Kaggle et comprend neuf caractéristiques permettant de distinguer les comptes réels des faux. Les résultats expérimentaux montrent que le modèle **Random Forest** atteint la meilleure précision ,prouvant ainsi son efficacité dans la détection des faux comptes. Cette étude met en évidence le rôle de l'intelligence artificielle dans l'amélioration de la sécurité sur les réseaux sociaux et ouvre la voie à de futures améliorations pour automatiser la détection des faux comptes.

Mots-clés : Faux comptes, Instagram, Apprentissage automatique, Cybersécurité, Ingénierie sociale, Intelligence artificielle.