République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche scientifique

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

Faculté Des Mathématiques Et Informatique

Rapport de projet de fin d'études

Présenté en vue de l'obtention

Du diplôme de master

Filière : Informatique

Spécialité : TIC

THEME :

Détection automatique des faux comptes sur les réseaux sociaux à l'aide de l'apprentissage profond

Présenté par :

- Boumaiza Amdjad
- Mohammedi Chahinez

Soutenu publiquement le 12/06/2025 devant le jury composé de :

M.	Boumaza Farid	Maître de Conférences	Université de BBA	Président
M.	Nouioua Mourad	Maître de Conférences	Université de BBA	Examinateur
M.	Naili makhlouf	Maître de Conférences	Université de BBA	Encadreur

Année Universitaire 2024-2025

شكر وتقدير

وَآخِرُ دَعْوَاهُمْ أَنِ الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ

ما سلكنا البدايات إلا بتيسيره، وما بلغنا النهايات إلا بتوفيقه، وما حققنا أسمى الغايات إلا بفضله، الحمد لله حبا وشكرا وامتنانا على البدء والختام

."لم تكن الرحلة قصيرة، ولم تكن الأمور يسيرة، ولكن بحول الله "نحن لها وإنَّ أَبَتْ، رُغمًا عنها أتينا بها

إلى "أبي الغالي" العزيز الذي حملت اسمه فخرًا، وبكل اعتزاز أنا لهذا الرجل ابنة، إلى من كلله الله بالهيبة والوقار، .لك الشكر الخالص، والدعاء الدائم، صاحب الإحسان والرحمة

إلى العظيمة أمي "الحبيبة" بركة الصباحات الدراسية، أنت نجاح الرحلة، وكفاح القلب، وإصرار التحدي، كم بدلت التعب بالراحة والفتور بالهمة، وخفوت الصوت بألحان القمم لك الشكر خالصًا، والدعاء مرارا يا من هللت عمري .بضيائك

إلى المشرف الفاضل، الذي لم يبخل عليّ بعلمه وتوجيهه، والذي كان لعطائه ودعمه الأثر البالغ في إنجاز هذه المذكرة، أتقدم إليك بخالص الشكر والتقدير، لما قدمته من وقت وجهد، وما غرسته في نفسي من ثقة وحافز للاستمرار فلك مني كل الامتنان والدعاء الصادق أن يجزيك الله عني خير الجزاء، وأن يجعل ما قدمته في ميزان حسناتك

إلى من شد الله بهم عضدي فكانوا خير معين إلى الشموع النيرة الذين انتظروا هذه اللحظة كثيرا ليفخروا بي كما أفخر ."بهم إلى "أخوتي

.إلى صديقاتي العزيزات كنتن السند والرفقة الطيبة، أنتن الأمل الذي ظل يضيء دربي

إلى جميع من كان له الأثر الجميل في حياتي. أهدي إليكم هذه الرسالة سائلة المولى أن ينفع بها .

Dédicace

À ma tendre mère Naïma, lumière de mon cœur et source de mes prières dans chaque moment de fatigue,

À mon père Ibrahim, mon soutien et la force qui a porté mes rêves,

À mes chers frères et sœurs : Imane, Khadija, Mohamed et Adel,
Compagnons de route et réconfort de mon âme,

À Arij, la fille de ma sœur bien-aimée, qui remplit la vie de douceur et de joie,

À mes chères amies Afnan et Feryal,Aya,Ibtissem ,zineb qui ont partagé avec moi les beaux moments et les défis,

Et à chems, la fille de ma cousine, qui a illuminé mon chemin par sa gentillesse et sa sincérité.

Et à ma chère grand-mère Barkahoum, mère de mon père,
Dont les prières sincères et constantes ont veillé sur moi avec amour et espoir.

Ce succès est le fruit de votre amour et de vos prières,
Je vous l'offre du fond du cœur,
Avec la promesse de toujours être digne de votre confiance.

Mohammedi Chahinez

Dédicace

A la lumière de ma vie, la source de tendresse, ma première supportrice et mon amour éternelle, ma mére que j'adore Hada.

A mon très cher pére Abderahman , pour ses encouragements, sen seutien, et surtout pour sen amour et sen sacrifice afin que rien n'entrave le déroulement de mes études.
A Mes chères sœurs Chaima et Asma, qui n'ont pas cessés de m'encourager tout au cours de réalisation de ce travail.

A mes frères Ayoub, Adem et Aymen.
A ma grande mère, mes oncles et cousines.

A mes chères copines Feriel,Chahinez, amel,merci pour tous les moments inoubliables et de m'a toujours encouragé et m'aimé.

A tous mes autres proches ;
Tous simplement, a tous ceux que j'aime et qui m'aiment.

Boumaiza Amdjad

Résumé

À l'ère du numérique, les réseaux sociaux sont devenus des vecteurs incontournables de communication, de diffusion d'information et de promotion, tant pour les particuliers que pour les entreprises. Cependant, cette omniprésence s'accompagne d'un phénomène préoccupant : la prolifération des faux comptes, en particulier sur des plateformes comme Instagram. Ces profils inauthentiques, souvent générés automatiquement ou créés dans un but malveillant, compromettent la sécurité des utilisateurs, faussent les statistiques d'engagement et servent parfois à des campagnes de désinformation ou de manipulation.

Afin de répondre à cette problématique, ce mémoire propose une approche basée sur l'apprentissage profond, et plus précisément sur l'utilisation des réseaux de neurones LSTM (Long Short-Term Memory), capables de modéliser les données temporelles et séquentielles propres au comportement des utilisateurs. En exploitant un ensemble de données simulé décrivant l'activité de plusieurs centaines de comptes Instagram, un modèle de classification a été entraîné et évalué avec succès.

Les résultats obtenus démontrent l'efficacité de cette approche en termes de précision, de robustesse et de capacité à détecter des comportements anormaux. Ce travail s'inscrit dans une démarche de renforcement de la cybersécurité et de lutte contre les menaces numériques via les technologies d'intelligence artificielle.

Abstract

In the digital age, social media platforms have become essential tools for communication, information dissemination, and brand visibility. However, this widespread use has given rise to a growing concern: the proliferation of fake accounts, particularly on Instagram. These inauthentic profiles, often automated or maliciously crafted, pose serious threats to user security, distort engagement metrics, and serve as vehicles for disinformation and fraudulent activities.

To address this challenge, this thesis presents a deep learning-based approach using Long Short-Term Memory (LSTM) neural networks, which are well-suited to modeling the sequential and behavioral data of social media users. A synthetic dataset representing Instagram accounts was used to train and evaluate the model.

The results highlight the method's ability to accurately classify accounts as genuine or fake, offering strong performance metrics and promising generalization capabilities. This research contributes to the broader field of cybersecurity and illustrates the potential of artificial intelligence in detecting online threats and enhancing digital platform integrity.

ملخص

في عصر الرقمنة، أصبحت وسائل التواصل الاجتماعي أدوات لا غنى عنها للتواصل، ونشر المعلومات، والترويج، سواء للأفراد أو المؤسسات. ومع ذلك، فإن هذا الانتشار الواسع ترافقه ظاهرة مقلقة تتمثل في الانتشار الكبير للحسابات الوهمية، وخاصة على منصات مثل إنستغرام. هذه الحسابات غير الحقيقية، والتي يتم إنشاؤها غالبًا بشكل تلقائي أو لأغراض خبيثة، تُهدد أمن المستخدمين، وتُشوّه إحصاءات التفاعل، وقد تُستخدم أحيانًا في حملات تضليل أو تلاعب بالرأي العام.

ولمواجهة هذه المشكلة، يقترح هذا البحث مقاربة قائمة على التعلم العميق، وتحديدًا باستخدام الشبكات العصبية من نوع الذاكرة القصيرة والطويلة المدى (LSTM)، التي تمتاز بقدرتها على نمذجة البيانات الزمنية والمتسلسلة المرتبطة بسلوك المستخدمين. من خلال استغلال مجموعة بيانات مُحاكية تصف نشاط مئات الحسابات على إنستغرام، تم تدريب نموذج تصنيف وتقييم أدائه بنجاح.

وتُظهر النتائج المُتحصّل عليها فعالية هذا النموذج من حيث الدقة والصلابة والقدرة على كشف السلوكيات غير الطبيعية. ويُعد هذا العمل مساهمة في تعزيز الأمن السيبراني ومكافحة التهديدات الرقمية باستخدام تقنيات الذكاء الاصطناعي

Table des matières

Liste des figures

Liste des Tableaux

Liste des abréviations

- **AI:** Artificiel Intelligence
- **ML:** Machine Learning
- **DL:** Deep Learning
- **RL :** Apprentissage par Renforcement
- **NLP :** Traitement du Langage Naturel
- **CNN :** Les réseaux de Neurones Convolutionnels
- **RNN :** Réseaux de Neurones Récurrents
- **GAN :** Réseaux de Neurones Génératifs
- **LSTM :** Mémoire longue à Court Terme
- **ROC:** Receiver Operating Characteristic
- **PRC:** Precision-Recall Curve

Introduction générale

Introduction générale

Les réseaux sociaux, notamment Instagram, occupent aujourd'hui une place centrale dans la vie quotidienne des individus, des entreprises et des institutions. Ils permettent la communication, le marketing, le partage d'informations et bien plus encore. Cependant, cette croissance fulgurante a également favorisé l'émergence de comportements nuisibles, dont la prolifération des faux comptes (comptes fictifs) qui représentent une menace réelle pour la sécurité numérique. Ces comptes sont souvent utilisés pour des activités malveillantes telles que l'escroquerie, la diffusion de fausses informations ou encore la manipulation de l'opinion publique.

Face à cette problématique, il devient impératif de développer des outils intelligents capables de détecter ces comptes frauduleux avec précision. Le Deep Learning (apprentissage profond) s'est imposé ces dernières années comme une solution puissante, notamment grâce à sa capacité à analyser de grandes quantités de données et à détecter des motifs complexes. Parmi les modèles les plus performants, le réseau de neurones à mémoire à long et court terme (LSTM - Long Short-Term Memory) se distingue particulièrement dans le traitement de données séquentielles et temporelles, telles que les activités d'un utilisateur sur une plateforme sociale.

Ce travail vise à concevoir et entraîner un modèle basé sur l'architecture LSTM afin de détecter les faux comptes sur Instagram en se basant sur l'analyse du comportement utilisateur. Nous passerons par toutes les étapes essentielles : le prétraitement des données, la structuration du modèle, l'entraînement, puis l'évaluation de ses performances.

À travers cette étude, nous souhaitons apporter une contribution concrète dans le domaine de la cybersécurité et de l'intelligence artificielle, en exploitant les avantages du Deep Learning pour relever un défi actuel lié à la fiabilité des réseaux sociaux.

Chapitre I : Détection des faux comptes via les réseaux sociaux

I.1. Introduction

À l'ère du numérique, les réseaux sociaux ont transformé en profondeur la manière dont les individus interagissent, communiquent et consomment l'information. Toutefois, cette démocratisation de l'expression en ligne s'accompagne d'effets pervers, notamment la prolifération des comptes inauthentiques, qui compromettent l'authenticité et la fiabilité des plateformes. Instagram, avec ses milliards d'utilisateurs actifs, est particulièrement exposé à ce phénomène. Ces faux comptes, souvent créés à des fins de fraude, de manipulation d'opinion ou de promotion commerciale malveillante, nuisent à la sécurité des utilisateurs et à la crédibilité des indicateurs de performance. Ce chapitre propose une analyse approfondie de ce problème : il commence par définir la notion de réseau social, en présente les enjeux sociétaux et technologiques, puis explore les stratégies de création de faux profils, leurs objectifs, et les risques associés. Enfin, il met en lumière les indicateurs comportementaux permettant leur détection et souligne l'importance croissante du recours à l'intelligence artificielle pour contrer efficacement ces menaces.

I.2. Les Réseaux sociaux

Une "société en réseau" désigne une structure sociale qui a émergé en raison de trois facteurs clés : les mouvements culturels et sociaux des années 1960 et 1970, l'essor des technologies de l'information numérique et la domination de l'économie néo-libérale. Dans une société en réseau, l'organisation sociale est caractérisée par le traitement et la gestion de l'information à travers des réseaux, en particulier Internet. Cela a entraîné un changement dans la manière dont les scientifiques communiquent et collaborent, Internet servant de plateforme pour l'échange d'idées, de ressources et de connaissances à l'échelle mondiale. De plus, la société en réseau a estompé les frontières entre les différents rôles et les publics dans la communication scientifique, permettant l'accessibilité et la diffusion de l'information scientifique à un public plus large. [1]

I.2.1. Définition

Un réseau social est un site web ou une application qui permet aux personnes de se connecter les unes aux autres sur une plateforme commune. Les utilisateurs peuvent partager des informations, exprimer des opinions, explorer des intérêts communs, rechercher des emplois, promouvoir leurs entreprises, établir des relations et interagir de diverses manières. Les personnes qui participent à un réseau social partagent souvent une

large gamme d'informations et de contenus, notamment des photos, des vidéos, des extraits sonores, des documents, des actualités, du matériel marketing ou des liens vers d'autres ressources. [2]

La figure suivante offre un aperçu de certaines des plateformes de médias sociaux les plus populaires, telles que Facebook, Instagram, Twitter.

Figure I. 1. Réseaux sociaux

I.2.2. Impacts des réseaux sociaux

Les réseaux sociaux permettent aux personnes de développer des relations qui ne seraient pas possibles en raison des distances géographiques et temporelles. Ils contribuent également à améliorer la productivité des entreprises pour les relations publiques, le marketing et la publicité. [3]

L'impact des sites de réseaux sociaux sur la sécurité des individus se manifeste par la menace potentielle de devenir victime de comportements criminels. Cela conduit également à une diminution de la performance et de la productivité au travail, en raison de l'addiction à l'utilisation des sites de réseaux sociaux, ainsi qu'à une baisse des performances scolaires des étudiants. [4]

I.3.Analyse des comptes Instagram inauthentiques

Un compte Instagram inauthentique constitue une menace potentielle, souvent assimilée à du « spam », en usurpant l'identité d'un utilisateur connu ou en prétendant être

une personne réelle. Ces comptes, autrefois appelés « Finstas » (contraction de « Fake Instagram »), connaissent une croissance rapide en parallèle avec l'augmentation du nombre d'utilisateurs sur la plateforme et la complexité croissante de la gestion de ses activités [5].

Les comptes utilisant de fausses informations ou partageant du contenu trompeur sont considérés comme ayant pour but explicite ou implicite de manipuler d'autres utilisateurs. Selon des estimations récentes de 2024, environ 10 % des 2 milliards de comptes présents sur Instagram seraient des comptes fictifs [6].

Ces comptes inauthentiques poursuivent divers objectifs :

1. **Activités malveillantes** : la création de faux comptes permet à certains acteurs de commettre des fraudes, des tentatives d'hameçonnage ou encore des actes de chantage envers d'autres utilisateurs.
2. **Diffusion de spam** : ils servent à promouvoir de manière répétée des produits ou services commerciaux dans les publications.
3. **Influence des tendances** : en multipliant les publications similaires mais légèrement reformulées, ces comptes cherchent à influer sur certaines tendances tout en échappant aux mécanismes de détection automatique.
4. **Augmentation artificielle de la popularité** : certains utilisateurs ont recours à de faux comptes pour accroître leur nombre d'abonnés, renforçant ainsi artificiellement leur influence perçue.
5. **Commerce d'interactions** : enfin, ces comptes alimentent un marché d'achat et de vente de faux abonnés et de mentions « J'aime », destinés à gonfler artificiellement l'engagement sur la plateforme. [6]

I.4.Détection des faux comptes Instagram

Les faux comptes sont fréquemment exploités pour des activités malveillantes telles que la collecte frauduleuse de données personnelles, la diffamation ou encore la diffusion de logiciels malveillants. Grâce aux avancées de l'intelligence artificielle (IA), les systèmes informatiques sont désormais capables de reproduire certains comportements humains. Les approches actuelles de détection de faux profils reposent essentiellement sur l'analyse de caractéristiques prédéfinies (approches basées sur les attributs) ou sur des représentations graphiques permettant de visualiser et de détecter les anomalies dans les réseaux sociaux.

Ces techniques mobilisent notamment l'apprentissage automatique (ML) ainsi que l'apprentissage profond (DL). [7]

Figure I. 2.Détection d'inauthenticité en ligne

➢ Indicateurs de détection des faux comptes sur Instagram

Dans ce qui suit, on cite une liste des principaux indicateurs qui permettent de détecter des faux comptes sur l'Instagram :

- **profile pic :** Indique si le compte possède une photo de profil (1 = oui, 0 = non)
- **nums/length username :** Ratio du nombre de chiffres par longueur totale du nom d'utilisateur
- **fullname words** : Nombre de mots présents dans le nom complet du compte
- **nums/length fullname :** Ratio de chiffres par rapport à la longueur du nom complet
- **name==username :** Indique si le nom d'utilisateur est identique au nom complet (1 = oui, 0 = non)
- **description length :** Nombre de caractères dans la description du profil
- **external URL :** Présence d'un lien externe dans la bio (1 = oui, 0 = non)
- **private :** Statut de confidentialité du compte (1 = privé, 0 = public)
- **#posts :** Nombre total de publications
- **#followers :** Nombre d'abonnés (followers)

- **#follows :** Nombre de comptes suivis (followings)**Importance de la détection des faux comptes dans les réseaux sociaux**

Dans le monde numérique d'aujourd'hui, le vol d'identité et les faux comptes représentent des menaces importantes pour les organisations. Ces comptes peuvent être utilisés pour tromper les clients et distribuer la désinformation pour endommager la réputation de la marque. Le défi consiste à gérer ces comptes rapidement et efficacement sur une variété de plateformes. [8]

Des profils incorrects sont souvent des rapports avec une intention frauduleuse à des fins telles que le spam, le phishing, la fraude ou la manipulation de l'opinion publique. Ces profils affectent la fiabilité et la sécurité de la plate-forme, conduisant à de nombreux résultats négatifs pour les utilisateurs et les fournisseurs de services. En conséquence, la détection et la dépréciation de faux profils pour les plateformes de médias sociaux sont devenues une priorité. [9]

I.5.Conclusion

Ce chapitre a mis en lumière l’ampleur du phénomène des faux comptes sur les réseaux sociaux, en particulier sur Instagram. Nous avons démontré que ces profils frauduleux représentent un danger tangible tant pour les utilisateurs que pour les plateformes elles-mêmes, en menaçant la sécurité, la transparence et la fiabilité des interactions en ligne. À travers l’étude des objectifs poursuivis par ces comptes, des schémas comportementaux typiques, ainsi que des indicateurs utilisés pour les identifier, il apparaît clairement qu’une simple modération manuelle est insuffisante face à une menace aussi dynamique et évolutive. Dans cette optique, les techniques d’apprentissage automatique et profond offrent des perspectives prometteuses pour automatiser la détection et renforcer la résilience des réseaux sociaux. Cette analyse théorique prépare ainsi le terrain pour la conception et l’évaluation, dans les chapitres suivants, d’un modèle basé sur le Deep Learning dédié à la détection des faux comptes.

Chapitre II : Deep Learning

II.1. Introduction

L'émergence de l'intelligence artificielle (IA) et de ses sous-domaines tels que l'apprentissage automatiqcue (Machine Learning) et l'apprentissage profond (Deep Learning) marque une révolution technologique majeure dans de nombreux secteurs. Pour comprendre et exploiter pleinement les possibilités offertes par ces technologies, notamment dans le contexte de la détection de comportements frauduleux sur les réseaux sociaux, il est essentiel d'en maîtriser les fondements théoriques. Ce chapitre vise ainsi à définir et à distinguer les concepts clés de l'IA, du Machine Learning et du Deep Learning, en retraçant leur évolution historique et en explorant leurs principes de fonctionnement. Il présente également les principaux types de réseaux neuronaux, avec une attention particulière portée sur les réseaux LSTM, au cœur de notre approche. Ce socle théorique constitue une étape incontournable pour aborder ensuite la mise en œuvre concrète des modèles de détection.

II.2. L'apprentissage profond

II.2.1. Historique

L'apprentissage profond (le Deep Learning) repose sur les réseaux de neurones artificiels, inspirés du fonctionnement du cerveau humain. Son histoire commence dès le milieu du XXe siècle avec des avancées théoriques majeures [10] :

- **1943** : McCulloch et Pitts développent le premier modèle mathématique de neurone artificiel basé sur la logique seuil.
- **1958** : Frank Rosenblatt introduit le perceptron, un réseau à deux couches pour la reconnaissance de formes.
- **1980** : Kunihiko Fukushima conçoit le Neocognitron, un réseau neuronal multicouche hiérarchique utilisé pour la reconnaissance de l'écriture manuscrite.
- **1989** : Les premiers algorithmes à réseaux neuronaux profonds sont développés, mais trop lents à entraîner pour un usage pratique.
- **1992** : Juyang Weng propose le Cresceptron, capable de reconnaître des objets 3D dans des scènes complexes.
- **Milieu des années 2000** : Le terme "deep learning" se popularise grâce aux travaux de Hinton et Salakhutdinov sur la préformation couche par couche.

- **2009** : Des chercheurs découvrent que, avec suffisamment de données, les réseaux profonds peuvent s'entraîner sans préformation, avec des taux d'erreurs bien plus faibles.
- **2012** : Les algorithmes atteignent des performances proches de celles de l'humain pour certaines tâches (ex : reconnaissance d'images).
- **2014** : Google rachète la startup britannique DeepMind pour 400 millions de livres.
- **2015** : Facebook intègre DeepFace, un système de reconnaissance faciale basé sur plus de 120 millions de paramètres.
- **2016** : AlphaGo de Google DeepMind bat le champion du monde de Go, démontrant la puissance du deep learning dans des environnements complexes

II.2.2. Définition

L'apprentissage profond est un sous-ensemble de l'apprentissage automatique qui utilise des réseaux neuronaux multicouches, appelés réseaux neuronaux profonds, pour simuler le pouvoir décisionnel complexe du cerveau humain. Une certaine forme d'apprentissage profond alimente la plupart des applications d'intelligence artificielle (IA) présentes dans nos vies aujourd'hui.[11]

II.2.3. Mode de fonctionnement

Inspirés de la structure du cerveau humain, constitué de millions de neurones interconnectés, les modèles d'apprentissage profond reposent sur des réseaux neuronaux artificiels composés de couches de nœuds capables de traiter et transmettre progressivement l'information. Dans le contexte de la détection de faux comptes Instagram, ces modèles permettent d'extraire automatiquement des caractéristiques pertinentes à partir de données telles que le nombre de publications, la longueur du nom d'utilisateur, la correspondance entre le nom et l'identifiant, la longueur de la biographie, la présence de liens externes ou encore les ratios entre abonnés et abonnements. En s'appuyant sur de vastes ensembles de données annotées, l'apprentissage en profondeur ajuste les poids et paramètres du réseau neuronal afin de réduire l'écart entre les prédictions du modèle et les étiquettes réelles (faux ou légitimes). Au fil des itérations, le modèle affine sa capacité de généralisation, atteignant ainsi des performances de classification élevées. Le langage Python constitue aujourd'hui un standard dans ce domaine grâce à sa simplicité et à ses bibliothèques spécialisées, telles

que TensorFlow et PyTorch, qui facilitent la mise en œuvre de modèles complexes pour l'analyse comportementale et relationnelle des profils sociaux [15]

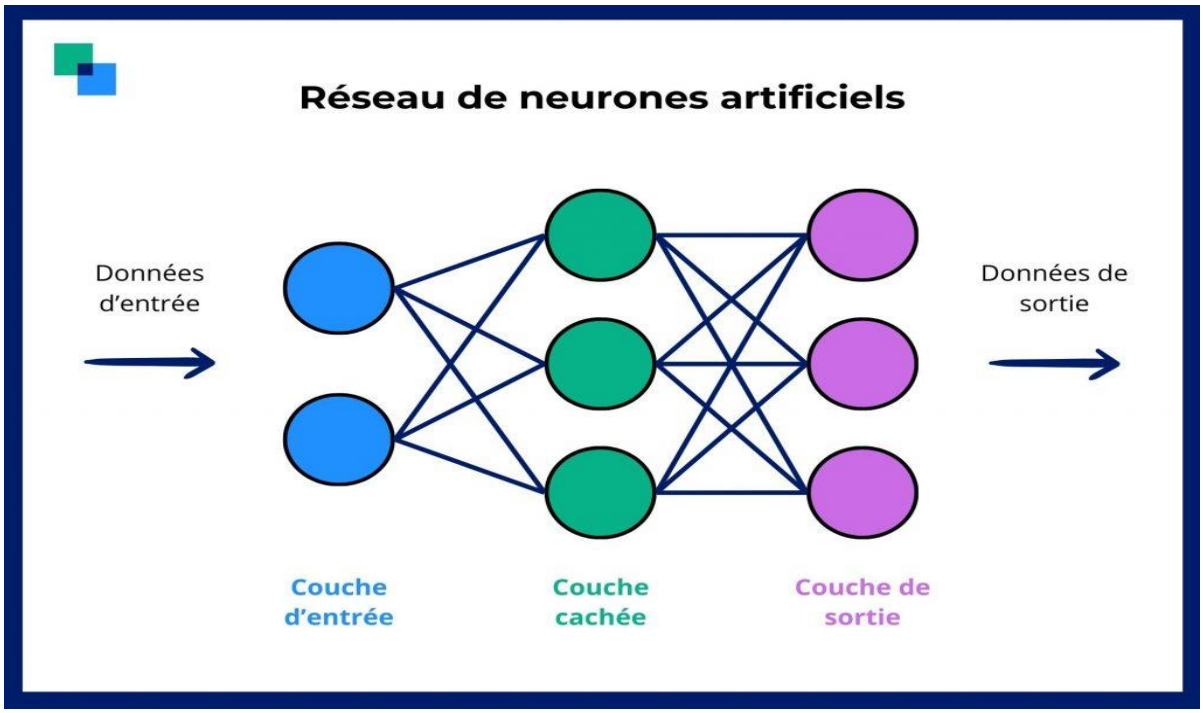


Figure II. 1.Mécanisme d'apprentissage profond.

II.2.4. L'apprentissage automatique

L'intelligence artificielle compte parmi ses branches l'apprentissage automatique (ML), qui représente un domaine spécifique de l'informatique et de la science des données. Cette discipline se caractérise par la capacité des systèmes à évoluer et à perfectionner leurs performances grâce à l'analyse de données, sans qu'il soit nécessaire de procéder à des modifications de programmation additionnelles [11]

II.2.4.1. Types d'apprentissage automatique

L'intelligence artificielle compte parmi ses branches l'apprentissage automatique (ML), qui représente un domaine spécifique de l'informatique et de la science des données. Cette discipline se caractérise par la capacité des systèmes à évoluer et à perfectionner leurs performances grâce à l'analyse de données, sans qu'il soit nécessaire de procéder à des modifications de programmation additionnelles.

a. L'apprentissage supervisé

L'apprentissage supervisé constitue une approche de l'apprentissage automatique dans laquelle le modèle développe ses compétences grâce à un corpus de données

préalablement annotées, où chaque exemple est associé à sa solution ou classification attendue.[11]

b. L'apprentissage non supervisé

Les méthodes d'apprentissage non supervisé — telles qu'Apriori, les modèles de mélanges gaussiens (GMM) et l'analyse en composantes principales (ACP) — extraient des informations significatives à partir d'ensembles de données dépourvus d'annotations. Cette approche favorise l'exploration des données, permet de détecter des schémas ou structures sous-jacents, et offre la possibilité d'élaborer des modèles prédictifs sans connaissance préalable des résultats souhaités.[11]

c. Apprentissage auto-supervisé

L'apprentissage auto-supervisé [11] constitue une approche permettant aux modèles de se former à partir de données brutes non étiquetées, évitant ainsi la dépendance aux vastes corpus de données préalablement annotées. Ces algorithmes, qu'on désigne aussi comme des méthodes d'apprentissage par prétexte ou prédictif, fonctionnent en déduisant certains éléments d'entrée à partir d'autres composantes, créant ainsi leurs propres étiquettes et transformant des défis non supervisés en problématiques supervisées. Cette technique s'avère particulièrement précieuse dans les domaines de la vision artificielle et du traitement automatique du langage, où l'ampleur des données étiquetées traditionnellement requises pour l'entraînement peut représenter un obstacle considérable, voire insurmontable.

d. Apprentissage par renforcement

L'apprentissage du renforcement est également une forme de programmation dynamique qui utilise des systèmes de récompense et des pénalités pour induire des algorithmes en améliorant les rendements humains (RLHF - renforcement d'apprentissage fait avec une rétroaction humaine). Pour mettre en œuvre l'apprentissage avec renforcement, les agents effectuent des actions dans un environnement spécifique pour atteindre des objectifs prédéfinis. Les agents sont récompensés ou punis pour leurs actions en fonction des mesures définies (général). [11]

e. Apprentissage semi-supervisé

La cinquième technique d'apprentissage automatisée représente un compromis entre l'apprentissage surveillé et non surveillance.

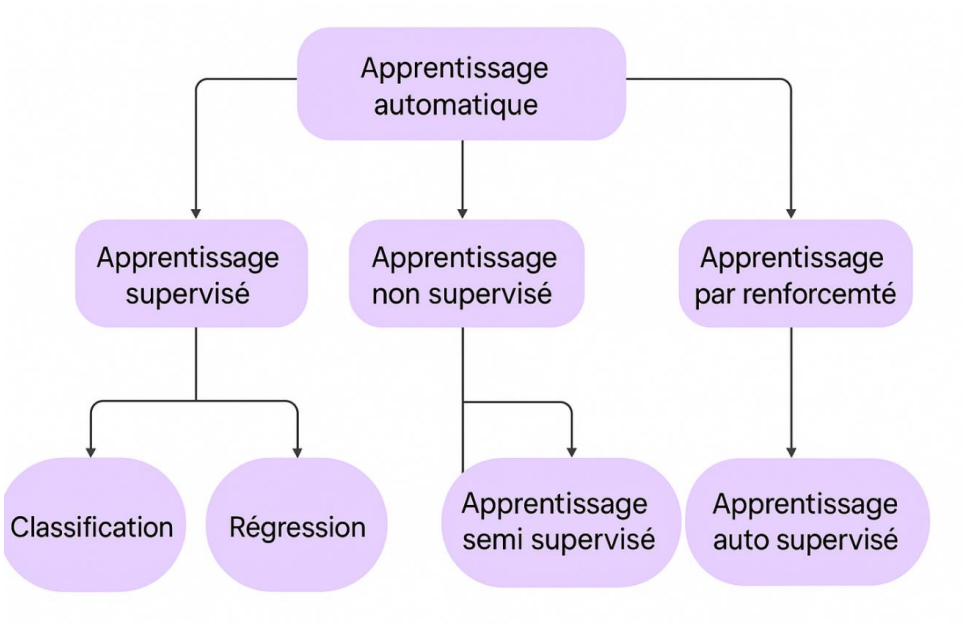
Les algorithmes d'apprentissage semi-spécialisés sont contrôlés à partir d'un petit ensemble de données marquées et d'une grande série de données intactes.

Les données marquées servent de guide pour orienter le processus d'apprentissage pour le reste des données détaillées. Par exemple, peut utiliser un modèle de semi-perméation illimité d'apprentissage pour identifier les groupes (ou clusters) dans les données, et utiliser l'apprentissage surveillé pour attribuer ces groupes.

Le réseau d'antagonistes génétiques (Geose Accident Network) est un outil d'apprentissage en profondeur qui tire des données rayées à travers deux réseaux de neurones, et est un exemple de l'utilisation de l'apprentissage semi-sous-couvert. Quel que soit le type utilisé, les modèles d'apprentissage automatique peuvent extraire des informations utiles des données de l'entreprise.

Il est important de mettre en place des pratiques d'intelligence artificielle responsables au sein d'une organisation en raison de la sensibilité à la sécurité face aux biais ou aux distorsions humaines disponibles dans les données.[11]

Figure II. 2.Types d'apprentissage automatique



II.2.5. L'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond

- **L'intelligence artificielle**

L'intelligence artificielle (IA), l'apprentissage automatique (Machine Learning, ML) et l'apprentissage profond (Deep Learning, DL) sont souvent utilisés de manière

interchangeable lorsqu'il s'agit de tout ce qui concerne l'IA. Bien que ces termes soient liés, ils ne sont pas interchangeables.

- **L’apprentissage automatique**

Alors que l'IA est un domaine large, l'apprentissage automatique est une application de l'IA qui permet aux machines d'apprendre sans être spécifiquement programmées. L'apprentissage automatique est utilisé plus explicitement comme un moyen d'extraire des connaissances à partir de données grâce à des méthodes plus simples telles que les arbres de décision ou la régression linéaire, tandis que l'apprentissage profond utilise des méthodes plus avancées basées sur les réseaux neuronaux artificiels.

- **L’apprentissage profond**

L'apprentissage profond nécessite moins d'intervention humaine, car les caractéristiques d'un ensemble de données sont extraites automatiquement, contrairement aux techniques d'apprentissage automatique plus simples qui nécessitent souvent qu'un ingénieur identifie manuellement les caractéristiques et les classificateurs des données et ajuste l'algorithme en conséquence. En essence, l'apprentissage profond peut apprendre de ses propres erreurs alors que l'apprentissage automatique nécessite une intervention humaine.

L'apprentissage profond nécessite également beaucoup plus de données que l'apprentissage automatique, ce qui nécessite à son tour une puissance de calcul beaucoup plus importante. L'apprentissage automatique peut généralement être effectué avec des serveurs équipés de processeurs (CPUs), tandis que l'apprentissage profond nécessite souvent des puces plus robustes telles que les unités de traitement graphique (GPUs). [13]

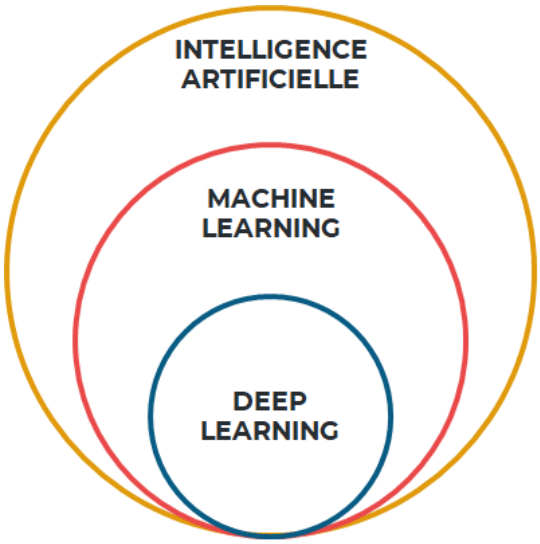


Figure II. 3.Relation entre l'IA, le ML et DL

II.2.6. Différences entre l'apprentissage profond et l'apprentissage automatique

La principale différence entre l'apprentissage profond et l'apprentissage automatique réside dans la structure de l'architecture du réseau neuronal sous-jacent. Les modèles d'apprentissage automatique « non profonds » traditionnels utilisent des réseaux neuronaux simples avec une ou deux couches de calcul. Les modèles d'apprentissage profond utilisent trois couches ou plus—mais généralement des centaines ou des milliers de couches pour entraîner les modèles.

Alors que les modèles d'apprentissage supervisé nécessitent des données d'entrée structurées et étiquetées pour produire des résultats précis, les modèles d'apprentissage profond peuvent utiliser l'apprentissage non supervisé. Avec l'apprentissage non supervisé, les modèles d'apprentissage profond peuvent extraire les caractéristiques, les traits et les relations dont ils ont besoin pour produire des résultats précis à partir de données brutes et non structurées. De plus, ces modèles peuvent même évaluer et affiner leurs résultats pour une précision accrue.

L'apprentissage profond est un aspect de la science des données qui alimente de nombreuses applications et services améliorant l'automatisation, en effectuant des tâches analytiques et physiques sans intervention humaine. Cela permet à de nombreux produits et services du quotidien—tels que les assistants numériques, les télécommandes vocales, la détection de fraude par carte de crédit, les voitures autonomes et l'IA générative de fonctionner. [13]

II.2.7. Avantages et inconvénients de l'apprentissage profond

- **Avantages de l'apprentissage profond**
  - Apprentissage automatique des caractéristiques - Le deep learning peut détecter et apprendre les éléments importants à partir des données sans avoir besoin que l'humain les définisse manuellement, ce qui est très utile pour des tâches comme la reconnaissance d'images
  - Gestion efficace de données massives et complexes - Ces modèles sont capables de traiter des volumes énormes de données très complexes, là où les algorithmes traditionnels montrent rapidement leurs limites
  - Excellente performance dans divers domaines - Les algorithmes de deep learning offrent des résultats remarquables dans des domaines comme la vision par ordinateur, la reconnaissance vocale et le traitement du langage naturel
  - Compréhension des relations complexes dans les données - Contrairement aux méthodes classiques, le deep learning peut identifier des relations non linéaires (plus subtiles) entre les variables.
  - Capacité à gérer différents types de données - Que les données soient structurées ou non (images, texte, audio...), les réseaux de deep learning peuvent les traiter efficacement.
  - Modèles puissants et adaptés à des problèmes complexes - Avec ses nombreuses couches, le deep learning peut résoudre des problèmes de grande ampleur et de haute complexité
  - Apprentissage automatique avec peu ou pas de supervision - Ces modèles peuvent apprendre à représenter les données sans avoir besoin d'étiquettes explicites, ce qui réduit l'effort humain
  - Souplesse et réutilisation dans d'autres contextes - Les modèles entraînés peuvent être ajustés pour d'autres tâches similaires, même avec peu de données supplémentaires
  - Robustesse face aux données manquantes - Les algorithmes peuvent apprendre à combler ou ignorer les valeurs manquantes, ce qui les rend utiles quand les données sont incomplètes [15]

- **Inconvénients**

- Besoins énormes en données - Le deep Learning nécessite une très grande quantité de données de qualité pour fonctionner efficacement. Cette exigence implique un

temps considérable et des ressources importantes pour collecter, nettoyer et préparer les données nécessaires à l'entraînement.

- Forte dépendance aux ressources matérielles - L'entraînement des modèles profonds sur de grands volumes de données demande des machines puissantes : processeurs (CPU), cartes graphiques (GPU), mémoire RAM, et espace de stockage élevé. Cela peut rendre leur déploiement coûteux.
- Risque de sur apprentissage (overfitting) - Le modèle peut apprendre « trop bien » les données d'entraînement, au point de perdre sa capacité à bien généraliser sur de nouvelles données. Cela diminue sa fiabilité et sa capacité à être utilisé dans des situations réelles.
- Difficulté d'interprétation des résultats - Contrairement aux algorithmes traditionnels plus transparents, les modèles de deep learning sont souvent perçus comme des « boîtes noires ». Il peut être très difficile d'expliquer comment et pourquoi une décision a été prise, ce qui est problématique dans des domaines sensibles.
- Problèmes juridiques et éthiques - Les modèles peuvent reproduire ou amplifier des biais présents dans les données d'entraînement. De plus, l'utilisation de données personnelles ou protégées par la propriété intellectuelle soulève des questions de respect de la vie privée et de conformité légale.
- Manque d'expertise métier (domaine) - Pour que le deep learning soit réellement efficace, il est indispensable d'avoir une bonne connaissance du domaine concerné (santé, finance, etc.). Sans cette expertise, il est difficile de définir le problème avec précision ou de choisir les bons algorithmes. [15]

II.2.8. Domaines d'application

L'apprentissage profond propose un large éventail d'applications dans plusieurs secteurs et domaines. Inclus parmi les applications les plus courantes :

- **Classification de texte**

  La classification de texte est une tâche fondamentale du traitement du langage naturel (NLP), qui consiste à attribuer des documents textuels à des catégories ou classes prédéfinies. Elle est utilisée dans plusieurs domaines comme le filtrage de spam, l'analyse de sentiments, la classification thématique, etc.

L'objectif est d'attribuer de manière précise une étiquette à chaque texte en fonction de son contenu

Modèles utilisés : BERT (Bidirectional Encoder Representations from Transformers), LSTM (Long Short-Term Memory) [15]

- **Traduction automatique**

  La traduction automatique est le processus de traduction automatique du texte d'une langue à une autre, tout en maintenant simultanément sa signification d'origine. Cette tâche joue un rôle clé dans la communication multilingue, l'emplacement du contenu et les activités commerciales internationales.

  Modèles utilisés : Transformer, NMT (Neural Machine Translation – Traduction automatique neuronale) [16]

- **Classification de documents**

  La classification des documents est la classification de l'ensemble du document en catégories prédéfinies en fonction de son contenu, Cette tâche est particulièrement utile pour organiser et gérer de grandes collections de documents, faciliter la recherche d'informations et améliorer les systèmes de recommandation de contenu.

  Modèles utilisés : Doc2Vec, HAN (Hierarchical Attention Networks – Réseaux hiérarchiques à mécanisme d'attention) [12]

II.3. Les réseaux de neurones

Les réseaux neuronaux artificiels, constituant une branche de l'apprentissage automatique, sont à la base des algorithmes d'apprentissage profond. Inspirés du fonctionnement biologique des neurones, ils sont organisés en couches successives : une couche d'entrée, une ou plusieurs couches cachées, et une couche de sortie. Chaque nœud, ou neurone artificiel, est associé à un poids et un seuil d'activation. Lorsqu'un nœud dépasse ce seuil, il transmet son signal à la couche suivante ; dans le cas contraire, aucune information n'est transmise.

Grâce aux données d'entraînement, ces réseaux ajustent progressivement leurs paramètres pour améliorer leur précision. Une fois optimisés, ils deviennent des outils performants pour le classement et le regroupement automatique de données.[13]

II.3.1. Les principaux types de réseaux de neurones

Voici quelque type de modèles, présentés dans l'ordre approximatif de leur développement. Chaque modèle améliore les limites d'un modèle précédent.

- **Les réseaux de neurones convolutifs (CNN)**

Les CNNs sont des réseaux de neurones spécialisés en vision par ordinateur et en classification d'images. Ils fonctionnent en extrayant progressivement des caractéristiques d'une image à travers différentes couches convolution, pooling et couches entièrement connectées. Ils surpassent les méthodes traditionnelles d'extraction manuelle de caractéristiques, offrant ainsi une approche plus rapide et évolutive pour la reconnaissance d'objets et le traitement d'images.

Malgré leurs nombreux avantages en termes de précision et d'efficacité, les CNNs ont aussi des limites : gourmands en calculs, coûteux en ressources (besoin de GPU puissants) et nécessitant des experts pour l'optimisation des hyper paramètres et la configuration du réseau.[13]

- **Les réseaux de neurones récurrents (RNN)**

Les RNNs sont des réseaux neuronaux adaptés aux données séquentielles et temporelles, notamment en traitement du langage naturel et en reconnaissance vocale. Contrairement aux réseaux classiques, ils utilisent des boucles de rétroaction pour mémoriser les informations passées et influencer la sortie actuelle, ils sont largement utilisés dans des applications comme la traduction automatique, la reconnaissance vocale et la prédiction des séries chronologiques (ex : marché boursier). [13]

- **Les réseaux antagonistes génératifs (GANs)**

Les réseaux antagonistes génératifs (GANs) sont des réseaux neuronaux utilisés en intelligence artificielle (IA) et dans d'autres domaines pour générer de nouvelles données similaires aux données d'entraînement originales. Par exemple, ils peuvent créer des images de visages humains qui n'existent pas réellement. Le terme "antagoniste" provient de l'interaction entre les deux parties du GAN : un générateur et un discriminateur. le générateur produit des contenus, tels que des images, vidéos ou sons, en ajoutant une transformation. Par exemple, un cheval peut être transformé en zèbre avec un certain degré de précision. La qualité du résultat dépend des données d'entrée et du niveau d'entraînement du modèle génératif. Le discriminateur joue le rôle d'adversaire. Il compare les résultats générés (fausses images) avec les vraies images du jeu de données et tente de distinguer les vraies des fausses. [13]

II.4. Modèle d'apprentissage profond LSTM

Le modèle d'apprentissage profond employé pour la résolution du problème de détection des faux comptes dans les réseaux sociaux est le réseau LSTM. Dans ce qui suit, une description détaillée de ce modèle avancé sera présentée :

II.4.1. Fonctionnement et architecture du LSTM

Les réseaux LSTM permettent de supprimer ou d'ajouter des conditions cellulaires. Ce processus est contrôlé par une structure appelée porte. La porte a remis les informations. Ils sont constitués de couches sigmoïdes neuronales et de chirurgie de multiplication ponctuelle, par conséquent, lors du passage de RNN à LSTM, nous introduisons de plus en plus de mécanismes de contrôle pour réguler le mélange des rivières et des entrées en fonction des poids apprises, par conséquent, les réseaux LSTM ont un impact plus important et entraînent de meilleurs résultats. Cependant, cela comporte plus d'ambiguïté et de coûts opérationnels. [14]

Dans l'architecture du LSTM, on distingue les trois composants suivants [17] :

1. **L'état de la cellule,** qui conserve la mémoire à long terme du réseau
2. **L'état caché**, qui contient les informations issues du temps précédent
3. Et **les données d'entrée** correspondant à l'instant présent

Les réseaux LSTM sont basés sur une cellule de mémoire, régulée par trois types de portes : la porte d'entrée, la porte d'oubli et la porte de sortie. Chaque porte joue un rôle clé en décidant quelles informations doivent être ajoutées, supprimées ou extraites de la cellule de mémoire.

1. **Porte d'entrée :** Elle détermine quelles nouvelles informations doivent être stockées dans la cellule de mémoire.
2. **Porte d'oubli :** Elle choisit quelles informations existantes doivent être effacées de la cellule de mémoire.
3. **Porte de sortie :** Elle contrôle quelles informations doivent être extraites et utilisées à partir de la cellule de mémoire.

Grâce à ce mécanisme, les LSTM sont capables de conserver ou d'éliminer de manière sélective des informations au fil du temps, ce qui leur permet d'apprendre et de gérer des dépendances sur de longues périodes. Le réseau maintient également un état

caché, qui représente sa mémoire à court terme, mis à jour à chaque étape à l'aide de l'entrée actuelle, de l'état caché précédent et de l'état de la cellule de mémoire. [18]

II.4.4. Avantages et inconvénients du LSTM

Le modèle LSTM comme tout autre modèle d'apprentissage à des avantages et des inconvénients, parmi ceux-ci :

- **Avantages**

Le réseau de neurones LSTM parvient à atténuer le problème de dispersion du gradient en intégrant des mécanismes supplémentaires appelés portes : porte d'entrée, porte de sortie et porte d'oubli. Ces portes offrent au réseau une capacité de mémoire étendue ainsi qu'un meilleur contrôle sur l'impact des expériences passées sur l'ajustement des poids. Autrement dit, grâce à sa capacité à réguler son propre processus d'apprentissage, un LSTM peut éviter que les gradients ne deviennent excessivement grands ou trop faibles, ce qui lui permet de surpasser un RNN classique.[19]

- **Inconvénients**

Les LSTM présentent néanmoins plusieurs limitations. Leur structure complexe, incluant diverses portes et cellules de mémoire, engendre des coûts computationnels élevés, ainsi qu'une augmentation du temps d'entraînement et de l'utilisation de la mémoire. Malgré les progrès réalisés par rapport aux RNN classiques, ils rencontrent toujours des difficultés à traiter de très longues séquences, en particulier pour capturer les dépendances sur le long terme. Le risque de sur apprentissage est également important, surtout lorsque les données d'apprentissage sont limitées. Par ailleurs, les LSTM sont souvent critiqués pour leur faible interopérabilité, rendant difficile l'explication de leurs prédictions. Leur mise en œuvre est aussi compliquée par l'instabilité lors de l'entraînement, la forte sensibilité aux hyperparamètres et les limitations liées à la parallélisassions.[20]

II.4.5. Applications des réseaux LSTM

Parmi les utilisations les plus répandues des réseaux LSTM il y a :

- La traduction automatique. Leur aptitude à considérer le contexte global d'une séquence leur permet de saisir les dépendances à long terme entre les mots et de produire des traductions plus fidèles.

- Les LSTM sont également employés pour la génération créative de texte, en s'appuyant sur de vastes ensembles de données pour élaborer de nouvelles phrases à la fois cohérentes et pertinentes.
- En outre, les réseaux LSTM ont démontré leur efficacité dans des domaines comme la reconnaissance vocale. Leur capacité à modéliser les dépendances temporelles complexes des signaux audios leur permet de générer des transcriptions précises, ce qui s'avère particulièrement utile dans des secteurs tels que la transcription médicale, l'assistance vocale et la commande vocale d'appareils électroniques.
- Enfin, les LSTM sont largement exploités pour des tâches de prédiction et de classification de séquences, telles que la prévision de séries temporelles, l'analyse des sentiments dans les textes, la détection d'anomalies, entre autres. Leur aptitude à capturer des schémas complexes et à modéliser des relations à long terme en fait des outils extrêmement polyvalents, capables de s'adapter à de nombreux types de problèmes. [21]

II.5. Conclusion

Ce chapitre a permis d'établir les bases conceptuelles nécessaires à la compréhension du projet. En clarifiant les différences entre intelligence artificielle, apprentissage automatique et apprentissage profond, et en présentant les différentes approches d'apprentissage ainsi que les types de réseaux neuronaux pertinents, nous avons posé le cadre théorique indispensable à l'élaboration d'un modèle performant. Le modèle LSTM, en raison de sa capacité à traiter des données séquentielles et temporelles, se révèle particulièrement adapté au problème de détection de faux comptes sur les réseaux sociaux. La compréhension approfondie de ces concepts permettra dans les chapitres suivants de mieux justifier les choix méthodologiques et techniques adoptés pour la conception, l'entraînement et l'évaluation de notre système de détection automatisé.

Chapitre 3 : Conception du système

III.1 Introduction

La détection de faux profils sur les réseaux sociaux représente un enjeu crucial dans la lutte contre la désinformation, les fraudes et les manipulations numériques. Instagram, en tant que plateforme visuelle et fortement utilisée à des fins commerciales et d'influence, est particulièrement exposé à la prolifération de comptes automatisés ou frauduleux. Afin de contribuer à la détection de ces comportements suspects, ce travail s'appuie sur l'élaboration et l'entraînement d'un modèle de classification fondé sur les réseaux de neurones à mémoire courte à long terme (LSTM), réputés pour leur capacité à modéliser des séquences de données complexes. Cette étude détaille l'ensemble du processus méthodologique, de la collecte des données à la sauvegarde du modèle, en passant par les étapes de prétraitement, de conception, d'entraînement et d'évaluation, avec un accent particulier sur l'adaptation des données au format requis par l'architecture LSTM.

III.2 Présentation de l'architecture fonctionnelle du système

Afin de concevoir un système fiable de détection des faux comptes Instagram, une démarche méthodologique rigoureuse a été adoptée, articulée autour de plusieurs étapes successives. Celles-ci englobent la préparation des données, la structuration des entrées pour l'apprentissage profond, la modélisation à l'aide d'un réseau LSTM, ainsi que l'évaluation et la sauvegarde du modèle. Chaque phase contribue de manière complémentaire à la performance et à la robustesse globale du système, comme détaillé ci-après.

A. Collecte et préparation des données

L'objectif principal de cette étape est de fournir des données pertinentes et structurées en vue d'assurer l'entraînement et l'évaluation du modèle de classification.
À cet effet, un jeu de données simulé a été élaboré, représentant l'activité de 577 comptes Instagram. Chaque compte est décrit à travers plusieurs types d'informations

Jeu de données utilisé

Le jeu de données utilisé dans ce projet est un ensemble simulé représentant l'activité de **577** comptes Instagram. Chaque enregistrement correspond à un compte unique et est décrit à travers un ensemble de variables numériques reflétant des caractéristiques comportementales et structurelles, telles que

- le nombre de publications,

- le nombre d'abonnés (followers),
- le nombre de comptes suivis (following),
- des ratios d'activité,
- ou encore d'autres indicateurs dérivés.

La variable cible, nommée fake, est binaire et permet de distinguer les comptes authentiques (0) des comptes frauduleux ou artificiels (1). Ce format rend l'ensemble parfaitement adapté à une tâche de classification binaire supervisée

La première étape de notre démarche consiste à importer un ensemble de données structuré au format CSV, permettant une intégration flexible avec les bibliothèques Python. La base contient des caractéristiques numériques représentant des comptes Instagram, avec une variable cible binaire nommée fake, indiquant si un compte est faux (1) ou authentique (0).

Les données ont été divisées en deux composantes principales : les variables explicatives (X) et la variable cible (y). Afin de garantir une homogénéité des échelles et d'accélérer la convergence du modèle d'apprentissage, un redimensionnement a été appliqué à l'aide du MinMaxScaler, qui transforme les valeurs dans l'intervalle [0, 1]

B. Division et équilibrage des données

Les données ont ensuite été réparties en un ensemble d'apprentissage (80 %) et un ensemble de test (20 %), en veillant à préserver la proportion des classes grâce à la stratification. Étant donné le déséquilibre inhérent aux jeux de données liés à la détection de comportements frauduleux, un ré échantillonnage a été effectué à l'aide de la méthode SMOTE cette technique permet la génération synthétique de nouvelles instances de la classe minoritaire, contribuant ainsi à améliorer la robustesse du modèle

C. Mise en forme des données pour le modèle LSTM

Les réseaux de neurones à mémoire courte à long terme (**LSTM**) exigent une structuration tridimensionnelle des données en (nombre d'échantillons, pas de temps, nombre de caractéristiques). Par conséquent, les matrices d'entrée ont été remodelées pour être compatibles avec cette exigence.

D. Conception du modèle LSTM

Le modèle a été développé à l'aide de l'API Séquentiel de Keras, et comprend les couches suivantes :

- Une couche **LSTM** avec 64 neurones, destinée à capturer les dépendances temporelles et structurelles entre les différentes variables du compte.
- Une couche de Dropout fixée à 30 %, permettant d'atténuer le surapprentissage en désactivant aléatoirement certaines connexions durant l'apprentissage.
- Une couche Dense intermédiaire (32 neurones, activation ReLU).
- Une couche de sortie sigmoïde, adaptée à la classification binaire.

La fonction de coût utilisée est l'entropie croisée binaire, optimisée à l'aide de l'algorithme Adam. Le critère d'évaluation retenu est l'exactitude **(**accuracy**)**

E. Entraînement et validation

L'entraînement a été réalisé sur un nombre maximum de 50 époques, avec une **validation croisée** interne sur 20 % des données d'entraînement. Un mécanisme d'arrêt anticipé (EarlyStopping) a été intégré afin d'interrompre l'apprentissage lorsque la performance de validation cessait de s'améliorer pendant cinq époques consécutives. Cette stratégie contribue à éviter le sur apprentissage.

F. Évaluation du modèle

L'évaluation du modèle constitue une étape déterminante pour juger de sa capacité à généraliser sur des données inédites. Dans le cadre de cette étude, le modèle LSTM a été testé sur l'ensemble de test (20 % des données), jamais observé lors de l'apprentissage. Plusieurs métriques issues de la classification binaire ont été mobilisées pour une analyse fine de la performance :

- **Précision globale (accuracy)**

Exactitude (Accuracy) : mesure la proportion globale de prédictions correctes. Elle indique dans quelle mesure le modèle classe correctement l'ensemble des instances, toutes classes confondues.

Accuracy (exactitude) = (VP + VN) / (VP + FN + FP + VN)

- **Rapport de classification** : comprenant la précision, le rappel, et la F-mesure

  Précision (Precision) : indique, parmi les instances prédites comme positives (faux comptes), combien sont réellement positives. Elle évalue la fiabilité des prédictions positives.

Precision = VP / (VP + FP)

Rappel (Recall ou Sensibilité) : mesure la capacité du modèle à identifier correctement les vrais positifs (faux comptes réellement détectés).

Recall (rappel) = VP / (VP + FN)

F-mesure (F1-score) : combine la précision et le rappel en une moyenne harmonique, utile notamment en cas de déséquilibre entre les classes.

F1-score = 2 × (Precision × Recall) / (Precision + Recall)

- **Matrice de confusion** : permettant de visualiser la distribution des prédictions correctes et erronées

Tableau III. 1.Les différentes catégories pour une matrice de confusion

Par ailleurs, l'évolution de la fonction de perte et de la précision au fil des époques a été représentée graphiquement afin d'évaluer la dynamique d'apprentissage.

G. III.2.7 Sauvegarde du modèle

	Prédit : Authentique (0)	Prédit : Faux (1)
Réel : Faux (1)	Vrai Positif (VP)	Faux Négatif (FN)
Réel : Authentique (0)	Faux Positif (FP)	Vrai Négatif (VN)

Enfin, le modèle final a été sauvegardé au format HDF5 (.h5), facilitant ainsi sa réutilisation dans un cadre applicatif ou pour une phase de déploiement.

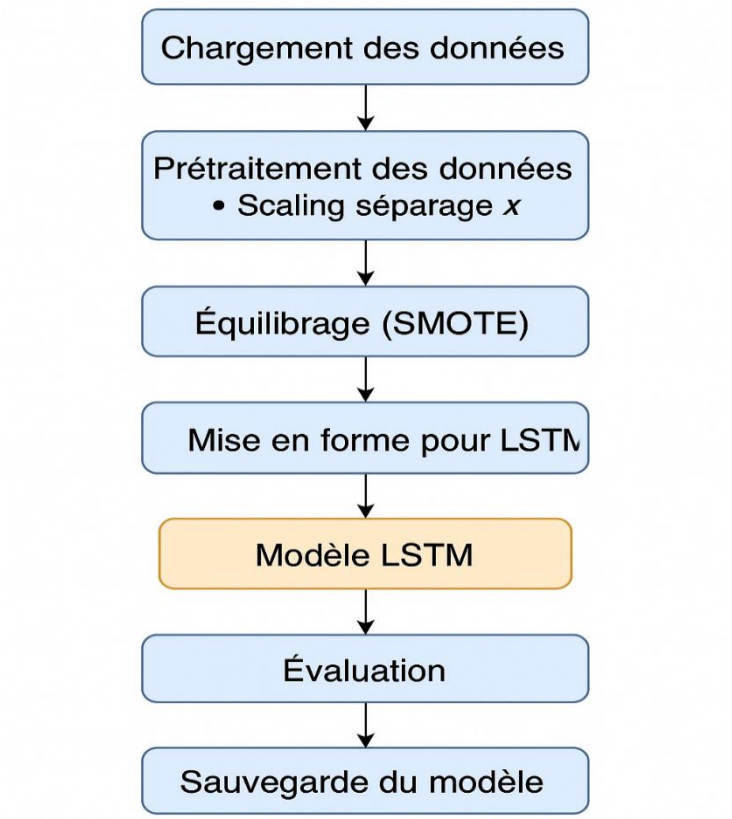


Figure III. 1.L'architecture générale du système

III.3 Architecture et mécanismes du modèle LSTM

Le réseau de mémoire à long terme (LSTM) démontre une efficacité notable dans l'analyse des données tabulaires décrivant les caractéristiques des comptes Instagram, grâce à sa capacité à apprendre des relations non linéaires complexes entre les différentes variables. Bien que ce type de réseau soit initialement conçu pour traiter des données séquentielles, il a été adapté dans ce travail aux données tabulaires en les remodelant sous forme de tenseurs tridimensionnels représentant une séquence de caractéristiques. Les mécanismes internes de gating — tels que les portes d'oubli, d'entrée et de sortie — permettent de réguler le flux d'informations au sein de la cellule neuronale, favorisant la rétention des informations pertinentes et l'élimination des données non significatives, réduisant ainsi les problèmes de disparition ou d'explosion du gradient typiques des réseaux neuronaux classiques. Par ailleurs, l'utilisation de techniques comme SMOTE pour l'équilibrage des classes et la normalisation préalable des données a renforcé la stabilité du modèle pendant l'apprentissage. Globalement, le modèle LSTM constitue un cadre flexible

et puissant pour la détection des faux comptes en exploitant les dynamiques internes des données tabulaires et en analysant les motifs cachés entre les attributs.

III.3. Conclusion

Le processus de détection des faux comptes Instagram, tel que mis en œuvre dans cette étude, démontre l'efficacité des réseaux LSTM lorsqu'ils sont correctement configurés et alimentés par des données pertinentes et équilibrées. Grâce à une préparation rigoureuse des données, à l'utilisation de techniques d'optimisation comme le redimensionnement et le suréchantillonnage, ainsi qu'à l'intégration de mécanismes de régularisation et de validation, le modèle obtenu présente des performances satisfaisantes en classification binaire. La sauvegarde du modèle au format HDF5 offre, par ailleurs, une base solide pour son intégration future dans des systèmes automatisés de détection de fraudes. Ce travail constitue ainsi une contribution concrète à l'application de l'apprentissage profond dans le domaine de la cybersécurité sur les réseaux sociaux.

Chapitre IV :
Implémentation et résultat

IV.1 Introduction

Ce chapitre présente l'environnement et les outils de développement utilisés pour la mise en œuvre du système de classification des comptes Instagram, visant à distinguer les comptes réels des faux. Il détaille les spécifications matérielles, les langages de programmation, les logiciels, ainsi que les librairies employées, notamment Python, TensorFlow, et PyTorch. Les étapes de prétraitement des données, la méthodologie de construction et d'entraînement des modèles (y compris le LSTM), ainsi que les techniques d'évaluation des performances sont également exposées. Enfin, une analyse comparative des résultats obtenus avec différentes méthodes de classification est réalisée, mettant en lumière les avantages et limites de chaque approche.

IV.2 Environnement et outils de développement

IV.2.1 Matériels utilisés

Le tableau ci-dessous présente les spécifications et les performances des équipements utilisés pour la mise en œuvre du système développé

Tableau IV. 1.Caractéristiques du matériel

Poste de travail N°01	Caractéristiques
Dell	PC
Windows 10 Professionnel	Système d'exploitation
Intel(R) Core(TM) i3-6006U CPU @ 2.00GHz 2.00 GHz	Processeur
4,00 Go	RAM
SE 64 bits	Type de système

IV.2.2 Langages, logiciels et librairies utilisés

IV.2.2.1 Langage de programmation

Langage Python

Python est un langage informatique régulièrement employé pour la création de sites internet et de programmes, l'automatisation de tâches ainsi que l'analyse de données.

Python présente une grande flexibilité, ce qui lui permet d'être utilisé pour une vaste gamme de projets et ne se concentre pas sur un type de problème unique. [19]

La clarté et l'harmonie de Python diminuent la difficulté d'apprentissage pour les développeurs débutants et facilitent le passage à des projets de machine learning avancés.

IV.2.2.2 Environnement de développement

Nous avons eu recours à deux environnements de développement pour réaliser ce projet :

Visual Studio Code - Dans ce projet, nous avons exploité l'outil Visual Studio Code (VS Code), qui se présente comme un éditeur de code source extrêmement flexible, léger et facilement adaptable. Il fournit une expérience de développement enrichissante et efficace pour divers types de projets logiciels, allant du développement web à l'informatique en nuage et même à la science des données.

Google Colaboratory - Colab est un service Jupyter Notebook hébergé qui ne nécessite aucune configuration et offre un accès gratuit aux ressources de calcul, notamment aux GPU et aux TPU. Colab est particulièrement adapté à l'apprentissage automatique, à la science des données et à l'éducation. [24]

IV.2.2.3 Les bibliothèques importées

Tensorflow - TensorFlow est une plateforme de code source ouvert créée pour le traitement numérique, l'intelligence artificielle à grande échelle, l'apprentissage profond, ainsi que d'autres missions d'analyse statistique et de prévision. Ce type de technologie permet aux programmeurs de déployer des modèles d'intelligence artificielle de manière plus rapide et simplifiée, car elle rend l'obtention des données, le déploiement des prédictions à grande échelle, et l'optimisation des résultats futurs plus accessibles. [25]

Pandas - Pandas est une librairie Python conçue pour traiter des collections de données, elle offre des outils pour examiner, assainir, découvrir et gérer les données, le terme "Pandas" se réfère à "Panel Data" et à "Python Data Analysis", et a été développée par Wes McKinney en 2008. [26]

Numpy - NumPy est une bibliothèque open source qui se concentrent sur le calcul mathématique et scientifique pour la programmation en Python. Son nom est dérivé de Numerical Python. Cette bibliothèque propose un large éventail de fonctions

mathématiques avancées, englobant le support pour des tableaux multidimensionnels, des tableaux masqués ainsi que des matrices. [27]

Matplotlib - Matplotlib est une bibliothèque de création de graphiques et de visualisation de données accessible sur plusieurs plateformes, notamment pour Python et son extension numérique NumPy. En tant que telle, elle représente une option open source intéressante en remplacement de MATLAB. Les développeurs ont aussi la possibilité d'utiliser les API de Matplotlib pour insérer des visualisations dans des applications dotées d'interfaces graphiques. [28]

Scikit-learn (sklearn) - Scikit-learn est sans doute la bibliothèque la plus précieuse pour la machine learning sous Python. La bibliothèque sklearn offre une variété de ressources efficaces pour l'apprentissage automatique et l'analyse statistique, en particulier pour des tâches telles que la classification, la régression, le regroupement et la réduction de dimension. [29]

Torch (PyTorch) - PyTorch est un framework d'apprentissage automatique (ML) open source basée sur le langage de programmation Python et la bibliothèque Torch. Torch est une bibliothèque ML open source utilisée pour créer des réseaux neuronaux profonds et écrite en langage de script Lua. C'est l'une des plateformes privilégiées pour la recherche en apprentissage profond.

Seaborn - Seaborn est une puissante bibliothèque de visualisation de données qui offre de nombreuses possibilités de personnalisation de l'apparence des graphiques. La personnalisation des graphiques Seaborn est essentielle pour créer des visualisations pertinentes et visuellement attrayantes.

IV.3. Base de données utilisées

IV.3.1. Informations générales sur l'ensemble des données

L'exploration de l'ensemble de données constitue une étape cruciale dans tout projet de science des données, car elle permet de mieux comprendre la structure, la qualité et les caractéristiques des variables disponibles avant de procéder à la phase de modélisation.

Dans ce projet, nous analysons un ensemble de données extrait d'Instagram, contenant diverses informations sur l'activité et le profil des utilisateurs. Chaque ligne

représente un compte Instagram, et l'objectif est de prédire si ce compte est réel ou faux à partir de plusieurs indicateurs comportementaux et structurels.

IV.4.1. Préparation des données pour la classification

Méthodologie de prétraitement et préparation du jeu de données

Dans le cadre du développement d'un système de classification binaire visant à distinguer les comptes Instagram authentiques des comptes frauduleux, une série d'étapes de préparation des données a été mise en œuvre afin d'assurer leur qualité et leur pertinence pour l'apprentissage automatique.

Dans un premier temps, un nettoyage initial du jeu de données a été réalisé afin de s'assurer de l'absence de valeurs manquantes ou aberrantes. Les attributs jugés non informatifs ou redondants ont été éliminés pour optimiser la représentation des données. Les variables catégorielles ont ensuite été encodées sous forme numérique afin d'être compatibles avec les modèles d'apprentissage profond.

Une étape de normalisation a ensuite été appliquée aux variables continues (telles que le nombre de publications, de followers ou encore le ratio numérique du nom d'utilisateur), afin d'homogénéiser les échelles et d'éviter qu'une variable à forte amplitude n'influence excessivement le processus d'apprentissage.

Enfin, le jeu de données a été scindé en deux sous-ensembles : un ensemble d'apprentissage et un ensemble de test, permettant une évaluation rigoureuse et objective des performances du modèle LSTM. Ce processus de préparation garantit la qualité des données en vue de leur exploitation optimale dans le cadre de la classification binaire.

Analyse de la distribution des classes

L'analyse de la distribution des étiquettes est une étape essentielle dans l'étude d'un ensemble de données supervisé, car elle permet de détecter d'éventuels déséquilibres susceptibles d'influencer les performances du modèle. Contrairement à d'autres bases comme celles de Reddit ou Twitter qui présentent souvent un fort déséquilibre entre classes positives et négatives, l'ensemble de données utilisé dans ce projet – composé de profils Instagram – affiche une distribution parfaitement équilibrée entre comptes réels et faux comptes.

Plus précisément, le jeu de données contient un total de 576 comptes, répartis équitablement entre :

- 288 comptes réels (étiquette = 0),
- 288 faux comptes (étiquette = 1).

Cette répartition est illustrée dans le tableau suivant :

Tableau IV. 2.Distribution des classes de l'ensemble de données

Classe	Nombre de comptes	Pourcentage (%)
Compte réel (0)	288	50.0 %
Faux Compte (1)	288	50.0 %

La figure suivante représente Répartition proportionnelle des profils authentiques et faux.

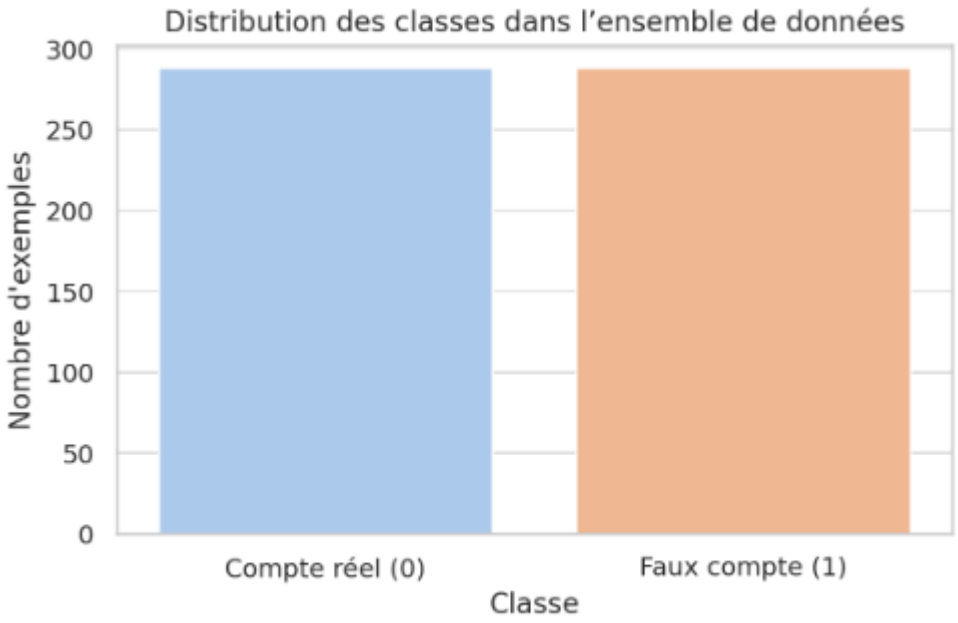


Figure IV. 1.Distribution des classes de données

La figure suivante Proportions des comptes réels et faux dans l'ensemble de données Instagram.

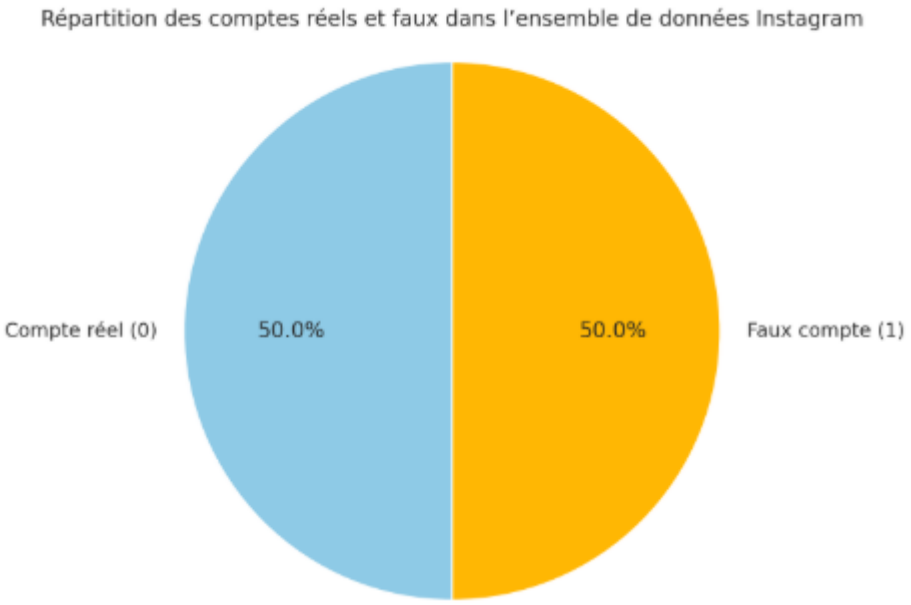


Figure IV. 2.Fréquences des comptes selon leurs classes

IV.5. Résultats et comparaison

IV.5.1. Résultats obtenus

A. Méthodes de base

Tableau IV. 3.Les résultats de performance des méthodes de base

Modèle	Accuracy	Precision (0)	Recall(0)	F1_Score (0)	Precision (1)	Recall(1)	F1_Score (1)
Random Forest	0.93	0.92	0.96	0.94	0.95	0.90	0.92
K-NN	0.87	0.86	0.91	0.89	0.89	0.82	0.86
Decision Tree	0.87	0.88	0.88	0.88	0.86	0.86	0.86
Naive Bayes	0.69	0.95	0.45	0.61	0.60	0.97	0.75

B. Apprentissage profond LSTM

Tableau IV. 4.Les résultats de performance de modèle LSTM.

Modèle	Accuracy	Precision(0)	Recall(0)	F1_Score (0)	Precision(1)	Recall(1)	F1_Score (1)
LSTM	0.87	0.81	0.97	0.88	0.96	0.78	0.86

IV.6. Analyse des résultats

IV.6.1. Évaluation des différents Classificateurs

1.Random Forest

Il se distingue par les meilleures performances globales avec une précision de 0,93. Il affiche un excellent équilibre entre précision et rappel pour les deux classes : *une* precision(0) de 0,92 et un recall(0) de 0,96 pour les vrais comptes, et une precision(1) de 0,95 et un recall(1) de 0,90 pour les faux comptes. Le score F1, synthèse des deux, est respectivement de 0,94 (classe 0) et 0,92 (classe 1), témoignant d'un modèle robuste.

Avantage :

- Robuste au bruit et aux données manquantes.
- Moins sujet au surajustement qu'un Decision Tree seul.

Inconvénients :

- Plus complexe et moins interprétable qu'un arbre de décision simple.

2. K-NN (K Plus Proches Voisins)

Le modèle K-NN atteint une précision globale de 0,87, avec des performances légèrement inférieures à celles du Random Forest. Il reste toutefois équilibré avec un *F1_Score(0)* de 0,89 et un *F1_Score(1)* de 0,86.

Avantages :

- Simple à comprendre et à implémenter.
- Pas d'hypothèse forte sur la distribution des données.

Inconvénients :

- Sensible aux données déséquilibrées.
- Nécessite un choix optimal de "k" (trop petit → bruit, trop grand → sous-performance).

3. Arbre de décision (decision Tree)

L'arbre de décision présente une performance équivalente à celle du K-NN avec une précision de 0,87. Ses scores de précision, rappel et F1 sont similaires pour les deux classes, ce qui indique un comportement cohérent mais sans exceller.

Avantages :

- Très interprétable (règles claires).
- Rapide à entraîner.

Inconvénients :

- Surajustement fréquent sans réglage (élagage, limitation de profondeur).
- Sensible aux petites variations dans les données.

4. Naive Bayes

Le Naive Bayes affiche les performances les plus faibles, avec une précision globale de seulement 0,69. Bien que ce modèle soit performant dans la détection des faux profils (recall(1) = *0,97*), il présente un déséquilibre marqué avec une faible capacité à bien identifier les vrais profils (recall(0) = 0,45), ce qui se traduit par un F1_Score(0) de seulement 0,61.

Avantages :

- Très rapide à entraîner et à prédire.
- Fonctionne bien avec des données textuelles (ex : filtrage de spam).

Inconvénients :

- Hypothèse d'indépendance rarement vérifiée en pratique.
- Peu performant sur les données complexes ou fortement corrélées.

En résumé les résultats de cette étude comparative démontrent que le choix d'un algorithme de classification doit être guidé par les exigences spécifiques du problème considéré. Le Random Forest se distingue comme l'approche la plus performante, offrant un équilibre optimal entre précision et rappel, ce qui en fait la solution privilégiée pour des tâches nécessitant une classification précise. Si l'interprétabilité des décisions est un critère primordial, l'**a**rbre de décision constitue une alternative pertinente, bien que ses performances soient légèrement inférieures. Le Naive Bayes, en raison de sa rapidité

d'exécution, reste adapté aux applications en temps réel ou au traitement de données textuelles, malgré une précision globalement moindre.

Enfin, la méthode K-NN représente une option simple à mettre en œuvre, à condition que les données fassent l'objet d'un prétraitement rigoureux pour en optimiser l'efficacité. Ainsi, la sélection finale du modèle doit s'appuyer sur une analyse approfondie des contraintes opérationnelles (temps de calcul, complexité) et des objectifs prioritaires (précision, interprétabilité ou rapidité).

IV.6.1.1. Courbes ROC et PRC des différents Classificateurs

A. COURBES ROC

L'aire sous la courbe (AUC) est une métrique synthétique permettant de quantifier la performance globale de chaque classificateur

Random Forest (AUC = 0.99) - Le modèle Random Forest affiche la meilleure performance avec une AUC de 0.99. Sa courbe ROC se rapproche fortement du coin supérieur gauche du graphique, ce qui reflète une excellente capacité à distinguer les faux comptes des vrais. Ce résultat est cohérent avec les propriétés du Random Forest, qui combine plusieurs arbres de décision pour réduire le surapprentissage et améliorer la généralisation

Naive Bayes (AUC = 0.95) - Le modèle Naive Bayes démontre une performance remarquable malgré sa simplicité. Son AUC de 0.95 indique une bonne séparation entre les deux classes, ce qui peut s'expliquer par l'hypothèse d'indépendance conditionnelle souvent bien adaptée aux données textuelles ou comportementales typiques des réseaux sociaux

K-Nearest Neighbors (AUC = 0.94) - Le modèle KNN, avec une AUC de 0.94, présente également de bonnes performances. Toutefois, sa courbe est légèrement inférieure à celle du Random Forest, ce qui peut refléter une sensibilité aux valeurs aberrantes ou au déséquilibre des classes

Decision Tree (AUC = 0.92) - Le modèle d'arbre de décision est celui qui montre les performances les plus modestes parmi les modèles testés, avec une AUC de 0.92. Bien que correcte, cette performance peut souffrir de l'overfitting typique des arbres non élagués

B. Courbes PRC

La courbe Precision-Recall (PRC) permet d'évaluer la performance d'un modèle en termes de précision (taux de vrais positifs parmi les positifs prédits) et de rappel (taux de vrais positifs détectés parmi les positifs réels). Contrairement à la courbe ROC, la PRC est plus informative lorsque les classes sont déséquilibrées — ce qui est souvent le cas dans la détection de comptes frauduleux, où les vrais comptes sont largement majoritaires.

Random Forest (AP = 0.98) - Le modèle Random Forest obtient le score le plus élevé en AP (0.98), confirmant sa supériorité observée dans l'analyse ROC. Il maintient une précision élevée sur l'ensemble du spectre de rappel, ce qui signifie qu'il est capable de détecter un grand nombre de faux comptes tout en commettant très peu d'erreurs. Cela en fait un outil extrêmement fiable dans des contextes de sécurité des réseaux sociaux, notamment pour la détection automatique sans supervision humaine continue.

Naive Bayes (AP = 0.95) - Avec un AP de 0.95, Naive Bayes démontre une grande efficacité malgré son modèle simplifié basé sur l'indépendance conditionnelle. La courbe montre une forte précision même à des niveaux élevés de rappel**,** ce qui signifie que le modèle identifie correctement la majorité des faux comptes sans générer un trop grand nombre de faux positifs.

KNN (AP = 0.93) - Le modèle KNN affiche une AP de 0.93, ce qui indique une performance solide. Sa courbe reste relativement stable jusqu'à des niveaux de rappel élevés, mais commence à **chuter en précision** lorsque le rappel dépasse 0.8, ce qui suggère une augmentation des faux positifs à mesure qu'on tente de capturer tous les cas.

Arbre de décision (AP = 0.88) - Le modèle à base d'arbre de décision obtient une AP de 0.88, soit la moins bonne performance parmi les quatre. La courbe montre une chute plus rapide de la précision avec l'augmentation du rappel, indiquant que ce modèle commet davantage de faux positifs lorsqu'il tente de maximiser la détection des faux comptes. Cela est cohérent avec la tendance de l'arbre de décision à surajuster les données, en particulier en l'absence de mécanismes de régularisation comme le pruning.

En résumé les résultats obtenus à travers les courbes ROC et Precision-Recall mettent en évidence des écarts significatifs de performance entre les modèles étudiés. Le **Random Forest** se distingue nettement par sa capacité à concilier haute précision et fort rappel, le plaçant comme le choix optimal pour la détection de faux comptes. Le **Naive Bayes**, bien que basé sur un modèle probabiliste simple, démontre une efficacité remarquable, notamment en termes de légèreté computationnelle. Le **KNN**, bien que performant dans une certaine mesure, souffre d'une dégradation de la précision à mesure que le rappel augmente. Enfin, le **Decision Tree** montre des résultats corrects, mais reste plus vulnérable au surapprentissage.

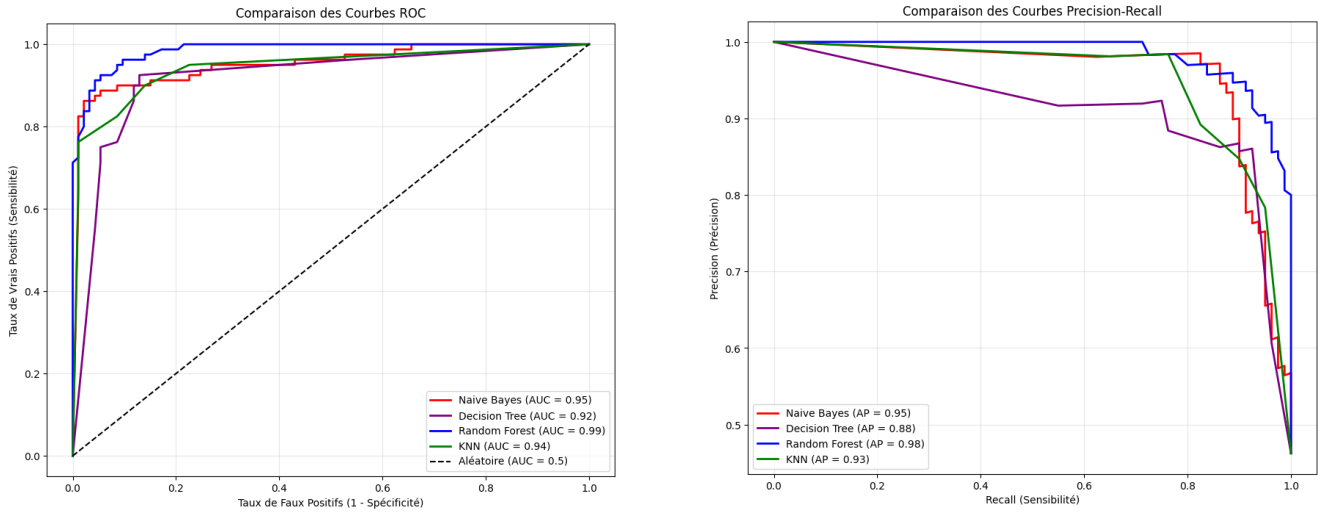


Figure IV. 3.Courbe ROC (gauche) et courbe PRC (droite) des méthodes de base.

La Figure suivante présente la matrice de confusion du modèle random forest, la méthode la plus performante dans l'approche de classification de base :

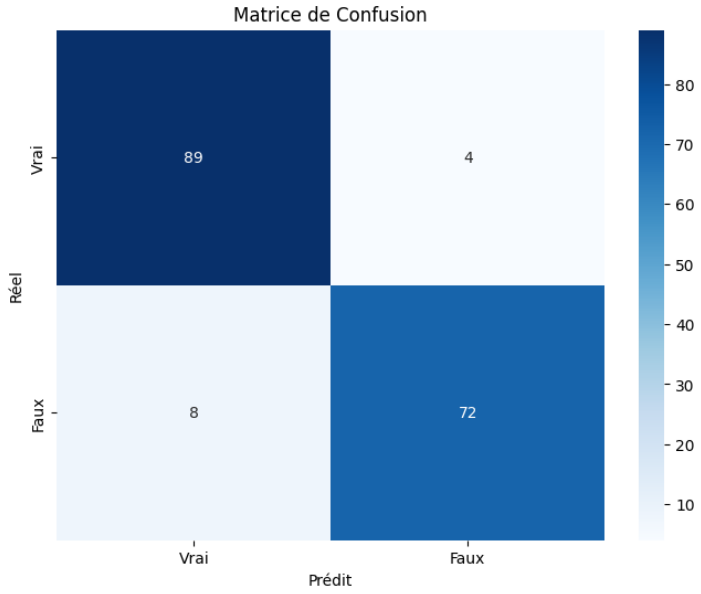


Figure IV. 4.Matrice de confusion du random forest

IV.6.2. Évaluation du modèle LSTM

Les résultats d'évaluations du LSTM sont comme suit :

- Précision : 0.81 pour la classe 0, 0.96 pour la classe 1
- Rappel : 0.97 pour la classe 0, 0.78 pour la classe 1
- F1-Score : 0.88 pour la classe 0, 0.86 pour la classe 1
- Exactitude (Accuracy) : 0.87

1. Exactitude globale (Accuracy)

Une exactitude de 0.87 signifie que le modèle a correctement classé 87 % de l'ensemble des comptes testés.

Ce taux est satisfaisant dans une tâche de classification binaire, mais il doit toujours être interprété avec prudence si les classes sont déséquilibrées.

2. Précision (Precision)

- Pour la classe 0, une précision de 0.81 signifie que 19 % des comptes prédits comme authentiques sont en réalité frauduleux.
- Pour la classe 1, une précision élevée de 0.96 indique que lorsqu'un compte est prédit comme étant frauduleux, le modèle est correct dans 96 % des cas

Le modèle LSTM est très fiable dans la prédiction des faux comptes (faible taux de faux positifs), ce qui est essentiel dans un système de sécurité automatisée pour éviter de sanctionner à tort des utilisateurs légitimes

3. Rappel (Recall)

- Pour la classe 0, un rappel de 0.97 reflète la capacité du modèle à identifier presque tous les comptes authentiques**.**
- Pour la classe 1, un rappel de 0.78 signifie que le modèle détecte 78 % des faux comptes, mais en laisse échapper environ 22 %

Le modèle est très performant pour reconnaître les comptes légitimes, mais il ne détecte pas tous les comptes frauduleux. Cette limitation pourrait affecter l'efficacité globale de la détection des comportements malveillants sur la plateforme.

4. F1-Score

- Les F1-scores de 0.88 (classe 0) et 0.86 (classe 1) montrent que le modèle conserve un équilibre raisonnable entre précision et rappel pour les deux classes

Cela indique une robustesse générale du modèle, sans déséquilibre excessif en faveur d'une classe au détriment de l'autre.

En conséquence le modèle LSTM démontre une excellente précision (0.96) pour la détection des faux comptes, ce qui limite fortement les faux positifs. Il est également le plus performant pour détecter les comptes authentiques (rappel classe 0 = 0.97). Toutefois, son rappel pour les faux comptes (0.78**)** est inférieur à celui des autres modèles, notamment Random Forest (0.90) et Naive Bayes (0.97). Cette faiblesse peut entraîner une sous-détection des fraudes, ce qui est problématique dans les contextes où la couverture maximale des comptes frauduleux est prioritaire. En comparaison, Random Forest émerge comme le modèle le plus équilibré, offrant à la fois une **très** bonne détection des faux comptes et une précision fiable. Ainsi, bien que le LSTM apporte une approche différente en privilégiant la sécurité des comptes authentiques, il ne surpasse pas les modèles classiques en performance globale**.**

IV.6.2.1. Courbe ROC et PRC du modèle LSTM

A. Courbe ROC

AUC de 0.92 : Cette valeur excellente indique que le modèle a une très forte capacité à distinguer entre les classes positives et négatives.

- Une AUC de 1.0 représente une séparation parfaite
- 0.92 se situe dans la plage d'excellente performance (généralement >0.9 est considéré excellent)

a. Courbe PRC

AP (Average Precision) de 0.93 : Cette métrique tout aussi élevée confirme les excellentes performances du modèle, particulièrement pour les problèmes avec déséquilibre de classes.

- L'AP est particulièrement utile quand les classes sont déséquilibrées
- Une valeur de 0.93 indique que le modèle maintient à la fois une haute précision et un bon rappel

Les résultats obtenus, avec une aire sous la courbe ROC (AUC) de **0,92** et une précision moyenne (AP) de **0,93**, attestent d'une performance élevée du modèle en classification. L'AUC élevée reflète une excellente capacité discriminatoire, permettant une séparation nette entre les classes positive et négative. Par ailleurs, la forte valeur d'AP confirme un équilibre robuste entre précision et rappel, particulièrement pertinent dans un contexte potentiellement déséquilibré. Ces indicateurs suggèrent que le modèle est bien calibré, avec un taux de faux positifs maîtrisé, ce qui en fait un outil fiable pour la tâche de prédiction envisagée.

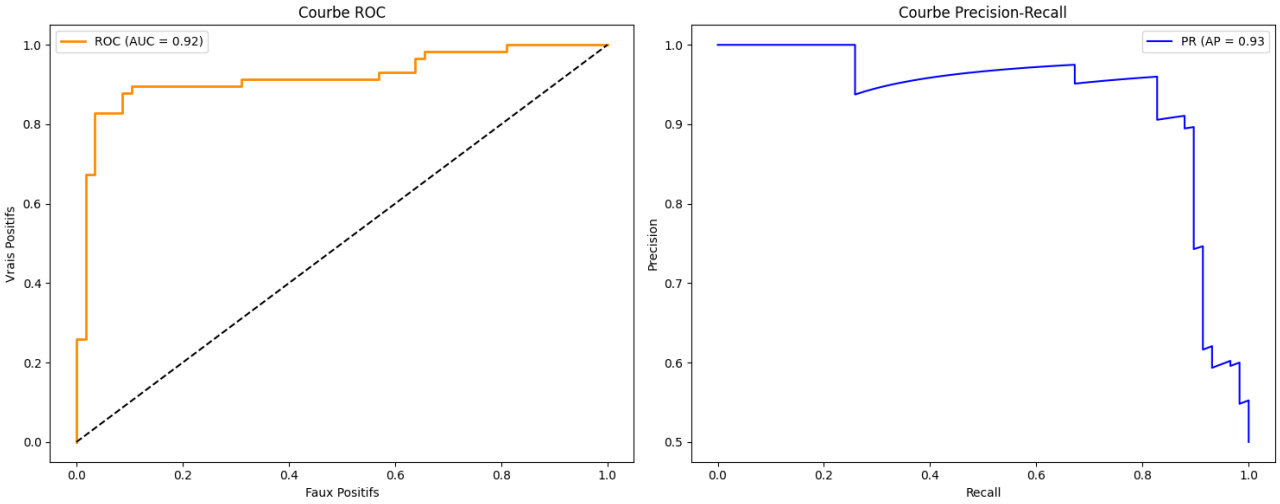


Figure IV. 5.Courbes ROC ET PRC du LSTM

Évolution de la précision

- La précision en validation atteint un niveau élevé (autour de 0,90), indiquant que le modèle généralise bien sur des données non vues.
- La courbe montre une progression stable, sans fluctuations brutales, ce qui suggère un apprentissage bien régularisé.
- L'absence de sur-ajustement (*overfitting*) est visible, car la précision ne chute pas après un certain nombre d'époques.

Évolution de la perte

- Les pertes d'entraînement et de validation diminuent de manière cohérente, confirmant que le modèle apprend efficacement.
- La convergence des deux courbes indique une bonne généralisation, sans divergence marquée qui signalerait un *overfitting*.
- La stabilisation des pertes en fin d'entraînement suggère que le modèle a atteint un optimum, et qu'un nombre d'époques plus élevé n'améliorerait pas significativement les performances.

Les résultats démontrent une convergence stable des métriques d'apprentissage, attestant de l'efficacité du modèle. La précision en validation, atteignant 0.90, reflète une forte capacité de généralisation, tandis que la décroissance simultanée des courbes de perte

(entraînement et validation) indique une optimisation adéquate sans surapprentissage. La stabilisation des performances au-delà de 20 époques suggère que le modèle a atteint un plateau de convergence, confirmant la robustesse de l'architecture et des hyperparamètres choisis. Ces observations valident la pertinence du modèle pour la tâche de classification considérée, tout en laissant entrevoir une possible optimisation marginale via un réglage fin du taux d'apprentissage ou un arrêt précoce.

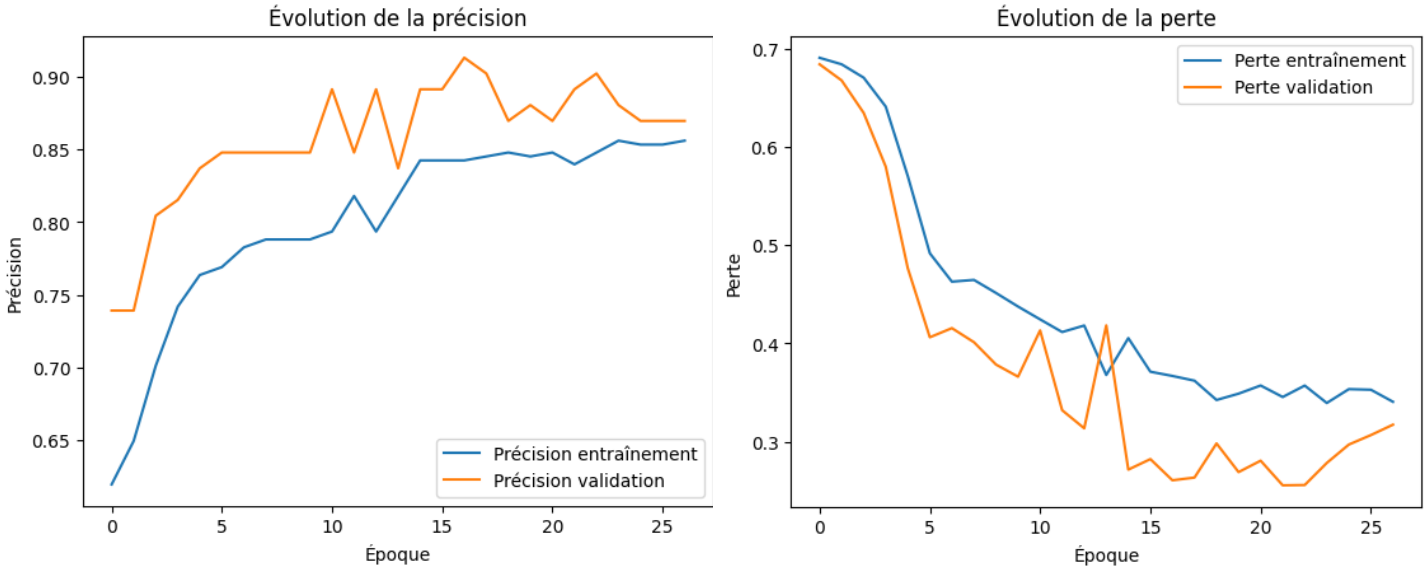


Figure IV. 6.Evolution de précision et perte

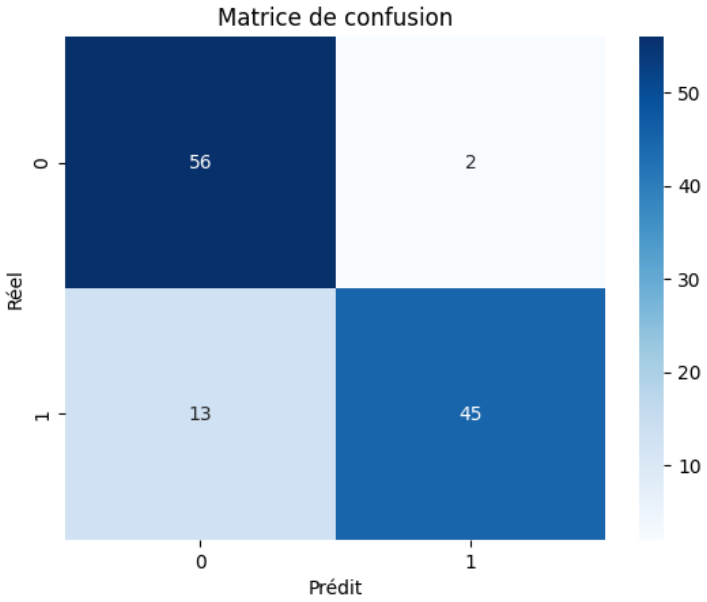


Figure IV. 7.Matrice de confusion du modèle LSTM

L'analyse approfondie des performances du modèle LSTM dans le cadre de cette tâche de classification binaire révèle des caractéristiques notables. Le modèle présente une précision élevée (0.96) pour la classe frauduleuse, limitant ainsi les faux positifs et réduisant les risques de sanctions injustifiées à l'encontre des utilisateurs légitimes. Parallèlement, il démontre une excellente sensibilité (rappel de 0.97) pour la classe authentique, assurant une identification quasi-complète des comptes légitimes. Ces résultats témoignent d'une bonne robustesse globale, comme en atteste l'équilibre des scores F1 (0.88 et 0.86 respectivement pour les classes 0 et 1).

Cependant, l'évaluation met en lumière une limitation significative : un rappel modéré (0.78) pour la détection des comptes frauduleux, impliquant qu'environ 22% des cas de fraude échappent au système. Cette lacune, comparée aux performances supérieures d'algorithmes classiques comme Random Forest en termes de rappel, souligne un compromis inhérent entre précision et sensibilité dans la conception du modèle.

En conclusion, bien que le LSTM présente des avantages indéniables pour les applications nécessitant une grande spécificité, son taux de détection incomplet des fraudes suggère la nécessité d'explorer des pistes d'amélioration. Celles-ci pourraient inclure l'optimisation des hyperparamètres, l'ajustement des seuils de classification, ou encore le développement d'une approche hybride combinant les forces du LSTM et d'autres algorithmes. Ces perspectives ouvrent la voie à des recherches futures visant à maximiser simultanément

précision et rappel, tout en tenant compte des contraintes opérationnelles spécifiques au domaine d'application.

IV.7. Conclusion

Ce chapitre a permis de décrire les outils et méthodes utilisés pour développer et évaluer un système de détection de faux comptes Instagram. Les résultats montrent que le modèle Random Forest offre les meilleures performances globales, avec un équilibre optimal entre précision et rappel. Bien que le modèle LSTM présente des avantages en termes de spécificité, son rappel modéré pour la détection des fraudes révèle des limites. Ces observations soulignent l'importance de choisir un modèle adapté aux exigences du problème, en tenant compte des contraintes opérationnelles. Des pistes d'amélioration, telles que l'optimisation des hyperparamètres ou l'exploration d'approches hybrides, pourraient être envisagées pour renforcer les performances futures. En somme, ce chapitre fournit une base solide pour la compréhension et l'optimisation des systèmes de classification binaire dans le domaine des réseaux sociaux.

Conclusion générale

Conclusion générale

À l'issue de ce mémoire, nous avons démontré la pertinence de l'usage de l'apprentissage profond, et en particulier des réseaux LSTM, pour la détection automatique des faux comptes sur les réseaux sociaux. En s'appuyant sur une base de données simulée représentative de comptes Instagram, nous avons pu entraîner un modèle performant, capable de distinguer avec une bonne précision les comptes authentiques des comptes frauduleux.

L'approche adoptée a permis de surmonter certaines limites des méthodes traditionnelles, notamment grâce à l'analyse séquentielle des comportements utilisateurs et à la capacité du LSTM à mémoriser les relations temporelles et structurelles. Les résultats obtenus confirment la robustesse de cette méthode, tout en ouvrant la voie à des améliorations futures, notamment via l'optimisation des hyper paramètres, l'exploitation de jeux de données réels plus volumineux, ou encore l'intégration d'approches hybrides combinant plusieurs modèles.

Ce travail constitue une contribution concrète au renforcement de la sécurité sur les plateformes numériques, et illustre le potentiel croissant de l'intelligence artificielle dans la lutte contre les menaces informationnelles.

Références

[1]ScienceDirect. (n.d.). *Network society*. ScienceDirect Topics. https://www.sciencedirect.com/topics/social-sciences/network-society

[2]Investopedia. (n.d.). *Social networking*. Investopedia. https://www.investopedia.com/terms/s/social-networking.asp

[3] Journal of Professional Studies Academy. (n.d.). *The impact of social networking sites*. *JPSA*, 13(1), Article 5. https://www.jpsa.ac.ae/journal/vol13/iss1/5/

[4] Plannthat. (n.d.). *How to spot a fake Instagram account (and what to do)*. Plann. https://www.plannthat.com/fake-instagram-account/

[5] Mdpi.com. (2023). *Detecting fake profiles on social media using deep learning models*. *Computers*, 13(11), 296. https://www.mdpi.com/2073-431X/13/11/296

[6] Khan, R., Naseem, R., & Ahmad, M. (2023). Detecting fake profiles on social media using deep learning models. *Computers*, *13*(11), 296. https://doi.org/10.3390/computers13110296

[7] Alharbi, N., Alkalifah, B., Alqarawi, G., & Rassam, M. A. (2024). *Countering Social Media Cybercrime Using Deep Learning: Instagram Fake Accounts Detection*. *Future Internet*, *16*(10), 367. https://doi.org/10.3390/fi16100367

[8] DigitalStakeout. (n.d.). *Impersonation & Fake Account Detection*. https://www.digitalstakeout.com/use-cases/impersonations-fake-account-detection

[9] Subhalakshmi, R. T. (2025). *Fake Profile Detection on Social Networking Websites Using Machine Learning*. *Journal of Science Technology and Research (JSTAR)*, 6(1), 1–18. https://philarchive.org/archive/SUBFPD

[10] Marr, B. (n.d.). *A short history of deep learning everyone should read*. Bernard Marr & Co. https://bernardmarr.com/a-short-history-of-deep-learning-everyone-should-read/

[11] IBM. (n.d.). *Deep learning*. IBM Think. https://www.ibm.com/think/topics/deep-learning

[12] ISO. (2022). *Apprentissage profond : la mécanique de la magie*. International Organization for Standardisation. https://www.iso.org/fr/news/ref2733.html

[13] Google Cloud. (n.d.). *Deep learning vs. machine learning: What's the difference?* Google Cloud. https://cloud.google.com/discover/deep-learning-vs-machine-learning

[14] MathWorks. (n.d.). *Long Short-Term Memory Networks*. MathWorks Discovery. https://www.mathworks.com/discovery/lstm.html

[15]Careerera. (n.d.). *Advantages and disadvantages of deep learning*. Careerera. https://www.careerera.com/blog/advantages-and-disadvantages-of-deep-learning

[16] Inayat, U. (n.d.). *Top 10 domains of deep learning*. LinkedIn. https://www.linkedin.com/pulse/top-10-domains-deep-learning-umair-inayat-vs9vf

[17] El Himer, Mohamed, and Abdelaziz El Himer. 2022. *Deep Learning Approach with LSTM for Daily Streamflow Prediction in a Semi-Arid Area: A Case Study of Oum Er-Rbia River Basin, Morocco*. Figure 3. ResearchGate. https://www.researchgate.net/figure/The-architecture-of-Long-Short-Term-Memory-LSTM-where-s-presents-the-sigmoid-function_fig3_366969980

[18] GeeksforGeeks. (2025, 5 avril). *What is LSTM – Long Short Term Memory?* https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/

[19] Coursera Staff. (2024, 10 avril). *What Is an LSTM Neural Network?* Coursera. https://www.coursera.org/articles/lstm-neural-network

[20] Kandadi, T., & Shankarlingam, G. (2025, 1 janvier). *Drawbacks of LSTM Algorithm: A Case Study*. SSRN. https://ssrn.com/abstract=5080605SSRN

[21] Daniel. (2023, 2 octobre). *Long Short Term Memory (LSTM) : de quoi s'agit-il ?* DataScientest. https://datascientest.com/long-short-term-memory-tout-savoir

[24] Google. (n.d.). *Google Colaboratory*. https://colab.google/

[25] Databricks. (n.d.). *Everything You Wanted To Know About TensorFlow*. https://www.databricks.com/glossary/tensorflow-guide

[26] W3Schools. (n.d.). *Pandas Tutorial*. https://www.w3schools.com/python/pandas/pandas_intro.asp

[27] Bigelow, S. J. (n.d.). *What is NumPy? Explaining how it works in Python*. TechTarget. https://www.techtarget.com/whatis/definition/What-is-NumPy-Explaining-how-it-works-in-Python

[28] ActiveState. (n.d.). *What is Matplotlib in Python? How to use it for plotting?*. https://www.activestate.com/resources/quick-reads/what-is-matplotlib-in-python-how-to-use-it-for-plotting/

[29] Analytics Vidhya. (2015, January 13). *Scikit-learn (sklearn) in Python*. https://www.analyticsvidhya.com/blog/2015/01/scikit-learn-python-machine-learning-tool/