



# Mémoire de fin d'étude

PRESENTÉ EN VUE DE L'OBTENTION  
DU DIPLOME DE : (Master)

Filière : Informatique  
Option : Technologies de l'information et de la  
communication

## Thème

*Suivre les rumeurs dans les  
réseaux sociaux*

Préparé par :

LAALAOUI Mouna

LAHRI Sarah

Devant le jury :

**Président :** Mr. Charikhi Mourad Enseignant à l'université de BBA

**Examineur :** Mr. Boumaaza Farid Enseignant à l'université de BBA

Encadreur :

NAILI makhlouf

## *REMERCIEMENT*

*Nos remerciements vont premièrement à Dieu tout puissant pour la volonté, la santé et la patience, qu'il nous a donnée durant toutes ces longues années.*

*Nous exprimons nos profondes gratitude à nos parents pour leurs encouragements, leur soutien et pour sacrifices qu'ils ont enduré.*

*Nous tenons également à exprimer nos sincères remerciements à notre encadrant, le Professeur NAILI Makhlouf, pour la présentation de ce sujet, et pour son suivi continu tout au long de la réalisation de cette thèse, qui n'a cessé de nous apporter ses conseils et commentaires.*

# *Dédicace*

*Nos remerciements vont premièrement à Dieu tout puissant pour la volonté, la santé et la patience, qu'il nous a donnée durant toutes ces longues années.*

*À celui dont le front était couvert de sueur et à celui qui m'a appris que le succès ne vient qu'avec patience et persévérance, à la lumière qui a éclairé mon chemin et à la lampe dont la lumière ne s'éteint jamais dans mon cœur, du sacrifice du précieux et le précieux, et duquel j'ai tiré ma force et mon estime de moi.*

## *Mon cher père*

*À celle qui a mis le paradis sous ses pieds et m'a soulagé les épreuves par ses prières. Au grand être humain qui a toujours souhaité ouvrir les yeux pour me voir un jour comme celui-ci.*

## *Chère maman*

*À mon ombre inébranlable et aux espoirs de mes jours, aux moments où je renforçais mon soutien pour eux, et ils étaient des sources auxquelles je buvais, aux jours les meilleurs et les plus purs de ma vie, à la prunelle de mes yeux, à ma chère frères et sœurs, à tous ceux qui ont été une aide et un soutien sur ce chemin, aux amis et compagnons fidèles des années, à ceux qui sont dans l'adversité et les crises, à ceux qui ont partagé leurs sentiments et leurs conseils sincères avec moi, avec vous, ma famille .*

*Mouna*

# *Dédicace*

*En toute modestie et en toute reconnaissance, je dédie tous mes efforts traduits dans ce Mémoire :*

*A*

*Mes chers parents qui m'ont offert la vie et orienté mes pas, et qui continuent à me guider vers le chemin de la réussite.*

*Mes sœurs Maria , Amira .*

*Mon frère Karim .*

*Mon ami CHAIMA et sa famille.*

*Mes amis et a toute ma famille et à tous ceux qui m'aime.*

*Sara*

## *Résumé*

L'apparition d'Internet a transformé les communications mondiales, rendant l'information accessible à des vitesses sans précédent. Avec l'émergence des médias sociaux, les gens peuvent désormais se connecter, partager et interagir immédiatement avec divers contenus. Toutefois, la facilité de l'échange d'informations facilite également la propagation rapide de rumeurs non vérifiées et de fausses informations. Dans ce travail, on vise à traquer et détecter les rumeurs, et à cette fin, on va présenter un modèle basé sur une approche de deep learning en utilisant les algorithmes LSTM et RNN afin d'obtenir la meilleure classification possible et des résultats plus précis et plus valides.

**Mots clés** : Réseaux sociaux, Suivi des rumeurs, Apprentissage profond, Apprentissage automatique.

## *Abstract*

The emergence of the Internet has transformed global communications, making information accessible at unprecedented speeds. With the emergence of social media, people can now connect, share and interact immediately with a variety of content. However, the ease with which information can be exchanged also facilitates the rapid spread of unverified rumors and false information. In this work, we aim to track and detect rumors, and to this end, we will present a model based on a deep learning approach using LSTM and RNN algorithms in order to obtain the best possible classification and more accurate and valid results.

**Key words:** Social networks, Rumor Tracking, Deep Learning, Machine Learning.

## ملخص

لقد أدى ظهور الإنترنت إلى إحداث تحول في الاتصالات العالمية، مما جعل الوصول إلى المعلومات متاحاً بسرعات غير مسبوقة. مع ظهور وسائل التواصل الاجتماعي، يمكن للأشخاص الآن الاتصال والمشاركة والتفاعل مع المحتوى المتنوع على الفور. ومع ذلك فإن سهولة تبادل المعلومات تسهل أيضاً الانتشار السريع للشائعات والمعلومات الكاذبة التي لم يتم التحقق منها. نهدف في هذا العمل إلى تتبع الشائعات وكشفها، وتحقيقاً لهذه الغاية، سنقدم نموذجاً يعتمد على منهج التعلم العميق باستخدام خوارزميات LSTM و RNN من أجل الحصول على أفضل تصنيف ممكن ونتائج أكثر دقة وأكثر صحة.

الكلمات المفتاحية: الشبكات الاجتماعية، تتبع الشائعات، التعلم العميق، التعلم الآلي

*Table des matières*

*Introduction générale* ..... 1

*Chapitre I : Les rumeurs dans les réseaux sociaux*

**I.1. Introduction**.....3

**I.2. Réseaux sociaux** ..... 3

**I.2.1. Historique** ..... 4

**I.2.2. Analyse des réseaux sociaux** ..... 5

**I.2.3. Avantages et inconvénients des réseaux sociaux.**..... 6

**I.2.4. La plateforme Twitter** ..... 8

**I.3. Les rumeurs** ..... 9

**I.3.1. Définition** ..... 9

**I.3.2. Concepts associés** ..... 10

**I.3.3. Caractéristiques de la rumeur** ..... 11

**I.3.4. Méthodes de détection des rumeurs** ..... 11

**I.3.5. Rumeur via les réseaux sociaux** ..... 12

**I.4. Conclusion** ..... 13

*Chapitre II : Deep Learning*

**II.1. Introduction** ..... 14

**II.2. L'apprentissage profond** ..... 14

**II.2.1. L'historique** ..... 14

**II.2.2 Définition** ..... 14

**II.2.3. Domaines d'application de l'apprentissage profonde** ..... 15

**II.2.4. Différence entre l'apprentissage automatique et l'apprentissage profond...** 16

**II.2.5. Principes de fonctionnement** ..... 18



---

II.2.6. Approches d'apprentissage profond .....	19
II.2.7. Les avantages de Deep Learning .....	20
II.2.8. Quand appliquer l'apprentissage profond ? .....	21
II.3. Les algorithmes utilisés .....	22
II.3.1. Réseaux de neurones récurrents (RNN) .....	22
II.3.2. Réseaux Long Short-Term Memory (LSTM) .....	23
II.3.4. Naive Bayes .....	28
II.4. Conclusion.....	30

*Chapitre III : Conception du système*

III.1. Introduction.....	31
III.2. Architecture générale .....	31
III.3. Architecture détaillée .....	32
III.3.1. Pre-traitement .....	32
III.3.2. Apprentissage.....	37
III.3.3. Test .....	38
III.3.4. Utilisation .....	39
III.4. Conclusion .....	39

*Chapitre IV : Implémentation et résultats*

IV.1. Introduction.....	40
IV.2. Présentation de outils utilisés .....	40
IV.2.1. Configuration matérielle .....	40
IV.2.2. Environnement logiciel .....	40

---

<b>IV.2.3. Langages utilisées .....</b>	<b>41</b>
<b>IV.2.4. Bibliothèques utilisées .....</b>	<b>41</b>
<b>IV.2.5. Ensemble des données .....</b>	<b>42</b>
<b>IV.3. Métriques d'évaluation utilisés .....</b>	<b>44</b>
<b>IV.4. Résultats et discussion .....</b>	<b>46</b>
<b>IV.4. 1. Comparaison de résultat des modèles implémentes .....</b>	<b>50</b>
<b>III.5. Conclusion .....</b>	<b>51</b>
<i>Conclusion générale</i>	<b>52</b>
<i>Références bibliographiques</i>	

---

*Liste des figures*

**Fig. I.1** : Social network .....4

**Fig. I.2** : Analyse des réseaux sociaux. ....6

**Fig. I.3** : Plateforme Twitter .....8

**Fig. I.4**: Propagation des rumeurs dans les réseaux sociaux. ....12

**Fig. II.1** : La relation entre Machine Learning et Deep Learning .....17

**Fig. II.2**: Différences entre Machine Learning et Deep Learning .....18

**Fig. II.3**: Les couches des réseaux profonds .....19

**Fig. II.4** : Réseau de neurones récurrents .....22

**Fig. II.5** : Le module répétitif d'un RNN standard contient une seule couche.....24

**Fig. II.6** : Le module répétitif d'une LSTM contient quatre couches en interaction .....24

**Fig. II.7** : l'état de la cellule de LSTM .....25

**Fig. II.8** : Porte d'oublie .....26

**Fig. II.9** : Opérateur d'ajout d'informations .....27

**Fig. II.10** : Mise à jour des états des cellules .....27

**Fig. II.11** : Générer la sortie .....28

**Fig. III.1** : Architecture générale .....32

**Fig. III.2** : Bibliothèques importées .....33

**Fig. III.3** : Extraction de caractéristiques et de cibles à partir d'un dataset .....33

**Fig. III.4** : Nettoyage des données .....34

**Fig. III.5** : Vectorisation de données .....35

**Fig. III.6** : Combinaison de données .....36

**Fig. III.7** : Division de données .....36

---

<b>Fig. III.8</b> : Tokenisation de textes .....	37
<b>Fig. III.9</b> : Remplissage des séquences .....	37
<b>Fig. IV.1</b> : Représentation des classes de l'ensemble de données .....	43
<b>Fig. IV.2</b> : Quelques lignes de dataset .....	44
<b>Fig. IV.3</b> : Résultat Accuracy et Loss de modèle RNN.....	48
<b>Fig. IV.4</b> : Résultat Accuracy et Loss de modèle LSTM.....	49
<b>Fig. IV.5</b> : Résultat de modèle Naïve bayes .....	50
<b>Fig. IV.6</b> : Résultat de comparaison d'algorithme .....	51

*Liste des tableaux*

**Table. I. 1:** Tableau de concept associé à la fausse information .....10

**Table. IV.1:** Description de dataset .....43

**Table. IV.2 :** Description des colonnes de dataset .....44

**Table. IV.3 :** Principe de la matrice de confusion .....45

**Table. IV.4 :** Résultats de l’algorithme RNN .....46

**Table. IV.5 :** Résultats de l’algorithme LSTM .....47

**Table. IV.6 :** Résultats de l’algorithme Naïve bayes .....47

**Table. IV.7 :** Comparaison des résultats .....50

*Liste des abréviations*

<b>DL</b>	Deep Learning .
<b>ML</b>	Machine Learning .
<b>CNN</b>	Réseaux de neurones convolutifs .
<b>RNN</b>	Réseaux de neurone récurrent .
<b>DNN</b>	Réseaux de neurone profond .
<b>LSTM</b>	Long Short – Term Memory .
<b>GRU</b>	unités récurrentes fermée .
<b>GAN</b>	Réseaux contradictoires génératif .
<b>RL</b>	Reinforcement Learning .
<b>DQN</b>	Réseaux Deep Q
<b>DRL</b>	Deep Reinforcement Learning .
<b>BPTT</b>	Back propagation Through Time .

# *Introduction générale*

### *Introduction générale*

En fait, l'Internet a changé non seulement la communication et l'interaction dans le monde, mais aussi la façon de créer, partager et consommer du contenu. Désormais, toute personne ayant accès à l'Internet peut apporter des informations ou des opinions, ce qui démocratise la création de contenu, mais entraîne également des défis en matière d'exactitude et de fiabilité des informations.

La facilité de publication sur Internet signifie que les rumeurs peuvent se propager plus rapidement et plus largement que jamais. Contrairement aux médias traditionnels, qui disposent habituellement de plusieurs niveaux de vérification des faits avant leur publication, les plateformes en ligne peuvent permettre à des informations non vérifiées de circuler rapidement, parfois sans surveillance. Ce phénomène peut se produire via les réseaux sociaux, tels que Facebook, Twitter, Instagram, en ajoutant une photo, un tweet ou un commentaire, ou via des sites d'informations produits par les utilisateurs.

Gérer ce flux d'informations nécessite d'évaluer la crédibilité des sources, de comprendre la différence entre l'opinion et les faits et de vérifier plusieurs sources avant d'accepter ou de partager des informations. Car les rumeurs ou les « fausses nouvelles » peuvent influencer l'opinion commune et même affecter les processus démocratiques comme les votes.

Les efforts visant à lutter contre la désinformation incluent des initiatives de la part des gouvernements, des entreprises technologiques et des organisations indépendantes de vérification des faits pour identifier, classer ou supprimer les rumeurs. Il s'agit cependant de trouver un équilibre entre le besoin de liberté d'expression et celui d'empêcher la propagation incontrôlée de rumeurs nuisibles et de désinformations.

Le défi de la détection des rumeurs est multiforme. Premièrement, les rumeurs se propagent rapidement sur diverses plateformes en ligne, ce qui rend difficile la traçabilité de leur source et la vérification de leur exactitude. D'autre part, les rumeurs évoluent et changent souvent au fur et à mesure qu'elles sont partagées, ce qui rend difficile l'identification de l'affirmation originale et l'évaluation de sa véracité. En outre, la propagation de rumeurs peut diffuser la panique, la désinformation et même nuire aux individus ou aux communautés. Par



## *Introduction générale*

---

conséquent, suivre et traiter efficacement les rumeurs nécessite une approche plus globale qui comprend une surveillance permanente, une collaboration entre les parties prenantes et une amélioration de l'éducation aux médias pour aider les individus à distinguer les faits réels des faits faussés.

Dans ce projet, l'étude vise à traquer les rumeurs. Il s'agit d'identifier, de vérifier et de gérer les rumeurs au sein des communautés, afin d'empêcher leur propagation et d'atténuer leurs effets négatifs. Ces objectifs sont résumés dans les points suivants :

**Identifier les fausses informations** : le suivi des rumeurs permet de distinguer les informations exactes des rumeurs fausses ou trompeuses.

**Prévenir la panique** : en traitant immédiatement les rumeurs et en fournissant des informations précises, le suivi des rumeurs permet d'éviter une peur ou une panique inutile au sein du public.

**Protection de la réputation** : le suivi des rumeurs permet aux organisations et aux individus de protéger leur réputation en luttant contre la désinformation avant qu'elle ne se propage.

Ce mémoire est composé de quatre chapitres :

- Le 1er chapitre présente le contexte de réseaux sociaux et le contexte général de la problématique traitée.
- Le 2e chapitre présente le domaine de deep learning qui est un concept clé dans le domaine de l'intelligence artificielle, ainsi que ces principaux algorithmes utilisés.
- Le 3e chapitre montre la conception du système décrivant l'architecture générale et l'architecture détaillée du système à modéliser (classificateur).
- Le 4e chapitre sera une description des modèles proposés, les algorithmes utilisés, les outils utilisés et les résultats obtenus.
- À la fin, nous terminerons ce travail par une conclusion générale.

# *Chapitre I*

## *Les rumeurs dans les réseaux sociaux*

## **I.1. Introduction**

Les plateformes de réseaux sociaux jouent un rôle central dans les changements mondiaux, facilitant la communication, le partage d'informations et la création de communautés en ligne. Cependant, outre leurs avantages, ces plateformes posent également des défis, notamment dans la gestion des rumeurs et de la désinformation. Les plateformes telles que Facebook, Twitter, Instagram, LinkedIn et TikTok font désormais partie intégrante de la vie quotidienne des gens, leur permettant de se connecter avec leurs amis, leur famille, leurs collègues. Dans ce qui suit, on va voir les principales caractéristiques de ces plateformes.

## **I.2. Réseaux sociaux**

Les réseaux sociaux en ligne sont des communautés virtuelles qui permettent aux individus d'interagir entre eux. De nombreux réseaux sociaux s'articulent autour de centres d'intérêt, d'activités et de loisirs. Toutefois, les plus grands réseaux sociaux en ligne (Facebook, Twitter et Instagram) ont une orientation plus générale.

Les médias sociaux offrent de nombreux moyens d'interagir avec les amis et la famille, de faire de nouveaux amis ou d'élargir le réseau professionnel (Fig. I.1). Ce type de site peut être un excellent moyen de renouer avec d'anciens collègues, de partager des photos et des informations avec des amis et de se tenir rapidement au courant de l'actualité.

Lors de l'utilisation d'un réseau social, il faut contrôler totalement les informations partagées et avec qui elles sont partagées. Ainsi de comprendre et d'être à l'aise avec les paramètres de confidentialité de chaque réseau social, afin de ne donner accès aux informations partagées qu'aux personnes appropriées [1].

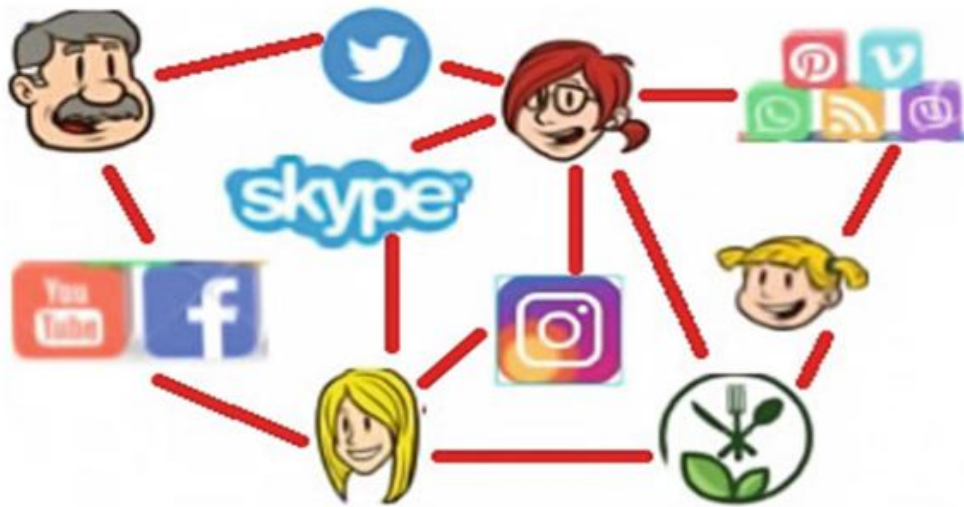


Fig. I.1 : Social network [2]

### I.2.1. Historique

À la fin des années 1890, Émile Durkheim et Ferdinand Tönnies anticipent le concept de réseaux sociaux dans leurs théories et études sur les groupes sociaux. Selon Tönnies [3], les groupes sociaux peuvent exister sous forme de liens sociaux personnels et immédiats reliant des individus partageant des valeurs et des croyances communes (communauté) ou sous forme de liens sociaux impersonnels, formels et instrumentaux (société). [4]Durkheim a fourni une explication non individualiste des faits sociaux, affirmant que les phénomènes sociaux surviennent lorsque les interactions entre individus représentent une réalité qui ne peut plus être expliquée par les attributs des acteurs individuels. [5]Georg Simmel, écrivant au début du 20e siècle, a noté la nature des réseaux et l'impact de la taille du réseau sur l'interaction et a examiné les possibilités d'interaction dans des réseaux lâches plutôt que dans des groupes.

Au cours des années 1930, plusieurs groupes de travail indépendants en psychologie, anthropologie et mathématiques ont observé des développements significatifs dans ce domaine. [6] [7]En psychologie, Jacob L. Moreno a commencé dans les années 1930 à enregistrer et analyser systématiquement les interactions sociales en petits groupes, notamment dans les salles de classe et les groupes de travail. En anthropologie, la théorie des réseaux sociaux s'appuie sur les théories de Bronislaw Malinowski [8] Alfred Radcliffe-Brown [9] [10]et Claude Lévi-Strauss. [11]Un groupe d'anthropologues sociaux associé à

Max Gluckman et à la Manchester School [12] [13] [14], reconnu pour avoir mené certains des premiers travaux de terrain sur l'analyse des réseaux, ces enquêtes ont examiné les réseaux communautaires en Afrique du Sud, en Inde et au Royaume-Uni. Au même moment [7] l'anthropologue britannique S.F. Nadel a proposé la théorie de la structure sociale qui a influencé l'analyse ultérieure des réseaux. En sociologie [15], les premiers travaux de Talcott Parsons (1930) ont été les premiers à adopter une approche relationnelle pour comprendre la structure sociale. Plus tard [16] [17], en s'appuyant sur la théorie de Parsons, les travaux du sociologue Peter Blau [18] [19] [20] donnent une forte impulsion à l'analyse des liens relationnels des unités sociales avec ses travaux sur la théorie de l'échange social.

Dans les années 1970, de plus en plus de chercheurs se sont consacrés à relier différentes écoles et concepts. Parmi ces chercheurs à cette époque, Charles Tilly, qui s'est concentré sur la sociologie politique et communautaire et les réseaux de mouvements sociaux, et Stanley Milgram, qui a proposé la thèse [20] de « six degrés de séparation », qui travaillaient également de manière indépendante au département des relations sociales de l'Université Harvard. Ainsi que, Mark Granovetter [21] et Barry Wellman [22] qui ont développé et popularisé l'analyse des réseaux sociaux.

À partir de la fin des années 1990, l'analyse des réseaux sociaux a fait l'objet de travaux menés par des sociologues, des politologues et des physiciens tels que Duncan J. Watts, Albert-László Barabási, Peter Bearman, Nicholas A. Christakis, James H. Fowler, et d'autres, développant et appliquant de nouveaux modèles et méthodes aux données émergentes disponibles sur les réseaux sociaux en ligne, ainsi qu'aux "traces numériques" en ce qui concerne les réseaux en face à face [23].

## **I.2.2. Analyse des réseaux sociaux**

L'analyse des réseaux sociaux (Social network analysis, ARS) est un processus d'analyse quantitative et qualitative d'un réseau social. L'analyse des réseaux sociaux mesure le flux de relations et les changements de relations entre les entités détentrices de connaissances. Les entités simples et complexes comprennent les sites web, les ordinateurs, les humains, les groupes, les organisations et les nations.

La structure ARS est composée d'entités de nœuds, telles que les personnes, et de liens, tels que les relations. L'avènement de la pensée et de l'informatique modernes a facilité

l'évolution progressive du concept de réseau social sous la forme de réseaux très complexes, basés sur des graphes, avec de nombreux types de nœuds et de liens. Ces réseaux sont la clé des procédures et des initiatives en matière de résolution de problèmes, d'administration et de fonctionnement des réseaux sociaux [24].



*Fig. I.2 : Analyse des réseaux sociaux [25]*

### **I.2.3. Avantages et inconvénients**

Les médias sociaux peuvent être un excellent moyen pour les gens de communiquer entre eux ou avec des groupes spécifiques de personnes, mais tout ce qui brille n'est pas de l'or [26]. Nous présentons ici les avantages et les inconvénients de l'utilisation humaine des réseaux sociaux.

#### **Pour les personnes**

##### **➤ Avantage**

Les réseaux sociaux nous permettent de nous connecter et de rester en contact avec nos amis, notre famille et nos collègues du monde entier. Cela permet de suivre facilement ce qui se passe dans la vie de chacun, même lorsque vous ne pouvez pas vous voir en personne. Les réseaux sociaux peuvent également être un excellent moyen de trouver de nouveaux amis et de se connecter avec d'autres personnes partageant les mêmes intérêts. D'autres avantages des réseaux sociaux sont la possibilité de collecter facilement des informations, telles que des

actualités, des événements et des publicités, ainsi que d'être à jour. Ils peuvent également nous aider à renouer avec de vieux amis ou à retrouver des personnes avec lesquelles nous avons perdu contact au fil des années.

### ➤ Inconvénient

Bien que les médias sociaux soient un excellent moyen de rester connecté, leur utilisation présente également de nombreux inconvénients potentiels, notamment des problèmes de confidentialité et une éventuelle utilisation abusive de nos informations personnelles.

Certains utilisateurs peuvent subir les conséquences négatives d'une utilisation excessive des médias sociaux, comme un isolement accru ou une diminution de la productivité au travail ou à l'école.

En outre, les réseaux sociaux peuvent parfois constituer un terrain fertile pour la cyber intimidation et d'autres comportements négatifs, comme le trolling ou la propagation de discours de haine.

### Pour les entreprises

#### ➤ Avantage

Les réseaux sociaux peuvent également être un outil précieux pour les entreprises, leur permettant de se connecter avec des clients potentiels et d'atteindre de nouveaux publics.

Des plateformes comme Facebook et Twitter permettent aux entreprises de partager facilement leurs actualités et de promouvoir leurs produits ou services, tandis que les réseaux professionnels comme LinkedIn les aident à se connecter avec d'autres professionnels du secteur.

En outre, les réseaux sociaux offrent un moyen de collecter des données et des analyses précieuses sur les utilisateurs, qui peuvent aider les entreprises à améliorer leurs produits et services.

#### ➤ Inconvénient

Malgré ces avantages, les réseaux sociaux présentent également quelques inconvénients pour les entreprises. Par exemple, ils présentent un risque de violation

du droit d'auteur ou de la marque si les entreprises partagent du contenu protégé par le droit d'auteur sans autorisation.

Ils peuvent également être utilisés par des concurrents pour diffuser des informations négatives ou par des clients mécontents publiant des avis négatifs.

#### I.2.4. La plateforme Twitter

Twitter est une plateforme de communication en temps réel, ce qui signifie que les messages publiés sont vus (et souvent traités) instantanément et que le flux de messages est constamment mis à jour (Fig.I.3). Il est considéré comme un site de microblog. Il s'agit d'une mini-version d'un site de blog, fusionné avec la liberté et la mobilité des SMS (messages texte), Twitter a débuté sous forme d'expérience en 2006. Lorsque la valeur de Twitter en tant que réseau de communication instantanée lors d'événements partagés tels que des tremblements de terre, des conférences et des festivals a émergé, il a commencé à grandir. C'était une force pour les organisateurs et les manifestants lors du printemps arabe et continu d'être un moyen de communication important. En décembre 2012, 32 % des internautes étaient sur Twitter, publiant plus de 175 millions de « tweets » quotidiennement [1].

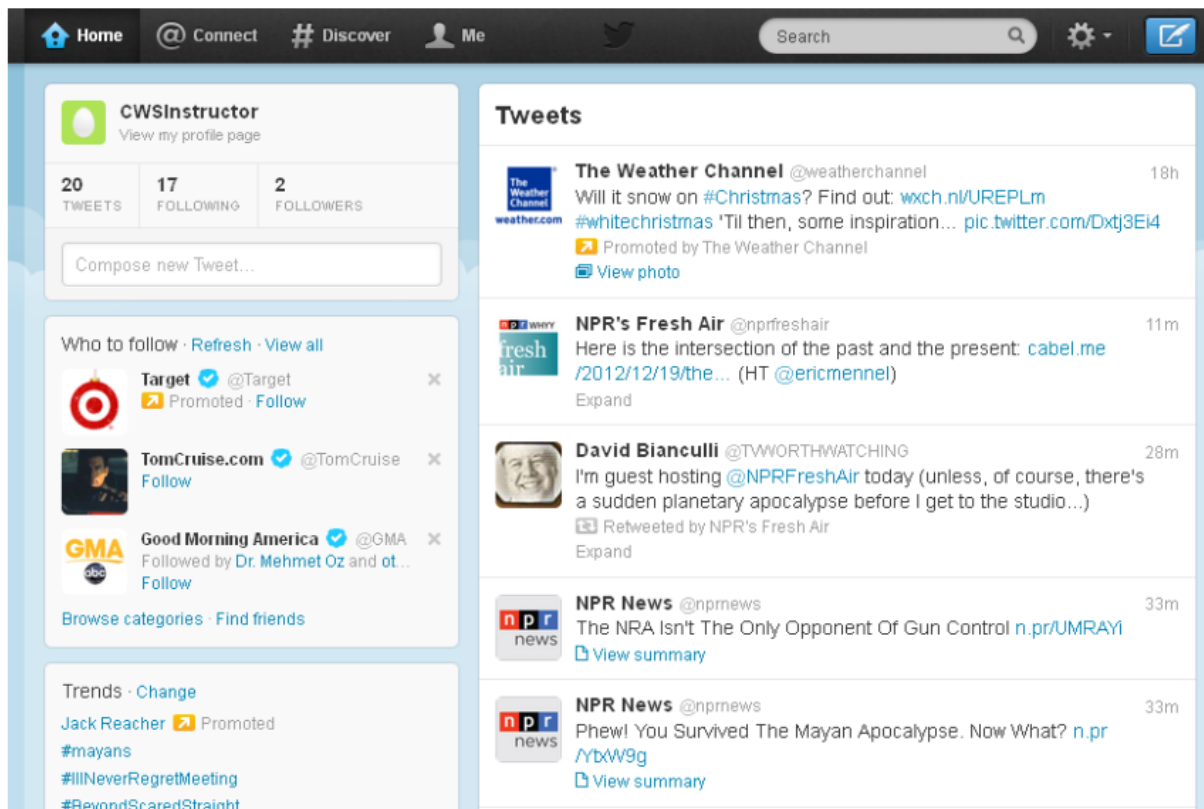


Fig.I. 1 : Plateforme Twitter [1]



### **I.3. Les rumeurs**

#### **I.3.1. Définition**

Une rumeur se définit comme une information non vérifiée transmise d'une personne à d'autres. Les rumeurs sont une réponse naturelle aux périodes incertaines ou menaçantes. Les rumeurs ne sont ni bonnes ni mauvaises en soi. Elles peuvent être vraies ou fausses, ou un mélange des deux [27]. Il existe deux types de rumeurs, à savoir :

- **Mésinformation** : informations erronées diffusées par des personnes sans intention de tromper ou de transmettre des informations erronées, par exemple en raison d'une mauvaise compréhension du message.
- **Désinformation** : informations incorrectes diffusées par des personnes dans le but de tromper ou de manipuler d'autres personnes.

Si les moyens et les motifs peuvent varier, l'impact est le même : les gens sont incapables de faire des choix éclairés sur leur avenir. Fonder ces choix sur des informations non vérifiées peut avoir des conséquences dévastatrices.

Catégories de rumeurs :

- Rumeurs de souhait.
- Rumeurs de peur.
- Rumeurs d'hostilité.

Les rumeurs de peur sont les plus répandues. Ils permettent aux gens de réagir soit en agissant physiquement, ou en s'isolant contre l'impact émotionnel d'un tel événement.

#### **Pourquoi les gens partagent-ils des rumeurs ?**

- Expliquer une situation ou un événement.
- Partagez des informations utiles ou divertissantes.
- Se définir en étant « au courant » ou en donnant une mauvaise image aux autres.
- Développer des relations en utilisant l'information comme monnaie d'échange pour se sentir connecté aux problèmes qui les concernent.
- Induire en erreur ou tromper, souvent pour des raisons économiques ou politiques.

### I.3.2. Concepts associés

Les études existantes relient souvent les fausses informations à des termes et concepts tels que les fausses informations malveillantes, fausses nouvelles, informations satiriques, désinformation, mésinformation et rumeur. Sur la base de ces termes et concepts définis, nous pouvons distinguer l'un des autres en fonction de trois caractéristiques [28]:

- L'authenticité (fausse / vraie),
- L'intention (trompeuse / divertissante),
- Si l'information est nouvelle ou non.

Le tableau suivant contient les détails

**Tab.I. 1 :** Tableau de concept associé à la fausse information [28]

	<b>Authenticité</b>	<b>Intention</b>	<b>Nouvelle ?</b>
<b>Nouvelle trompeuse</b>	Fausse	Trompeuse	Oui
<b>Fausse nouvelles</b>	Fausse	Inconnu	Oui
<b>Information satirique</b>	Non unifié	Divertissante	Oui
<b>Désinformation</b>	Fausse	Trompeuse	Inconnu
<b>Mésinformation</b>	Fausse	Inconnu	Inconnu
<b>Rumeur</b>	Inconnu	Inconnu	Inconnu

### Définition

**Information :** une information est définie par Yves- François LE COADIC comme : « une connaissance inscrite (enregistrée) sous forme écrite (imprimé ou numérisée), orale ou audiovisuelle sur un support spatiotemporel [29].

**Opinion :** Une opinion est un jugement de valeur individuel sur un objet par une personne dans un instant spécifique. Une opinion n'est donc pas une information, car elle est particulièrement applicable et valable pour la personne qui l'exprime. Elle est subjective [30].

**Rumeur :** une rumeur est un bruit informel. Il existe, il persiste, il s'évapore. Il repart aussi vite qu'il est arrivé. Il s'agit d'une ou de plusieurs informations qui circulent et qui sont transmises par un vecteur (individus ou internet). Il n'y a pas forcément de source déterminée [31].

### **I.3.3. Caractéristiques de la rumeur**

Tout d'abord, on peut décrire la rumeur par des caractéristiques réparties en 3 classes [32] :

- **La situation :** la rumeur apparaît dans une situation de crise, mais elle n'est pas toujours le signe d'un dysfonctionnement social, les canaux formels de communication ne véhiculent qu'une information réduite sur certains événements ou aspects de cette situation, c'est-à-dire que devant la privation d'information les individus vont créer des rumeurs.
- **Le processus de transmission :** la rumeur ne se transmet oralement de personne à personne, par le bouche-à-oreille, mais aussi par les médias (presse, radios, internet). Les canaux sont donc formels (médias) et informels (bouche-à-oreille). Cette communication a lieu entre des individus également impliqués dans la situation.
- **Le contenu :** le contenu de la rumeur connaît différentes distorsions au cours de son processus de transmission. Ce contenu traduit la pensée de désir de la population, elle témoigne de l'exercice d'une pensée sociale, la rumeur devient une sorte d'écran projectif où se déchiffre une dynamique socio affective. Il entretient un rapport avec l'actualité.

### **I.3.4. Méthodes de détection des rumeurs**

L'analyse de la véracité des informations, notamment la vérification des rumeurs, est un domaine de recherche important. Ce processus consiste à évaluer l'authenticité d'une information suspecte pour déterminer si elle est vraie ou fausse. Un système efficace de vérification des rumeurs comprend plusieurs étapes essentielles, allant de la détection initiale de l'information comme rumeur jusqu'à l'analyse finale pour établir sa véracité. Cette démarche est cruciale pour éviter la propagation de fausses informations, notamment sur les réseaux sociaux, où les rumeurs peuvent se répandre rapidement.



Fig. I. 2 : Propagation des rumeurs dans les réseaux sociaux

La méthode de résolution des rumeurs comprenant quatre tâches [33]

- **Détection des rumeurs** : déterminante si une réclamation est une vérification de valeur plutôt que l'expression d'une opinion.
- **Suivi des rumeurs** : rassembler les sources et les opinions sur une rumeur au fur et à mesure qu'elle se déroule.
- **Classification des positions** : déterminer l'attitude des sources ou des utilisateurs envers la véracité de la rumeur.
- **Vérification de la rumeur** : comme l'étape ultime où la valeur de véracité de la rumeur est prédite.

### I.3.5. Rumeur via les réseaux sociaux

Une rumeur via les réseaux sociaux est une nouvelle qui est diffusée via un site de réseautage.

Dans un laps de temps très court, elle va prendre une ampleur considérable de manière virale. Sur internet, et dans les réseaux sociaux en particulier, quatre caractéristiques sont ajoutées à celles reconnues habituellement à la rumeur :

- **Instantanéité** : les internautes ont très facilement accès à leur messagerie instantanée et aux réseaux sociaux, grâce à des outils technologiques de plus en plus abordables. Cela favorise évidemment la diffusion rapide d'une information.
- **Décentralisation** : sur internet, la diffusion de l'information n'est pas réglementée par les principes déontologiques du journalisme professionnel comme ce l'est dans les médias d'information traditionnels. Tout un chacun a accès aux outils qui lui permettent d'éditer du contenu en ligne. L'information surgit sur nos profils Facebook, nos chronologies Twitter, dans le flux de nos mails ou sur le blog d'une personnalité influente. Nous sommes tous susceptibles d'être les « relayeurs » d'une rumeur.
- **Visibilité** : les réseaux sociaux sont basés sur le principe d'une liste de contacts qui s'échangent des informations. Relayer les propos « d'amis » dans sa propre sphère est donc le premier pas vers une visibilité accrue par les relais successifs.
- **Internationalité** : les réseaux « d'amis » ne connaissant pas de frontières, les nouvelles, et donc les rumeurs, font rapidement le tour du globe de mail en mail ou de profil en profil. La confirmation ou l'infirmité d'une rumeur ne parviendra donc pas forcément aux oreilles « électroniques » de chaque internaute. Internet et les réseaux sociaux laissent donc libre cours aux croyances et offrent ainsi une plus longue vie aux rumeurs que lorsque celles-ci se propagent par le bouche-à-oreille humain.

#### **I.4. Conclusion**

Le phénomène des rumeurs, compte tenu de la domination des réseaux sociaux, est une arme à double tranchant : d'une part, il s'agit d'un outil permettant de communiquer et de diffuser des informations de manière simple et rapide. D'autre part, ils constituent une plateforme fertile pour la propagation de rumeurs et de fausses nouvelles à une vitesse énorme, susceptibles de causer de graves dommages aux individus et aux sociétés. Dans ce chapitre, les concepts de réseaux sociaux et leurs caractéristiques, ainsi que les rumeurs et leurs modes de détection sont présentés.

# *Chapitre II*

## *Deep Learning*

## **II.1 Introduction**

Le Deep Learning ou apprentissage profond en français est l'un des domaines d'apprentissage les plus importants, les plus difficiles et les plus innovants, car il permet à l'apprentissage automatique multi-niveau d'explorer de grandes quantités de données et de découvrir des modèles complexes. Il est intensivement entraîné et appliqué à un grand nombre d'applications. Dans ce chapitre, nous présenterons d'abord les concepts liés au Deep Learning.

## **II.2. L'apprentissage profond**

L'apprentissage profond (Deep Learning) représente une nouvelle frontière dans le domaine de l'apprentissage automatique et il a été introduit dans le cadre de la phase de développement de ce domaine. L'apprentissage profond appartient à la branche de l'intelligence artificielle, dans laquelle les algorithmes utilisés s'inspirent de la structure et du fonctionnement du cerveau humain. Ces algorithmes sont capables d'extraire plusieurs niveaux de représentation, ce qui leur permet de comprendre et d'analyser des relations complexes de manière très efficace.

### **II.2.1. Historique**

L'évolution des réseaux de neurones artificiels se poursuit depuis leur première proposition en 1943. D'abord, par l'apparition du premier modèle de "neurone formel" binaire, suivi du développement du "perceptron" en 1957, marquant le début du réseau de neurones artificiels. Ensuite, les avancées ultérieures, telles que le « perceptron multicouche » dans les années 1980. L'introduction du terme "Deep Learning" est attribuée aux travaux de Yann LeCun sur les réseaux de neurones convolutifs. Depuis 2012, la majorité des entreprises utilisant la machine Learning se tournent vers le Deep Learning dans divers domaines [34] [35].

### **II.2.2. Définition**

Le Deep Learning fait partie de la Machine Learning qui utilise des réseaux de neurones artificiels pour apprendre à partir de nombreuses données sans avoir besoin de programmation explicite. Ces réseaux s'inspirent du cerveau humain et peuvent être utilisés pour des tâches telles que la reconnaissance d'images, la compréhension de la parole et le

traitement du langage. Il existe différents types de réseaux d'apprentissage profond, comme les réseaux de neurones à action directe, les réseaux de neurones convolutifs et les réseaux de neurones récurrents. Le Deep Learning a besoin de beaucoup de données étiquetées et d'ordinateurs puissants pour fonctionner correctement, mais il peut obtenir de très bons résultats dans de nombreuses applications [36].

### II.2.3. Domaines d'application de l'apprentissage profond

Parmi les domaines d'application de l'apprentissage profond [37], il y a :

- **Assistants virtuels :**

les assistants virtuels sont des applications basées sur le cloud qui comprennent les commandes vocales en langage naturel et effectuent des tâches pour l'utilisateur. Amazon Alexa, Cortana, Siri et Google Assistant sont des exemples typiques d'assistants virtuels. Ils ont besoin d'appareils connectés à Internet pour fonctionner avec toutes leurs capacités.

- **Les chatbots**

Ils peuvent résoudre les problèmes des clients en quelques secondes. Un chatbot est une application d'IA permettant de discuter en ligne via texte ou synthèse vocale. Il est capable de communiquer et d'effectuer des actions similaires à celles d'un humain. Les chatbots sont beaucoup utilisés dans l'interaction client, le marketing sur les sites de réseaux sociaux et la messagerie instantanée avec le client. Il fournit des réponses automatisées aux entrées des utilisateurs. Il utilise des algorithmes d'apprentissage automatique et d'apprentissage profond pour générer différents types de réactions.

- **Soins de santé**

Le Deep Learning a trouvé son application dans le secteur de la santé. La détection et le diagnostic assistés par ordinateur des maladies ont été possibles grâce au Deep Learning. Il est largement utilisé pour la recherche médicale, la découverte de médicaments et le diagnostic de maladies potentiellement mortelles telles que le cancer et la rétinopathie diabétique grâce au processus d'imagerie médicale.

- **Divertissement**



Des entreprises telles que Netflix, Amazon, YouTube et Spotify proposent des recommandations pertinentes de films, de chansons et de vidéos pour améliorer leur expérience client. Tout cela grâce au Deep Learning. En fonction de l'historique de navigation, des intérêts et du comportement d'une personne, les sociétés de streaming en ligne proposent des suggestions pour l'aider à faire des choix de produits et de services. Des techniques d'apprentissage profond sont aussi utilisées pour ajouter du son aux films muets et générer automatiquement des sous-titres.

- **Agrégation de nouvelles et détection de fausses nouvelles**

Le Deep Learning permet aussi de personnaliser l'actualité en fonction de la personnalité des lecteurs. Il regroupe les informations d'actualité en fonction de paramètres sociaux, géographiques et économiques et des préférences individuelles d'un lecteur. Les réseaux de neurones aident à développer des classificateurs capables de détecter les informations fausses et biaisées et de les supprimer de votre flux. Ils vous avertissent également d'éventuelles violations de la vie privée.

- **Traitement du langage naturel**

Un autre domaine important dans lequel le Deep Learning montre des résultats prometteurs est le NLP (Natural Language Processing). C'est la procédure permettant aux robots d'étudier et de comprendre le langage humain.

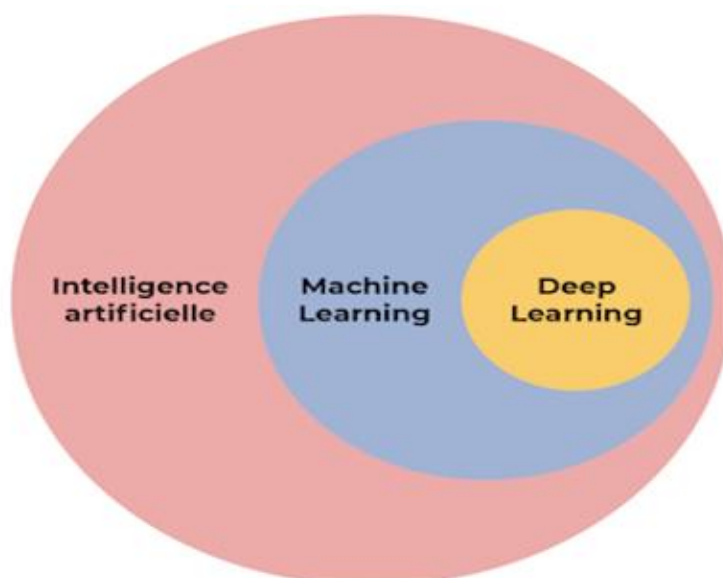
De nombreux défis associés à la compréhension du langage humain sont résolus par la PNL basée sur le Deep Learning en apprenant aux ordinateurs à fournir des réponses appropriées aux entrées linguistiques.

### **II.2.4. Différences entre l'apprentissage automatique et l'apprentissage profond**

#### **L'apprentissage automatique**

L'apprentissage automatique est une technologie d'intelligence artificielle qui permet aux ordinateurs d'apprendre sans programmation explicite. Cependant, pour apprendre et se développer, les ordinateurs ont besoin de données pour l'analyse et la formation. L'apprentissage automatique comprend une variété de méthodes pour créer automatiquement des modèles à partir de données. Ces méthodes sont en fait des algorithmes qui fournissent

des solutions à des problèmes complexes en traitant de grandes quantités d'informations. Cela offre ainsi une possibilité d'analyser et de mettre en évidence les corrélations qui existent entre deux ou plusieurs situations données, et de prédire leurs différentes implications [38].



**Fig. II. 1 :** La relation entre Machine Learning et Deep Learning [38]

Il existe de nombreuses différences entre l'apprentissage automatique et l'apprentissage profond ; en voici quelques différences [39] :

- **Intervention humaine :** l'apprentissage automatique nécessite une intervention humaine plus soutenue pour obtenir des résultats. L'apprentissage en profondeur est plus complexe à mettre en place, mais ne nécessite qu'une intervention minimale par la suite.
- **Matériel :** les programmes d'apprentissage automatique sont généralement moins complexes que les algorithmes d'apprentissage profond et peuvent souvent être exécutés sur des ordinateurs classiques, mais les systèmes d'apprentissage profond nécessitent du matériel et des ressources beaucoup plus puissants. Par exemple, l'utilisation accrue des unités de traitement graphique (GPU).
- **Temps :** les systèmes d'apprentissage automatique peuvent être mis en place et exploités rapidement, mais la puissance de leurs résultats peut être limitée. Les systèmes d'apprentissage en profondeur prennent plus de temps à mettre en place, mais peuvent générer des résultats plus précis.

- **Approche** : l'apprentissage automatique a tendance à nécessiter des données structurées et utilise des algorithmes traditionnels tels que la régression linéaire. L'apprentissage profond utilise des réseaux neuronaux et peut traiter de grands volumes de données non structurées.
- **Applications** : l'apprentissage automatique est déjà utilisé dans les boîtes aux lettres électroniques, les banques et les cabinets médicaux. La technologie de l'apprentissage en profondeur permet de créer des programmes plus complexes et plus autonomes, comme des voitures qui se conduisent toutes seules ou des robots qui pratiquent des opérations chirurgicales avancées.

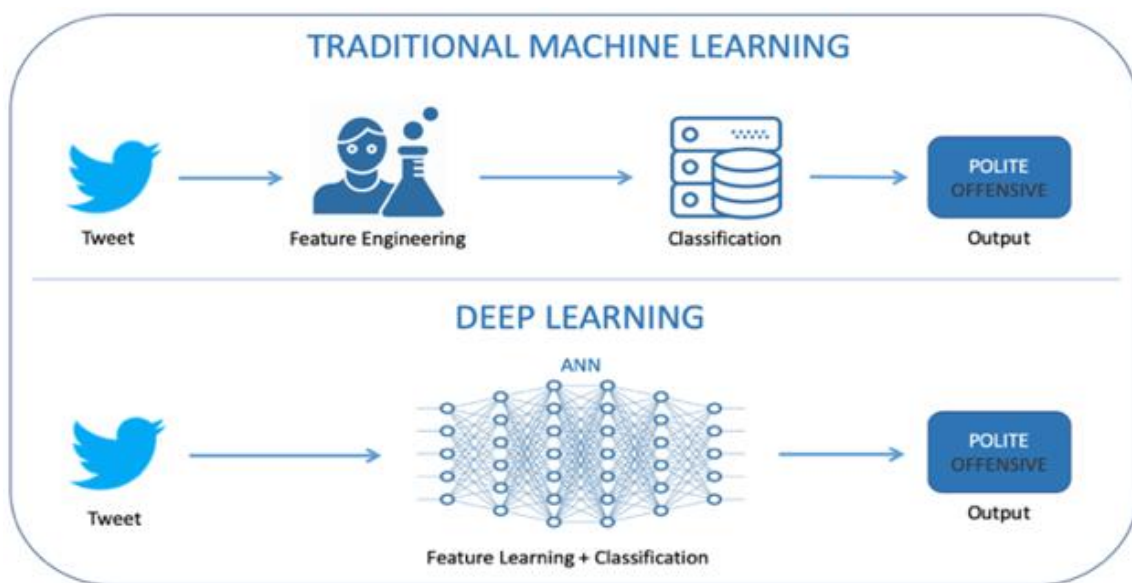


Fig. II. 2: Différences entre machine learning et deep learning [40]

### II.2.5. Principe de fonctionnement

Les structures de données complexes qu'ils rencontrent servent de base aux réseaux d'apprentissage profond. Ces réseaux construisent des modèles informatiques composés de nombreuses couches (couche d'entrée, couche cachée, couche de sortie) pour fournir plusieurs niveaux d'abstraction [41].

Le réseau neuronal profond se compose de :

- Couche d'entrée** : un réseau neuronal artificiel est composé de plusieurs nœuds qui reçoivent des données en entrée. Ces nœuds forment la couche d'entrée du système.
- Couche cachée** : la couche d'entrée traite et envoie les données aux couches suivantes du réseau neuronal. Les couches cachées traitent les informations à divers niveaux,

ajustant leur fonctionnement en fonction des nouvelles données reçues. Le Deep Learning implique l'utilisation de plusieurs couches cachées pour analyser un problème sous différents angles.

- c. **Couche de sortie :** la couche de sortie est formée par les nœuds qui génèrent les résultats. Par exemple, les modèles de Deep Learning qui donnent des réponses « oui » ou « non » sont des nœuds dans cette couche. En revanche, ceux qui produisent une variété plus large de réponses ont un nombre plus élevé de nœuds.

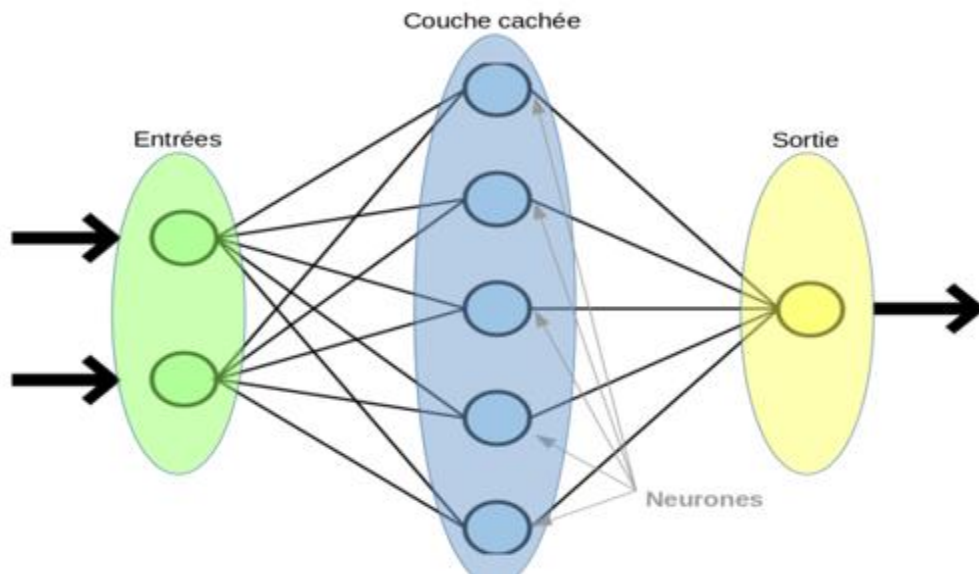


Fig. II. 3 : Les couches des réseaux profonds

## II.2.6. Approches d'apprentissage profond

L'apprentissage profond présente plusieurs approches utilisées pour apprendre les réseaux de neurones profonds. Les plus importantes de ces approches sont :

### a. Apprentissage supervisé

C'est une technique qui utilise des données étiquetées. Dans le cas d'une approche DL supervisée, l'environnement contient un ensemble d'entrées et de sorties résultantes. Les paramètres du réseau sont ensuite mis à jour de manière itérative pour obtenir une estimation améliorée de la sortie préférée. Après avoir obtenu un résultat de formation approprié, le modèle construit acquiert la capacité d'obtenir des solutions correctes aux requêtes de l'environnement [42].

### b. Apprentissage semi-supervisé

Il s'agit d'un processus d'apprentissage basé sur des ensembles de données semi-étiquetés. L'un des avantages de cette technique est qu'elle minimise la quantité de données étiquetées nécessaires. Le classificateur de documents texte est l'un des exemples les plus populaires d'applications d'apprentissage semi-supervisé. En raison de la difficulté d'obtenir une grande quantité de documents texte étiquetés, l'apprentissage semi-supervisé est idéal pour les tâches de classification de documents texte [42].

### **c. Apprentissage non supervisé**

Il s'agit d'un apprentissage qui a lieu sans données étiquetées disponibles (c'est-à-dire qu'aucune étiquette n'est requise). Ici, le modèle apprend les caractéristiques importantes ou la représentation interne requise pour découvrir une structure ou des relations non spécifiées dans les données d'entrée. Les techniques de réseau génératif, la réduction de dimensionnalité et le clustering sont souvent comptés dans la catégorie de l'apprentissage non supervisé [42].

### **d. Apprentissage par renforcement**

Dans lequel l'agent apprend en interagissant avec l'environnement qui fournit des signaux de rétroaction pour les actions que l'agent tente d'effectuer. Cette exigence est plus facile à réaliser, permettant d'appliquer l'apprentissage par renforcement à plus de situations que l'apprentissage supervisé. Les agents d'apprentissage par renforcement génèrent instantanément leurs propres données, grâce à leurs actions, et les récompenses de l'environnement. Les réseaux Deep Q (DQN) sont un exemple couramment utilisé dans les jeux ou la robotique [43].

## **II.2.7. Avantages de Deep Learning**

L'apprentissage profond présente un outil précieux pour toute une série d'applications. La valeur de l'apprentissage profond dépend du contexte, des objectifs et des ressources. Voici quelques-uns des avantages de l'apprentissage profond [44].

- **Extraction automatisée des caractéristiques** : l'apprentissage profond permet l'identification automatique des caractéristiques pertinentes des données. Il n'est donc pas nécessaire d'élaborer manuellement des caractéristiques, ce qui prend du temps et il est spécifique à un domaine particulier.

- **Précision élevée** : les modèles d'apprentissage profond peuvent être entraînés à atteindre une grande précision pour diverses tâches, telles que le traitement de la parole et la reconnaissance d'images.
- **Large applicabilité** : l'apprentissage profond est polyvalent et peut être utilisé dans divers domaines, de la conduite autonome à l'analyse financière.
- **Apprentissage continu** : les modèles d'apprentissage profond peuvent être améliorés et mis à jour au fil du temps à l'aide de nouvelles données. Cela leur permet de s'adapter aux changements dans les modèles et les tendances de la distribution des données.
- **Représentation hiérarchique des données** : Les réseaux profonds sont capables d'apprendre des modèles hiérarchiques. Ils peuvent saisir à la fois les caractéristiques simples des données et les relations complexes, ce qui leur permet d'obtenir des informations plus nuancées et plus riches.
- **Traitement en temps réel** : Certaines architectures d'apprentissage profond optimisées pour le traitement parallèle permettent d'analyser les données en temps réel. Cela permet des applications telles que la traduction linguistique instantanée ou l'analyse vidéo.

### II.2.8. Quand appliquer l'apprentissage profond ?

L'intelligence artificielle est utile dans de nombreuses situations et est égale ou meilleure que les experts humains dans certains cas, ce qui signifie que DL peut être une solution aux cas suivants :

- Les experts humains ne sont pas disponibles.
- Les humains sont incapables d'expliquer les décisions prises en utilisant leur expertise (compréhension du langage, décisions médicales et reconnaissance vocale).
- La solution au problème est mise à jour au fil du temps (prévision de prix, préférence en matière de stock, prévision météo et suivi).
- Les solutions nécessitent une adaptation en fonction de cas précis (personnalisation, biométrie).
- La taille du problème est extrêmement importante et dépasse nos capacités de raisonnement inadéquates (analyse des sentiments, correspondance des publicités avec Facebook, page Web de calcul rangs) [42].

## II.3. Les algorithmes utilisés

### II.3.1 Réseaux de neurones récurrents

Réseaux de neurones récurrents (RNN) ont été proposés dans les années 1980 pour modéliser des séries chronologiques. Cette structure est similaire aux réseaux de neurones traditionnels, mais ces réseaux permettent de conserver l'information en se l'envoyant [45].

#### Définition

Le réseau neuronal récurrent (RNN) est un type de réseau neuronal dans lequel la sortie de l'étape précédente sert d'entrée à l'étape actuelle. Dans les réseaux neuronaux traditionnels, toutes les entrées et sorties sont indépendantes les unes des autres. Toutefois, lorsqu'il s'agit de prédire le mot suivant d'une phrase, les mots précédents sont nécessaires et il faut donc s'en souvenir. C'est ainsi qu'est né le RNN, qui résout ce problème à l'aide d'une couche cachée. La caractéristique principale et la plus importante du RNN est son état caché, qui se souvient de certaines informations concernant une séquence. Cet état est également appelé état de mémoire, car il se souvient de l'entrée précédente du réseau. Il utilise les mêmes paramètres pour chaque entrée, car il effectue la même tâche sur toutes les entrées ou couches cachées pour produire la sortie. Cela réduit la complexité des paramètres, contrairement à d'autres réseaux neuronaux [45].

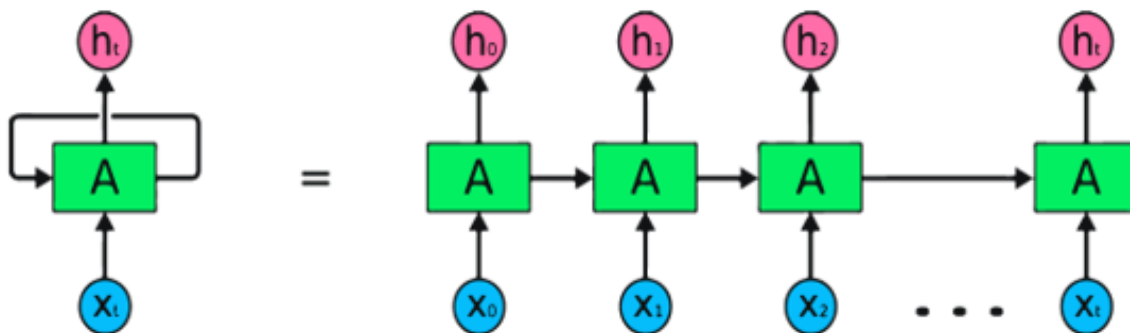


Fig. II. 4 : Réseau de neurones récurrents [45]

#### Comment fonctionne un réseau neuronal récurrent ?

Le réseau neuronal récurrent se compose de plusieurs unités à fonction d'activation fixe, une pour chaque pas de temps. Chaque unité possède un état interne appelé état caché de l'unité. Cet état caché représente les connaissances passées que le réseau détient actuellement

à un moment donné. Cet état caché est mis à jour à chaque pas de temps pour signifier le changement dans les connaissances du réseau sur le passé. L'état caché est mis à jour à l'aide de la relation de récurrence suivante [45] :

**La formule de calcul de l'état actuel :**

$$h_t = (h_{t-1}, x_t)$$

**Avec :**

- $h_t$  : État actuel .
- $h_{t-1}$  : État précédent .
- $x_t$  : État d'entrée.

**Formule d'application de la fonction d'activation (tanh) :**

$$h_t = \tanh (w_{hh}h_{t-1} + w_{xh}x_t)$$

**Avec :**

- $w_{hh}$ : Poids du neurone récurrent .
- $w_{xh}$ : Poids du neurone d'entrée.

**La formule de calcul de la sortie :**

$$y_t = w_{hy}h_t$$

**Avec :**

- $y_t$  : Sortie du neurone .
- $w_{hy}$  : Poids de la couche de sortie .

### II.3.2. Réseaux Long Short-Term Memory

Les réseaux LSTM (Long Short-Term Memory) sont un type particulier de RNN, capable d'apprendre des dépendances à long terme. Ils fonctionnent extrêmement bien sur une grande variété de problèmes, et sont maintenant largement utilisés [46].

Les LSTM sont capables de gérer et d'interpréter des données séquentielles, notamment des séries chronologiques, du texte et de la parole. Ces réseaux trouvent de nombreuses applications dans le traitement du langage naturel, la reconnaissance vocale et la prévision de séries chronologiques. Ils peuvent conserver ou supprimer sélectivement les



informations si nécessaire, surmontant ainsi les lacunes des RNN traditionnels. Ils utilisent des cellules de mémoire qui peuvent stocker des informations pendant de longues périodes et des portes qui régulent le flux d'informations entrant et sortant de ces cellules.

Tous les réseaux neuronaux récurrents se présentent sous la forme d'une chaîne de modules répétitifs. Dans les RNN standard, ce module répétitif présente une structure très simple, telle qu'une seule couche de tanh [46].

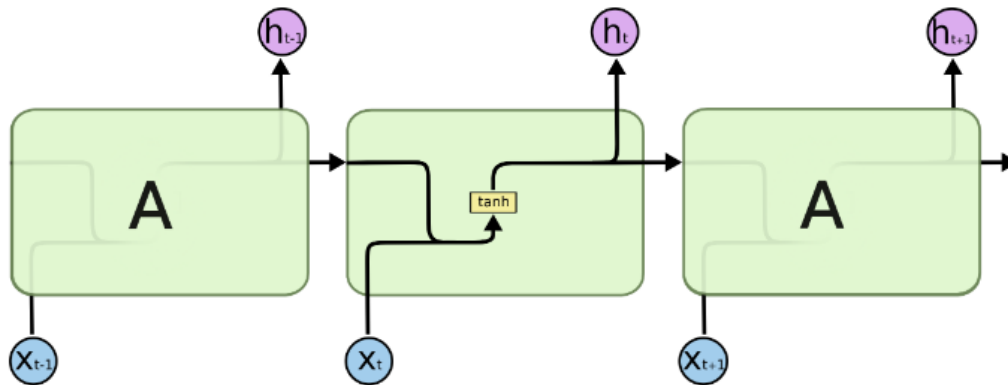


Fig. II. 5 : Le module répétitif d'un RNN standard contient une seule couche [46]

Les LSTM ont également cette structure semblable à une chaîne, mais le module répétitif a une structure différente. Au lieu d'avoir une seule couche de réseau, il y en a quatre, interagissant de manière très particulière.

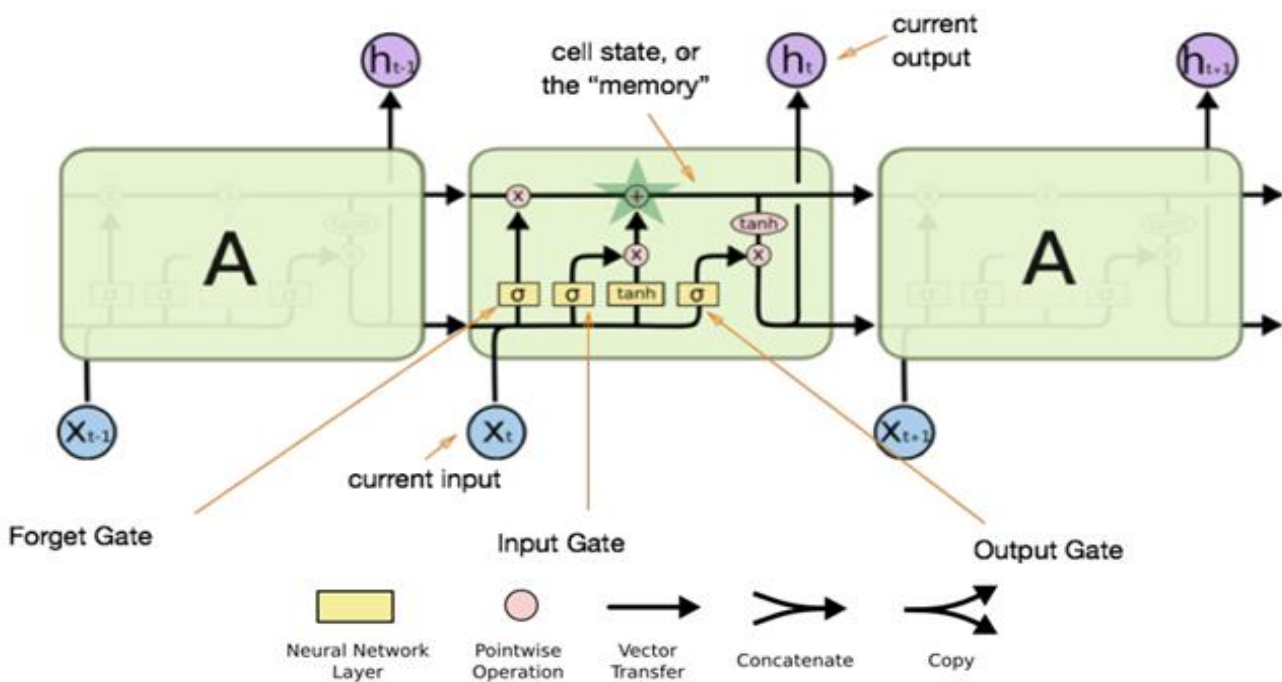


Fig. II. 6. Le module répétitif d'une LSTM contient quatre couches en interaction [46]

Dans la figure ci-dessus, chaque ligne porte un vecteur entier, de la sortie d'un nœud aux entrées d'autres nœuds. Les cercles roses représentent des opérations ponctuelles, comme l'addition de vecteurs, tandis que les boîtes jaunes sont des couches de réseau de neurones apprises. Les lignes qui fusionnent indiquent une concaténation, tandis qu'une ligne qui bifurque indique que son contenu est copié et que les copies sont placées à des endroits différents.

Il y a trois portes vers la cellule, qui sont :

- **Input gate** : décide si l'entrée doit modifier le contenu de la cellule.
- **Forget gate** : décide s'il faut remettre à 0 le contenu de la cellule.
- **Output gate** : décide si le contenu de la cellule doit influencer sur la sortie du neurone.

Ces portes sont des fonctions sigmoïdes qui déterminent la quantité d'informations à transmettre ou à bloquer depuis la cellule. Les fonctions sigmoïdes prennent des valeurs et les affichent dans la plage  $[0, 1]$ . En termes de fonction de porte, une valeur de 0 signifie ne rien laisser passer, et une valeur de 1 signifie tout laisser passer. [47]

### Fonctionnement du réseau LSTM

La clé des LSTM est l'état de la cellule, la ligne horizontale passant par le haut du diagramme. L'état cellulaire est un peu comme un tapis roulant. Il s'étend tout au long de la chaîne, avec seulement quelques interactions linéaires mineures. Il est très facile pour les informations de circuler sans changement.

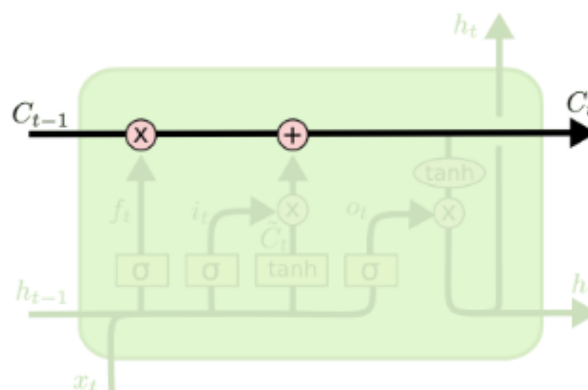


Fig. II. 7 : l'état de la cellule de LSTM

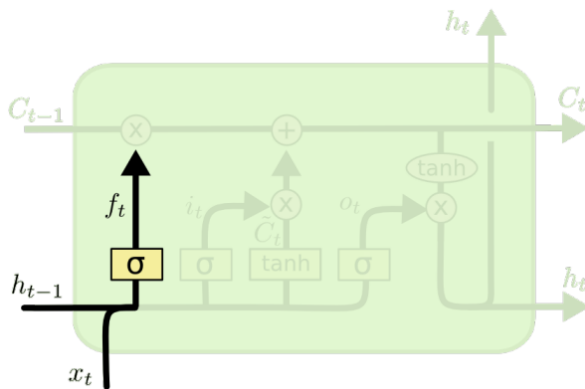
La LSTM a la capacité de supprimer ou d'ajouter des informations à l'état de la cellule, régulée par des structures appelées « portes » qui sont un moyen de laisser passer des informations de manière facultative. Ils sont composés d'une couche de réseau de neurones sigmoïde et d'une opération de multiplication ponctuelle.

La couche sigmoïde produit des nombres entre zéro et un, décrivant la quantité de chaque composant à laissez-passer. Une valeur de zéro signifie « ne rien laisser passer », tandis qu'une valeur d'un signifie « tout laisser passer ».

## Les étapes de LSTM

### a. Porte de l'oubli

La première étape de notre LSTM consiste à décider quelles informations nous allons rejeter de l'état de la cellule. Cette décision est prise par une couche sigmoïde appelée « couche de porte d'oubli », qui prend en compte  $h_{t-1}$  et  $x_t$  et émet un nombre entre 0 et 1 pour chaque nombre dans l'état de la cellule  $C_{t-1}$ . Un 1 représente « conserver complètement » tandis qu'un 0 représente « se débarrasser complètement ». [48].



$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

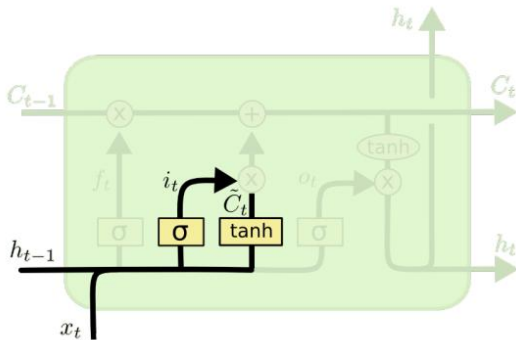
Fig.II. 8 :Porte d'oublie

Tel que :

- $h_{t-1}$  : Sortie à l'instant  $t-1$
- $x_t$  : Entrée courant à l'instant  $t$
- $b$  : Bais
- $w$  : Poids
- $\sigma$  : Fonction sigmoïde

**b. Nouvelle cellule candidate et porte d'entrée**

L'étape suivante consiste à décider quelles nouvelles informations nous allons stocker dans l'état de la cellule. Elle comporte deux parties. Tout d'abord, une couche sigmoïde appelée « couche d'entrée » décide des valeurs à mettre à jour. Ensuite, une couche tanh crée un vecteur de nouvelles valeurs candidates,  $\tilde{C}_t$ , qui pourraient être ajoutées à l'état.



$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

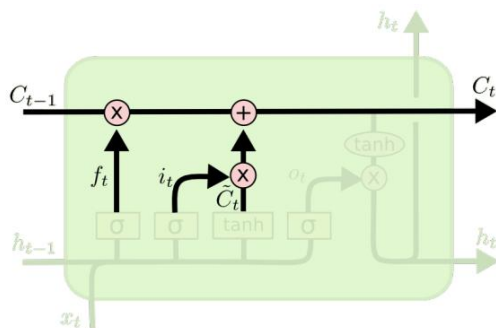
Fig. II. 9 : Opérateur d'ajout d'informations

Tel que :

- **tanh** : Fonction d'activation tangente hyperbolique.
- **$\tilde{C}_t$** : Valeur candidate .

**c. Mise à jour des états des cellules**

Maintenant, mettre à jour l'ancien état de la cellule,  $C_{t-1}$ , dans le nouvel état de la cellule  $C_t$ . Nous multiplions l'ancien état par  $f_t$ . Ensuite, nous l'ajoutons  $i_t * \tilde{C}_t$ . Il s'agit des nouvelles valeurs candidates, mises à l'échelle en fonction de l'importance de la mise à jour de chaque valeur d'état.



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

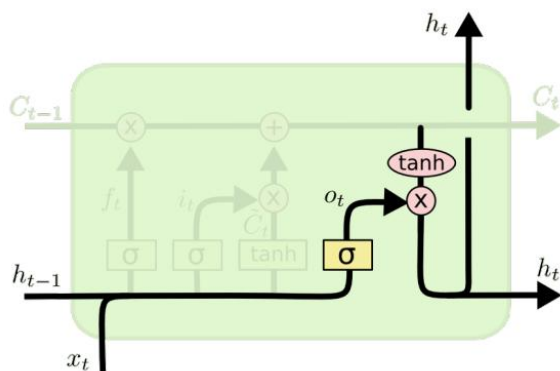
Fig. II. 10 : Mise à jour des états des cellules

Tel que :

- **$C_t$** : État interne .

## d. Générer une sortie

Enfin, décider ce que va être produit. Cette sortie sera basée sur l'état de notre cellule  $C_t$ , mais sera une version filtrée. Tout d'abord, nous exécutons une couche sigmoïde qui décide quelles parties de l'état cellulaire nous allons générer. Ensuite, nous mettons l'état de la cellule via  $\tanh$  et le multiplions par la sortie de la porte sigmoïde  $o_t$ .



$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

Fig. II. 11 : Générer la sortie.

Tel que :

- $h$ : Sortie

### II.3.4. Naïve Bayes

L'algorithme naïf de Bayes est basé sur le théorème de Bayes développé par Thomas Bayes au XVIIIe siècle. Le théorème de Bayes fournit une méthode mathématique pour mettre à jour les probabilités sur la base de nouvelles preuves. Ils sont largement utilisés pour leur simplicité et leur efficacité dans l'apprentissage automatique.

#### Définition

Naïve Bayes est une famille de « classificateurs probabilistes » simples basés sur l'application du théorème de Bayes avec de fortes hypothèses d'indépendance (naïve) entre les caractéristiques. En d'autres termes, un classificateur de Bayes naïf suppose que la présence d'une caractéristique particulière dans une classe n'est pas liée à la présence d'une autre caractéristique.

Les modèles de Naïve Bayes sont faciles à construire et particulièrement utiles pour les très grands ensembles de données. Outre leur simplicité, les modèles Naive Bayes sont connus pour leurs performances supérieures à celles des méthodes de classification les plus

sophistiquées. Il s'agit de l'une des approches les plus pratiques pour certains types de problèmes, notamment la classification de documents et le filtrage de spam [49]

### **Théorème de Bayes**

Le théorème de Bayes (alternativement la loi de Bayes ou la règle de Bayes) décrit la probabilité d'un événement, basée sur la connaissance préalable des conditions qui pourraient être liées à l'événement. La formule est comme suit :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Dont :

- **P(A|B)** : la probabilité conditionnelle que l'évènement A se produise, étant donné que B s'est produit. Ceci est également connu comme la probabilité postérieure.
- **P(B|A)** : la probabilité conditionnelle que l'évènement B se produise, étant donné qu'A s'est produit.
- **P(A) et P(B)** : probabilité de A et B sans égard l'un à l'autre.

### **Avantage et Inconvénient de Naïves Bayes**

Dans ce qui suit, on va voir quelques avantages et inconvénients de la méthode naïve Bayes [49]:

#### **Avantages**

- Facile à mettre en œuvre et efficace sur le plan des calculs.
- Efficace dans les cas où il existe un grand nombre de caractéristiques.
- Il donne de bons résultats même avec des données de formation limitées.
- Il donne de bons résultats en présence de caractéristiques catégorielles.
- Pour les caractéristiques numériques, les données sont supposées provenir de distributions normales.

#### **Inconvénients**

- Suppose que les caractéristiques sont indépendantes, ce qui n'est pas toujours le cas dans les données réelles.

- Peut être influencé par des attributs non pertinents.
- Peut attribuer une probabilité nulle à des événements non vus, ce qui conduit à une mauvaise généralisation.

#### **II.4. Conclusion**

En conclusion, le Deep Learning reste l'une des technologies les plus marquantes dans le domaine de l'intelligence artificielle. Nous avons expliqué le concept de Deep Learning avec certaines de ses caractéristiques et domaines d'application, puis nous avons expliqué les réseaux de neurones récurrents (RNN), les réseaux Long Short-Term Memory (LSTM) et le modèle de naïve bayes.

# *Chapitre III*

## *Conception du système*



### **III.1. Introduction**

Dans ce chapitre, nous fournirons une description complète de notre système, en nous concentrant sur l'aspect conceptuel du traitement et la mise en œuvre du système.

### **III.2. Architecture générale**

Le système de classification à développer est conçu pour le suivi des rumeurs en utilisant des techniques de Deep Learning et machine Learning. Il prend en entrée une base brute de tweets et de leurs caractéristiques, puis les transforme en une base de données utilisable pour l'apprentissage. Ce processus de transformation, appelé prétraitement, comprend diverses opérations telles que le nettoyage, le filtrage et l'encodage. La base de données prétraitée est ensuite divisée en deux parties : une pour l'entraînement et une pour les tests.

Le module d'entraînement utilise la partie d'entraînement de la base de données et un algorithme d'apprentissage pour créer un modèle de décision. Ce modèle est ensuite appliqué à la partie de test de la base de données. Si le modèle atteint un niveau de performance acceptable, c'est-à-dire s'il peut reconnaître efficacement les rumeurs, il est conservé pour une utilisation future par le module de suivi des rumeurs. Dans le cas contraire, les paramètres de l'algorithme d'apprentissage sont révisés dans le but d'améliorer sa capacité à détecter les rumeurs.

Ce processus itératif permet d'améliorer continuellement la capacité du système à suivre et à identifier les rumeurs de manière précise et efficace, fournissant ainsi une assistance précieuse dans la gestion de l'information et la prévention de la propagation de rumeurs.

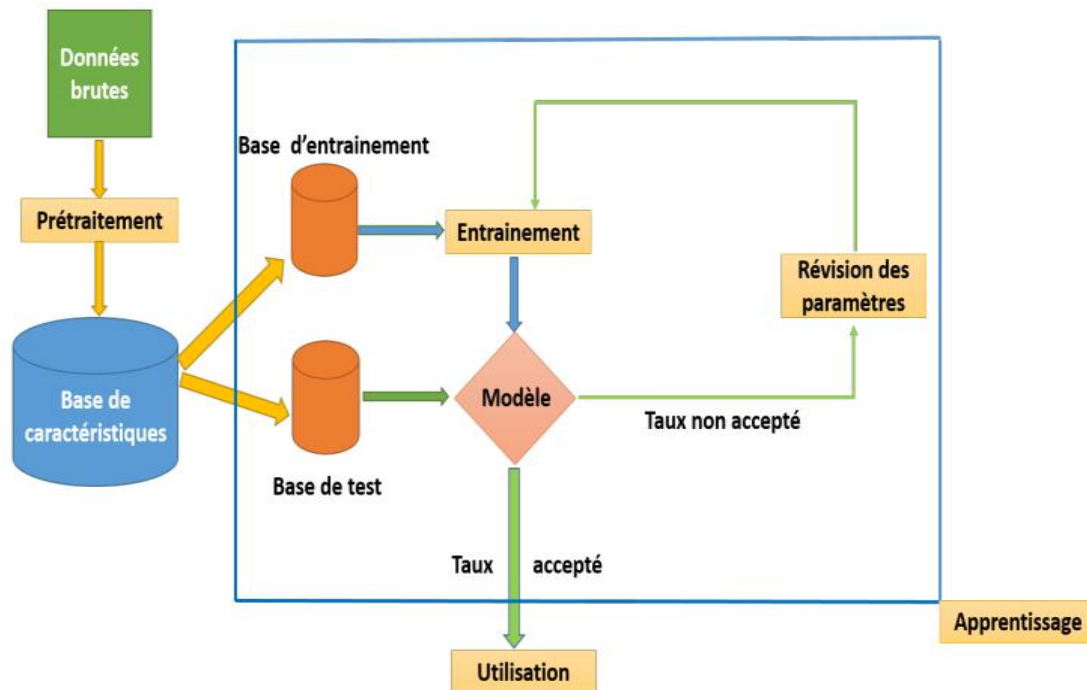


Fig. III. 1 : Architecture générale

### III.3. Architecture détaillée

Dans ce qui suit, nous détaillons chacune des phases de notre système.

#### III.3.1. Pré-traitement

Ici, l'objectif est d'extraire les meilleures caractéristiques permettant de détecter les rumeurs. On commence par le prétraitement des données du *dataset* brute qui sont [50]:

##### a. Importation des bibliothèques

Le processus décrit l'importation de bibliothèques nécessaires à la manipulation de données, au prétraitement, à la modélisation, et au traçage graphique. Les données sont chargées à partir d'un fichier *CSV* en utilisant la méthode `pandas.read_csv()`, où le chemin du fichier est spécifié par `dataset_path`. Pour visualiser et comprendre la structure du *dataset*, les premières lignes sont affichées avec la méthode `df.head()`.

```
1 import seaborn as sns
2 import pandas as pd
3 from sklearn.preprocessing import StandardScaler
4 from sklearn.feature_extraction.text import TfidfVectorizer
5 import re
6 import string
7 import numpy as np
8 from sklearn.model_selection import train_test_split
9 from tensorflow.keras.models import Sequential
10 from tensorflow.keras.layers import Embedding, LSTM, Dense, GlobalMaxPooling1D
11 from tensorflow.keras.preprocessing.text import Tokenizer
12 from tensorflow.keras.preprocessing.sequence import pad_sequences
13 from sklearn.metrics import precision_score, f1_score
14 import matplotlib.pyplot as plt
15 # Charger l'ensemble de données à partir d'un fichier CSV
16
17 dataset_path = 'dataset.csv'
18 df = pd.read_csv(dataset_path)
19
20 # Imprimer les premières lignes de l'ensemble de données
21 print("Original dataset:")
22 df.head()
```

Fig. III. 2 : Bibliothèques importées

### b. Extraction de caractéristiques et de cibles à partir d'un dataset

Le passage décrit la préparation d'un ensemble de données pour la classification en utilisant Python et pandas, où  $X$  contient les caractéristiques en excluant la colonne *is\_rumor* du *DataFrame* *df*, et  $y$  est la variable cible composée des valeurs de la colonne `is_rumor`. Cette étape est cruciale pour appliquer des modèles prédictifs en apprentissage automatique.

```
1 # Extraire les caractéristiques (X) et la variable cible (y) de l'ensemble de données
2 X = df.drop(columns=['is_rumor']) # Assuming 'target' is the column to be predicted
3 y = df['is_rumor']
4
```

Fig. III. 3 : Extraction de caractéristiques et de cibles à partir d'un dataset

### c. Nettoyage des données

Le nettoyage des données fait référence au processus de prétraitement et de transformation des données brutes afin de les rendre aptes à une analyse ultérieure, telle que l'analyse descriptive (visualisation des données) ou l'analyse prescriptive (élaboration d'un modèle). Des données propres, précises et fiables doivent être utilisées pour la post-analyse, car "de mauvaises données conduisent à de mauvais modèles prédictifs".

1. Convertir le texte en minuscules .
2. Supprimer le texte entre crochets .

3. Remplacer les caractères autres que les mots (à l'exception des espaces) par un espace .
4. Supprimer les URL .
5. Supprimer les balises HTML .
6. Supprimer la ponctuation .
7. Supprimer les caractères de nouvelle ligne .
8. Supprimer les mots contenant des chiffres .

```
1 def wordopt(text):
2     text = text.lower()
3     text = re.sub('[\.*?\\]', ' ', text)
4     text = re.sub("\\W", " ", text)
5     text = re.sub('https?://\S+|www.\S+', ' ', text)
6     text = re.sub('<.*?>+', ' ', text)
7     text = re.sub('[%s]' % re.escape(string.punctuation), ' ', text)
8     text = re.sub('\n', ' ', text)
9     text = re.sub('\w*\d\w*', ' ', text)
10    return text
11
12 X["text"] = X["text"].apply(wordopt)
13
```

**Fig. III. 4 :** Nettoyage des données

#### d. Vectorization

La vectorisation accélère le code Python en remplaçant les boucles par des opérations vectorielles. Le *TfidfVectorizer* de *sklearn* convertit des textes en une matrice *TF-IDF*, facilitant le traitement des données. Les données transformées et les étiquettes sont affichées pour vérification, et le nombre de valeurs nulles est examiné pour garantir la qualité des données avant toute analyse ou apprentissage [51] .

```

1 vectorization = TfidfVectorizer()
2 X_vect=vectorization.fit_transform(X)
3 print(X)
4 print(X_vect)
5 print(y)
6 null_values = y.isnull().sum()
7 print("Total null values in y the DataFrame:", null_values)
8 null_values = X.isnull().sum()
9 print("Total null values in X the DataFrame:", null_values)

```

	text	user.handle	\
0	charlie hebdo became well known for publishing...	BBCDaniels	
1	now dead in a shooting there today rt bbcda...	robbylevy	
2	bbcdaniels bbcworld i m guessing this is bei...	ModerateInAll	
3	bbcdaniels bbcworld why would you mention th...	GabTarquini	
4	bbcdaniels bbcworld perps identified	freethought41	
...	...	...	
62440	anyonops xplant so that means its ok to torc...	RianAlden	
62441	rianalden not at all but they need to change...	Xplant	
62442	xplant anyonops absoluteky but it pains me...	RianAlden	
62443	xplant anyonops i m curious how many of thes...	RianAlden	
62444	xplant anyonops you get people showing up ...	RianAlden	

	topic
0	charliehebdo
1	charliehebdo
2	charliehebdo
3	charliehebdo
4	charliehebdo

Fig. III. 5 : Vectorisation de données

#### e. Combinaison

Cet extrait de code prépare les données pour l'apprentissage automatique en fusionnant, nettoyant et reséparant un ensemble de données. Il combine d'abord les caractéristiques  $X$  et la variable cible  $y$ , supprime les lignes contenant des valeurs nulles dans les colonnes clés, puis sépare à nouveau les caractéristiques et la cible. Enfin, il vérifie et affiche les dimensions nettoyées de  $X$  et  $y$ , assurant ainsi l'intégrité des données pour l'apprentissage d'un modèle efficace.

```

1 # Combiner X et y en un seul DataFrame
2 combined_df = pd.concat([X, y], axis=1)
3
4 # Supprimer les lignes contenant des valeurs nulles dans y et text
5 combined_df.dropna(subset=['is_rumor', 'text'], inplace=True)
6
7 # Séparer X et y après avoir supprimé les valeurs nulles
8 X = combined_df.drop(columns=['is_rumor'])
9 y = combined_df['is_rumor']
10
11 # Vérifier si les valeurs nulles ont été supprimées
12 print("Shape after removing null values:", X.shape, y.shape)
13 null_values = y.isnull().sum()
14 print("Total null values in y the DataFrame after removing:", null_values)
15 null_values = X.isnull().sum()
16 print("Total null values in X the DataFrame after removing:", null_values)

```

```

↳ Shape after removing null values: (62443, 3) (62443,)
Total null values in y the DataFrame after removing: 0
Total null values in X the DataFrame after removing: text      0
user.handle          0
topic                12777
dtype: int64

```

Fig. III. 6 : Combinaison de données

#### f. Division des données

La méthode *split* () en Python divise une chaîne de caractères en une liste de mots en utilisant les espaces blancs comme séparateurs par défaut. Elle peut également limiter le nombre de séparations. Le code utilise la fonction *train\_test\_split* de *sklearn.model\_selection* pour diviser le jeu de données en ensembles d'entraînement et de test, attribuant 80% des données à l'entraînement et 20% au test. Cette division permet d'évaluer les performances du modèle sur des données inconnues [52].

```

[20] 1 # Diviser l'ensemble de données en deux ensembles, l'un pour la formation et l'autre pour le test
      2 X_train, X_test, y_train, y_test = train_test_split(X["text"], y, test_size=0.2, random_state=42)

```

Fig. III. 7 : Division de données

#### g. Tokenisation

La tokenisation, fondamentale en traitement du langage naturel (NLP) et en apprentissage automatique, consiste à découper un texte en *tokens* plus petits comme des mots ou des caractères. Ce processus, réalisé avec la classe *Tokenizer* de *keras.preprocessing.text*, limite le vocabulaire aux 10 000 mots les plus fréquents. Les données d'entraînement sont ajustées au *tokenizer* avec *fit\_on\_texts*, puis transformées en séquences d'entiers avec *texts\_to\_sequences*. Cette approche assure une

représentation cohérente et numérique des données textuelles, facilitant leur traitement par les modèles [53].

```
[21] 1 # Tokéniser les données textuelles et les convertir en séquences d'entiers
      2 max_words = 10000 # Définir le nombre maximum de mots à conserver dans le vocabulaire
      3 tokenizer = Tokenizer(num_words=max_words)
      4 tokenizer.fit_on_texts(X_train)
      5 X_train_seq = tokenizer.texts_to_sequences(X_train)
      6 X_test_seq = tokenizer.texts_to_sequences(X_test)
```

Fig. III. 8 : Tokenisation de textes

#### h. Remplissage des séquences

Le remplissage des séquences est une étape cruciale de prétraitement en traitement du langage naturel et en apprentissage profond, visant à uniformiser la longueur des séquences d'entrée. Cette technique, réalisée avec la fonction *pad\_sequences* de *Keras*, ajoute des zéros pour compléter les séquences plus courtes et tronque les séquences plus longues. Ainsi, les données d'entraînement et de test sont transformées en versions uniformes, essentielles pour les modèles d'apprentissage profond nécessitant des entrées de longueur constante.

```
[22] 1 # Les séquences doivent être protégées par un tampon afin d'assurer une longueur uniforme.
      2 max_sequence_length = 100 # Définir la longueur maximale de la séquence
      3 X_train_padded = pad_sequences(X_train_seq, maxlen=max_sequence_length)
      4 X_test_padded = pad_sequences(X_test_seq, maxlen=max_sequence_length)
```

Fig. III. 9 : Remplissage des séquences

### III.3.2. L'apprentissage

L'apprentissage automatique et l'apprentissage profond sont des sous-ensembles de l'intelligence artificielle qui permettent aux systèmes d'apprendre à partir de données et de prendre des décisions. L'apprentissage automatique utilise des algorithmes pour analyser les données et en tirer des enseignements, tandis que l'apprentissage profond utilise des réseaux neuronaux en couches pour analyser différents niveaux de caractéristiques des données.

Les algorithmes clés dans ces domaines comprennent la mémoire à long terme (LSTM) et les réseaux neuronaux récurrents (RNN), qui sont capables de traiter des séquences de données pour des tâches telles que le traitement du langage. En outre, l'algorithme de Naïve Bayes offre une approche plus simple pour la modélisation prédictive et la classification, particulièrement efficace pour les grands ensembles de données textuelles. Ces technologies permettent diverses applications, de la reconnaissance vocale à l'analyse des sentiments, améliorant la capacité des systèmes à effectuer des tâches complexes de manière autonome.

Au cours de cette phase d'entraînement, le modèle apprend à comprendre les modèles et les relations au sein des données. Cela implique une préparation des données, où les données sont nettoyées, normalisées et divisées en caractéristiques (entrées) et en étiquettes (sorties). L'architecture du modèle est définie, ce qui inclut la configuration des couches et des fonctions d'activation pour les modèles d'apprentissage profond ou le choix de l'algorithme approprié pour les modèles d'apprentissage automatique traditionnels, comme Naïve Bayes. Le modèle subit ensuite une propagation vers l'avant, où les données d'entrée sont transmises au modèle pour produire des sorties. Les sorties sont comparées aux étiquettes réelles à l'aide d'une fonction de perte, et l'erreur est rétropropagée (dans l'apprentissage profond) pour mettre à jour les poids du modèle. Ce processus est optimisé sur plusieurs itérations ou époques afin de minimiser la perte.

### **II.3.3. Test**

Dans la phase de test, l'objectif est d'évaluer les performances du modèle formé sur de nouvelles données inédites afin de s'assurer qu'il peut bien se généraliser. Les données de test sont prétraitées de la même manière que les données de formation, et le modèle formé est appliqué à ces données de test pour faire des prédictions. Des mesures de performance telles que l'exactitude, la précision, le rappel et le score F1 sont calculées pour évaluer l'efficacité du modèle. Par exemple, dans l'apprentissage profond avec les RNN ou les LSTM, des données séquentielles telles que du texte ou des séries temporelles sont préparées, et le modèle est formé à l'aide de la rétropropagation dans le temps (BPTT). Les performances du modèle sont ensuite évaluées sur un ensemble de séquences test à l'aide de mesures telles que la perplexité ou l'erreur quadratique moyenne. Dans l'apprentissage automatique traditionnel avec Naïve Bayes, le modèle calcule les probabilités et les vraisemblances préalables pendant la formation et est ensuite évalué sur un ensemble de données distinct à l'aide de mesures telles



---

que la précision et le rappel. Les phases de formation et de test sont toutes deux cruciales pour développer des modèles robustes qui fonctionnent bien sur de nouvelles données.

### **III.3.4. Utilisation**

C'est la dernière étape et la plus importante de notre système. Après avoir atteint le meilleur taux de reconnaissance ou après avoir construit le meilleur modèle, le modèle nous permet de reconnaître les rumeurs, qu'elles soient fausses ou vraies.

Nous nous sommes concentrés sur la mise en œuvre de divers algorithmes d'apprentissage automatique afin de suivre et d'analyser efficacement les rumeurs. Le cœur du système s'articule autour de l'exploitation de techniques d'apprentissage profond, en particulier les réseaux neuronaux récurrents (RNN) et les réseaux de mémoire à long terme (LSTM), ainsi que les algorithmes traditionnels d'apprentissage automatique tels que Naïve Bayes. Les RNN et les LSTM sont particulièrement bien adaptés à cette tâche en raison de leur capacité à traiter des données séquentielles et à saisir les dépendances temporelles, ce qui les rend idéaux pour comprendre l'évolution et la propagation des rumeurs au fil du temps. Ces modèles peuvent analyser des modèles dans les données textuelles, en identifiant les indices linguistiques et les informations contextuelles qui indiquent une rumeur. En outre, l'algorithme Naïve Bayes, connu pour sa simplicité et son efficacité, complète les modèles d'apprentissage profond en fournissant un cadre probabiliste pour la classification des textes basée sur la fréquence des mots. Ensemble, ces algorithmes forment un système puissant capable de détecter, de suivre et d'analyser les rumeurs avec une grande précision, fournissant ainsi des informations précieuses pour atténuer la propagation de la désinformation.

### **III.4. Conclusion**

Le système conçu dans ce mémoire se compose de trois phases : la phase de prétraitement permettant de préparer les données initiales, la phase d'apprentissage permettant de former le modèle de prédiction pour améliorer les résultats, et la phase de prédiction permettant de tester le modèle avec des données réelles. Nous avons passé en détail dans ce chapitre par les différents constituants du système proposé.

Dans le chapitre suivant, nous allons décrire le fonctionnement de notre application mettant en œuvre le système proposé.

# *Chapitre IV*

## *Implémentation et résultats*

## IV.1. Introduction

Dans ce chapitre, nous allons implémenter et comparer les deux algorithmes, RNN et ISTM, pour la classification de texte. Montrez les outils (programmes, langages, bibliothèques, données) utilisés dans notre système. Ensuite, nous présentons et discutons les résultats obtenus.

## IV.2. Présentation des outils utilisés

### IV.2.1. Configuration matérielle

Ce travail a été implémenté sur un PC, caractérisé comme suit :

- **Processeur** : Intel(R) Celeron(R) CPU N2940 @ 1.83GHz 1.83GHz
- **RAM** : 4.00 GB
- **Système d'exploitation** : Windows 10 Professionnel 64 bits

### IV.2.2. Environnement logiciel

#### Google Colab

Collaborative ou « Colab » est une plateforme gratuite pour des sessions d'une durée limitée sur le Web (environ 12 heures/jour). C'est un produit développé par l'équipe de recherche de Google. Il vous donne la possibilité d'écrire et de mettre en œuvre n'importe quoi. Code Python que vous choisissez via le navigateur. Il utilise l'environnement « Jupyter Notebook », destiné à la formation et à la recherche dans le domaine de l'apprentissage automatique pour diverses routines de programmation en Python [54].

Colab offre les avantages suivants :

- Développer des compétences en programmation dans le langage de programmation Python.
- Ce système vous donne la possibilité de former des modèles de machine Learning (ML) directement dans le Cloud, sans avoir à configurer l'environnement de développement (Jupyter Notebook). Mais la fonctionnalité qui distingue Colab des autres services est l'accès gratuit à l'unité de traitement graphique (GPU).

- Les applications d'apprentissage profond sont développées à l'aide de bibliothèques Python populaires telles que Keras, TensorFlow, PyTorch et OpenCV.

### IV.2.3. Langages utilisés

#### Python

Ces dernières années, Python est devenu le langage de programmation le plus utilisé par les informaticiens. Il a pris la tête dans la gestion d'infrastructure, l'analyse de données et le développement de logiciels. Python permet aux développeurs de se concentrer sur leur travail plutôt que sur les détails techniques. Il a libéré les développeurs des contraintes imposées par les anciens langages. En conséquence, le développement de code avec Python est plus rapide qu'avec d'autres langages [55].

### IV.2.4. Bibliothèques utilisées

Dans cet environnement on a installé des packages qui nous ont facilité la programmation, qui sont :

#### Pandas

Cette bibliothèque écrite pour le langage de programmation Python pour la manipulation et l'analyse de données. En particulier, il propose des structures de données et des opérations pour manipuler des tableaux numériques et des séries chronologiques. Pandas est un outil d'analyse et de manipulation de données open source rapide, puissant, flexible et facile à utiliser. [56]

#### NumPy

Le package fondamental nécessaire au calcul scientifique avec Python. Outre ses utilisations scientifiques évidentes, NumPy peut également être utilisé comme conteneur de données génériques multidimensionnelles efficace. Des types de données arbitraires peuvent être définis. Cela permet à NumPy de s'intégrer de manière transparente et rapide à une grande variété de bases de données. [57]

### **TensorFlow**

Il s'agit d'une plateforme ouverte créée par l'équipe Google Brain pour rechercher dans le domaine de la Machine Learning. Il fournit un écosystème complet et flexible d'outils, de bibliothèques et de ressources communautaires qui permettent aux chercheurs de faire progresser le domaine de l'apprentissage automatique et aux développeurs de créer et de déployer facilement des applications utilisant cette technologie. [58]

### **Keras**

C'est une API de réseau neuronal écrite en Python. C'est une bibliothèque qui donne accès aux fonctions proposées par de nombreuses bibliothèques dédiées au Deep Learning, notamment TensorFlow et Theano. Il dispose également de fonctions et de procédures qui facilitent le traitement associé au Deep Learning.

### **Matplotlib**

Une bibliothèque du langage de programmation Python destinée à tracer et visualiser les données et les résultats de traitements sous forme de graphiques. [59]

### **Sklearn**

Une bibliothèque d'apprentissage automatique gratuite et open source pour le langage de programmation Python. Il contient de nombreux outils puissants pour l'apprentissage automatique et la modélisation statistique, notamment la classification, la régression et le clustering, et est conçu pour s'interfacer avec les bibliothèques Python numériques et scientifiques NumPy et SciPy.

#### **IV.2.5. Ensemble des données**

Notre objectif est de développer un programme d'apprentissage profond pour suivre les rumeurs. Le modèle se concentrera sur l'identification des rumeurs, sur la base de plusieurs tweets. Cet ensemble de données de classification des tweets contient un ensemble de 62 445 lignes de commentaires dans lesquelles les informations sont étiquetées 1 ou 0 pour les rumeurs vraies ou fausses, respectivement.

Tab.IV. 1 : Description de dataset

Les données	Étiquette	Nombre	Type
Données fausses	0	48619	Texte anglais
Données vraies	1	13824	Texte anglais
Total	/	62445	Texte anglais

La figure suivante montre le nombre de fausses rumeurs en bleu et de vraie information en rouge :

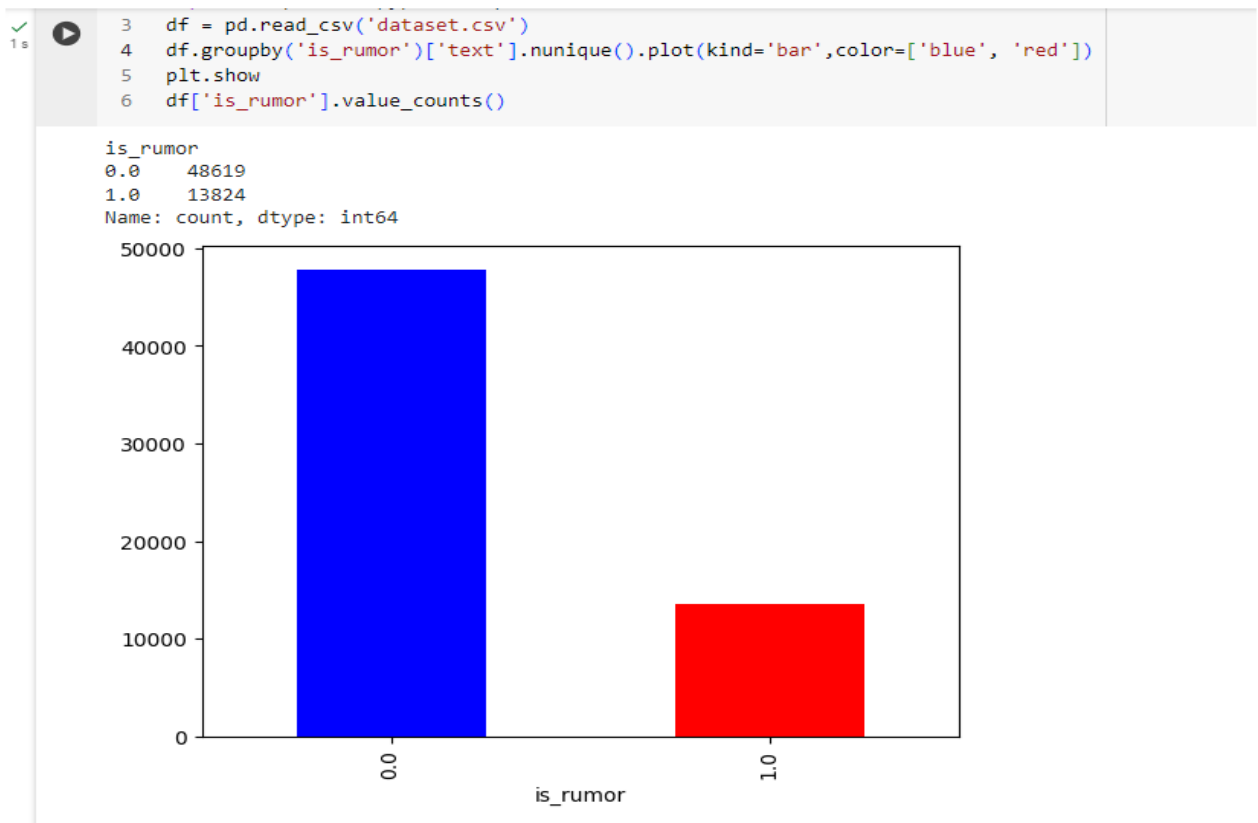


Fig.IV. 1 : Représentation des classes de l'ensemble de données

Il s'agit d'une description d'un ensemble de données textuelles. L'ensemble de données se compose d'attributs suivants :

Tab.IV. 2 : Description des colonnes de dataset

<b>Text</b>	Commentaires des utilisateurs de Twitter.
<b>is_rumor</b>	Il permet de vérifier si un texte est une rumeur ou non.
<b>user.handle</b>	Utilisateur qui a envoyé le tweet.
<b>Topic</b>	Le sujet sur lequel les données ont été collectées ; dans ce cas, c'est Charlie Hebdo et le crash de l'aile allemande.

Dans la fig. III.2, nous représentons la data frame :

Original dataset:

	text	is_rumor	user.handle	topic
0	Charlie Hebdo became well known for publishing...	0.0	BBCDaniels	charliehebdo
1	Now 10 dead in a shooting there today RT "@BBC...	0.0	robbylevy	charliehebdo
2	@BBCDaniels @BBCWorld I'm guessing this is bei...	0.0	ModerateInAll	charliehebdo
3	@BBCDaniels @BBCWorld why would you mention th...	0.0	GabTarquini	charliehebdo
4	@BBCDaniels @BBCWorld perps identified?	0.0	freethought41	charliehebdo
...	...	...	...	...
62440	@AnonyOps @Xplant So that means its ok to torc...	1.0	RianAlden	ferguson
62441	@RianAlden not at all, but they need to change...	1.0	Xplant	ferguson
62442	@Xplant @AnonyOps Absoluteky. But it pains me...	1.0	RianAlden	ferguson
62443	@Xplant @AnonyOps I'm curious how many of thes...	1.0	RianAlden	ferguson
62444	@Xplant @AnonyOps You get 15,000 people showin...	1.0	RianAlden	ferguson

62445 rows × 4 columns

Fig.IV. 2 : Quelques lignes de dataset

### IV.3. Métriques d'évaluation utilisés

Pour évaluer les algorithmes de classification, nous avons utilisé quatre mesures d'évaluation : l'accuracy, la précision, le score F1 et le rappel. Chaque formule est basée sur une matrice de confusion.

La matrice de confusion donne une représentation matricielle. Les classes cibles individuelles sont prédites par le modèle. Comme dans le cas de notre implémentation, nous distinguons 4 cas de classification possibles.

- **Vrai positif (VP)** : Instances correctement prédites comme positives.
- **Vrai négatif (VN)** : Instances correctement prédites comme négatives.
- **Faux positif (FP)** : Instances qui ont été incorrectement prédites comme positives.
- **Faux négatif (FN)** : Instances qui ont été incorrectement prédites comme négatives.

**Tab.IV. 3.**Principe de la matrice de confusion

	Positive	Négative
Positive	VP	FP
Négative	FN	VN

**1- Accuracy**

La précision est le rapport entre les éléments bien classés et le nombre total d'exemples. La précision est calculée à partir de la matrice de confusion :

$$\text{Accuracy} = \frac{VP+VN}{FP+FN+VP+VN}$$

**2- Précision**

La précision d'un modèle évalue la pertinence des éléments détectés en comparaison avec tous les éléments identifiés comme positifs. Elle est obtenue en divisant le nombre de vrais positifs par le nombre total de positifs détectés.

Pour calculer la précision, nous utiliserons la formule ci-dessous :

$$\text{Precision} = \frac{VP}{FP+VP}$$



### 3- Recall

Le rappel mesure l'efficacité d'un modèle de classification à identifier toutes les instances pertinentes d'un ensemble de données. Il s'agit du rapport entre le nombre de vraies instances positives (VP) et la somme des vraies instances positives et des fausses instances négatives (FN).

### 4- F1-score

.Le F1-score est une mesure qui combine à la fois la précision et le rappel (Recall) d'un modèle. Il est calculé comme une moyenne pondérée des scores de précision et de rappel. En d'autres termes, le F1 score équilibre la contribution de la précision et du rappel pour évaluer la performance globale du modèle. Nous pouvons définir le F1 score comme ci-dessous (P=précision, R=Recall) :

$$\text{F1-Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

## III.5. Résultats et discussion

Dans cette section, les résultats d'exécution des modèles implémentés, ainsi qu'une comparaison des performances de ces modèles, vont être montré et analysé :

**Tab.IV.** 4. Résultats de l'algorithme RNN.

Classe	Précision	Recall	F1-Score	Accuracy	Support
Négative	0.68	0.10	0.18		9754
Positive	0.80	0.99	0.88		2735
Avg/ Total				0.79	12489

Les résultats sont présentés en considérons les quatre classifieur construisant dans la phase d'apprentissage. La Table IV.4 montre les résultats pour le classifieur RNN.

Alors que la Table IV.5 montre les résultats pour le classifieur LSTM.

Tab.IV. 5. Résultats de l'algorithme LSTM.

Classe	Précision	Recall	F1-Score	Accuracy	Support
Négative	0.80	0.98	0.88		9754
Positive	0.64	0.10	0.18		2735
Avg/ Total				0.83	12489

Les résultats du classifieur Naïve bayes sont affichés dans la Table IV.6.

Tab.IV. 6. Résultats de l'algorithme Naïve bayes.

Classe	Précision	Recall	F1-Score	Accuracy	Support
Négative	0.83	0.99	0.90		9754
Positive	0.92	0.26	0.41		2735
Avg/ Total				0.83	12489

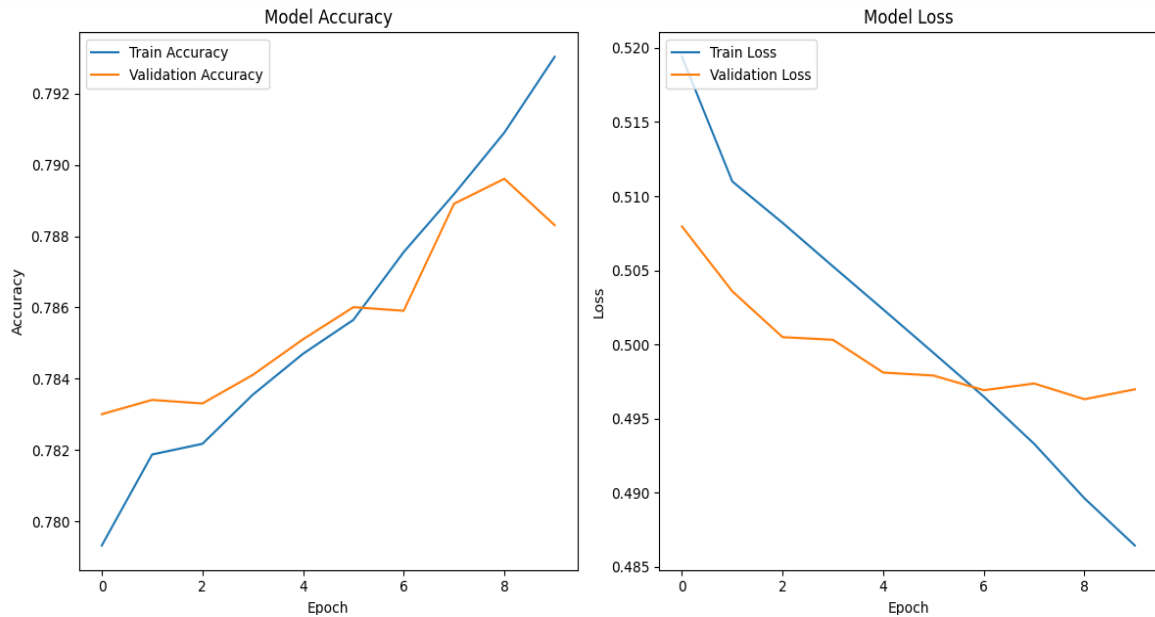
Performances de formation d'un modèle de réseau neuronal récurrent (RNN).

La figure suivante montre les résultats de la formation du modèle RNN pour la classification des rumeurs.

Le modèle semble avoir atteint une précision de test de 79,05 %.

Dans le graphique, il apparaît que la précision de l'apprentissage et la perte diminuent, ce qui est bon signe. Cela indique que le modèle apprend des données d'apprentissage et améliore ses performances.

Cela indique que le modèle RNN a réussi à apprendre à classer les rumeurs avec un degré de précision assez élevé.

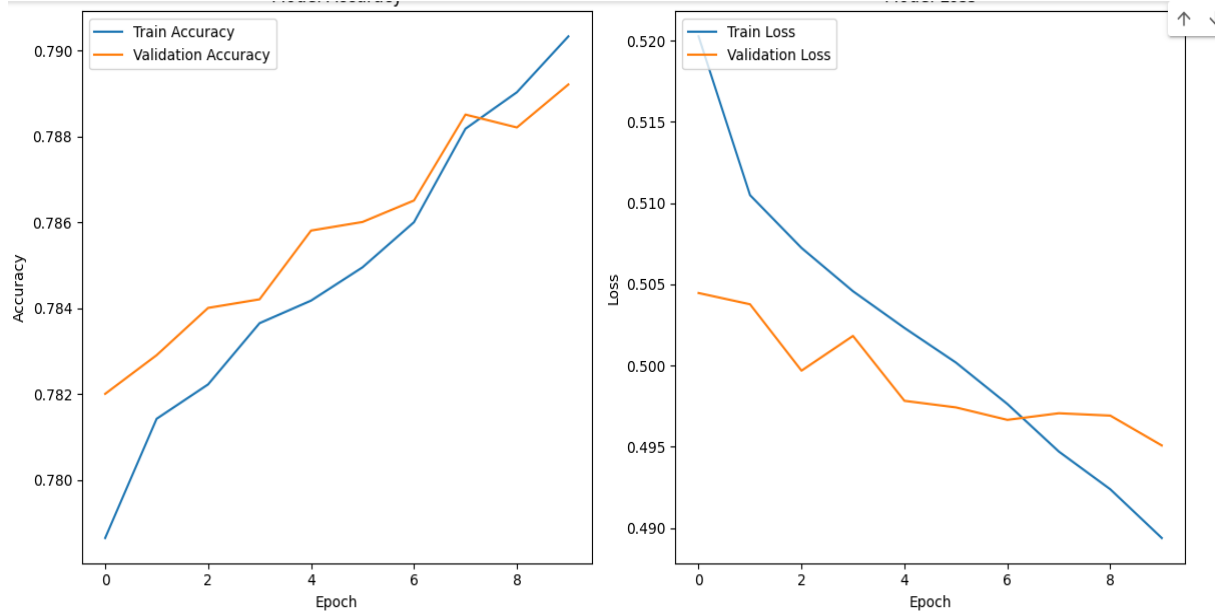


**Fig.IV. 3** : Résultat Accuracy et Loss de modèle RNN

Les deux graphiques de l'image montrent la précision de la formation et de la validation à l'aide du modèle d'apprentissage profond avec l'algorithme LSTM

- La précision d'un modèle reflète ses performances sur un ensemble spécifique de données.
- La perte de modèle représente la différence entre les valeurs prédites et les valeurs réelles.

Le modèle présente de bonnes performances sur les données d'apprentissage, avec une précision d'apprentissage d'environ 0,79. Mais la précision de validation est plus faible, environ 0,78, ce qui peut indiquer que le modèle s'adapte bien aux données d'apprentissage. Cela signifie que le modèle n'est performant que sur les données sur lesquelles il a été entraîné.



**Fig.IV. 4 :** Résultat Accuracy et Loss de modèle LSTM

La figure suivante représente une matrice de confusion, qui est un outil de visualisation utilisé dans l'apprentissage automatique pour évaluer les performances d'un modèle de classification. Elle indique le nombre de fois où le modèle a catégorisé correctement ou incorrectement des points de données.

Cette figure représente un outil d'évaluation du modèle de classification développé, tel que les lignes représentent les classifications réelles (rumeur ou non rumeur) et les colonnes représentent les classifications prédites. Les cellules diagonales (bleues) indiquent le nombre de classifications correctes, tandis que les cellules hors diagonale (jaunes et vertes) indiquent le nombre de classifications incorrectes. En analysant la matrice de confusion, nous avons constaté que l'algorithme de Naïve Bayes donne les meilleurs résultats.

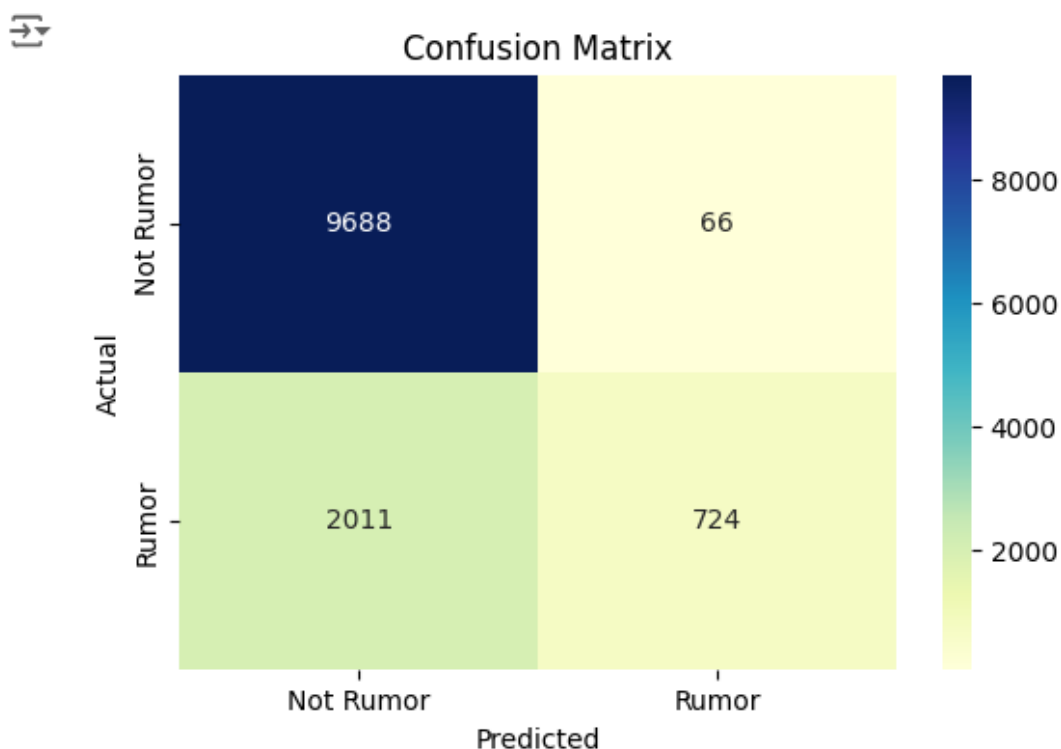


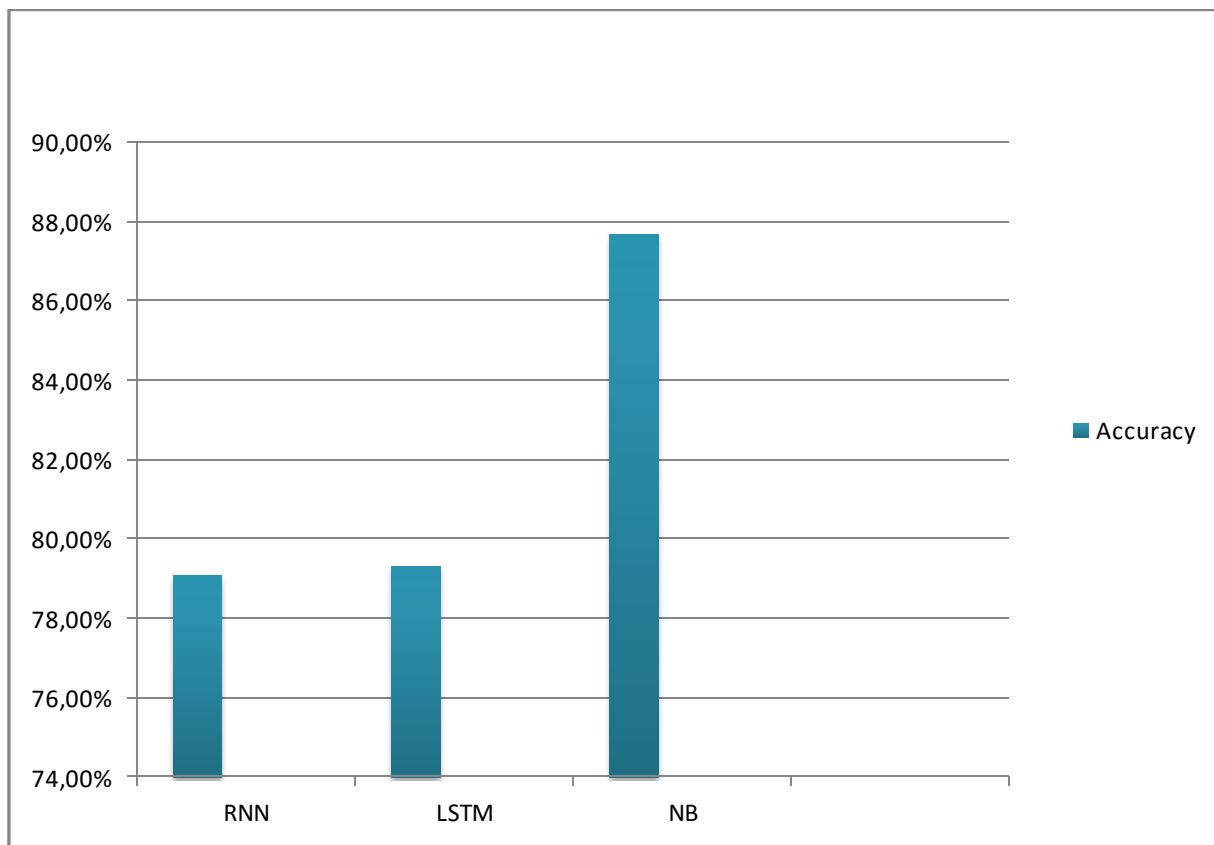
Fig.IV. 5. Résultat de modèle Naïve bayes

IV.5.1. Comparaison de résultat des modèles implémentés

Nous comparerons les meilleurs résultats des algorithmes d'apprentissage profond obtenus ci-dessus avec les algorithmes d'apprentissage automatique et proposerons le meilleur algorithme en fonction « accuracy » de la classification.

Tab.IV. 7 : Comparaison des résultats

Le modèle	L' accuracy
RNN ( Recurrent neural network)	79.09%
LSTM (Long Short Term Memory)	79.33 %
NB (Naïve Bayes)	87.69%



**Fig.IV. 6 :** Résultat de comparaison d'algorithme

De ces résultats, nous remarquons que l'algorithme NB a obtenu le meilleur résultat de réussite de 83,37 %. Il convient toutefois de noter que les algorithmes RNN et LSTM ont également obtenu des scores élevés, supérieurs à 79 %. Par conséquent, pour choisir le meilleur algorithme en fonction de la précision, il est recommandé de considérer l'algorithme NB comme le premier choix, suivi des algorithmes LSTM et RNN.

## IV.6. Conclusion

Ce dernier chapitre contient une description générale des moyens utilisés pour développer le système de classification basé sur l'apprentissage profond à savoir le modèle LSTM pour la détection des rumeurs. Ainsi que, une étude comparative entre ce modèle et les modèles RNN et Naïve bayes.

# *Conclusion générale*

### *Conclusion générale*

Le contrôle des rumeurs est l'un des principaux défis de l'ère de l'information numérique où les rumeurs peuvent se propager rapidement et affecter la société de manière significative.

Ce projet de fin d'études présente une méthode de suivi et de détection des rumeurs dans les sites de médias sociaux en utilisant des approches d'apprentissage automatique et d'apprentissage profond.

Dans le premier chapitre, nous avons étudié les réseaux sociaux, les rumeurs et leurs caractéristiques. Le deuxième chapitre, nous avons parlé de l'apprentissage automatique et de l'apprentissage profond, des algorithmes utilisés et des différences entre eux. Le troisième chapitre, nous avons parlé de la conception du système, en décrivant l'architecture générale et l'architecture détaillée, qui est le traitement initial des textes dans la base de données et l'apprentissage pour la formation du système. Puis, la phase de test. Enfin, dans le dernier chapitre, nous avons commencé par définir quelques concepts sur les moyens utilisés, puis nous avons décrit les modèles développés en se basant sur les algorithmes RNN, LSTM et Naïve Bayes et expliqué les résultats obtenus. En plus, faire une comparaison entre les performances de ces modèles.



*Références*

*Bibliographies*

### *Bibliographies*

- [1] C. Hill, «SOCIAL NETWORKING,» April 2017.
- [2] Tzung-Pei Hong, Deep Learning for Social Media Data Analytics, Springer, 2022.
- [3] F. Tönnies, Community and Society, East Lansing, Michigan State University Press, 1887.
- [4] E. Durkheim, De la division du travail social: étude sur l'organisation des sociétés supérieures, New York: Free Press, 1893.
- [5] G. Simmel, Soziologie, Leipzig: Duncker & Humblot., 1908.
- [6] P. J. Carrington et J. Scott, The Sage Handbook of Social Network Analysis., 2011.
- [7] J. P. Scott, Social Network Analysis: A Handbook (2nd edition), Thousand Oaks, CA: Sage Publications., 2000.
- [8] B. Malinowski, The Family Among the Australian Aborigines: A Sociological Study. London, University of London Press, 1913.
- [9] A. R. Radcliffe-Brown, The social organization of Australian tribes, Sydney, Australia: University of Sydney Oceania monographs, 1930.
- [10] A. R. Radcliffe-Brown, On social structure". Journal of the Royal Anthropological Institute, 1940.
- [11] C. Lévi-Strauss, Les structures élémentaires de la parenté, Paris: Boston: Beacon Press, 1967.
- [12] J. Barnes, Class and Committees in a Norwegian Island Parish, Human Relations., 1954.
- [13] L. C. Freeman et B. Wellman, A note on the ancestral Toronto home of social network analysis, Connections, 1995.
- [14] M. Savage, Elizabeth Bott and the formation of modern British sociology, The Sociological Review, 2008.
- [15] S. F. Nadel, The Theory of Social Structure, London: Cohen and West., 1957.
- [16] T. Parsons, The Structure of Social Action: A Study in Social Theory with Special Reference to a Group of European Writers, New York: The Free Press., 1949.
- [17] T. Parsons, The Social System, New York: The Free Press, 1951.
- [18] P. Blau, Bureaucracy in Modern Society, New York: Random House, Inc., 1956.
- [19] P. Blau, «"A Theory of Social Integration,» 1960.
- [20] B. Hogan, The Networked Individual: A Profile of Barry Wellman, 2012.
- [21] M. Granovetter, Introduction for the French Reader, Sociologica.
- [22] B. Wellman, Structural analysis: From method and metaphor to theory and substance, Cambridge, UK: Cambridge University Press, 1988.
- [23] «Social network,» [En ligne]. Available: [https://en.wikipedia.org/wiki/Social\\_network#cite\\_note-Introduction\\_for\\_the\\_French\\_Reader-26](https://en.wikipedia.org/wiki/Social_network#cite_note-Introduction_for_the_French_Reader-26).
- [24] «Social Network Analysis,» 2012. [En ligne]. Available: <https://www.techopedia.com/definition/3205/social-network-analysis-sna>.
- [25] «Social-Network-Analysis,» [En ligne]. Available: <https://github.com/chrisbetze/Social-Network-Analysis>.

- [26] «Advantages and Disadvantages of Social Networks,» [En ligne]. Available: <https://www.linkedin.com/pulse/advantages-disadvantages-social-networks-johan-smith>.
- [27] «Rumor- Tracking-and-Management-Guide-Developed-by-Internews,» [En ligne]. Available: <https://covid19communicationnetwork.org/wp-content/uploads/2022/05/Rumor-Trackingand-Management-Guide-Developed-by-Internews-with-support-from-RCCE-0dfad167.pdf>.
- [28] R. Z. Xinyi Zhou, «A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities».
- [29] M. O. P. Christine DU, intelligence économique et management des connaissances, thèse professionnelle.
- [30] N. K. a. J. Voas, The Economics of “Fake News”, IT Professional, 2017.
- [31] F.-B. HUYGHE, Faux, rumeurs et désinformation dans le cyberspace, Edition L’IRIS, 2013.
- [32] A. e. E. M. Adeline Michel, les rumeurs en tant que phénomène..
- [33] S. M. A. P. L. M. a. T. S. Gustavo Aguilar, «A multi-task approach for named entity recognition in social media data,» 2017.
- [34] L. Audubon, «Le Deep Learning: définition, histoire et cas d'usage,» 2022.
- [35] «deep-learning,» [En ligne]. Available: <https://www.sicara.fr/fr/parlons-data/deep-learning>.
- [36] «deep-learning-tutorial,» [En ligne]. Available: <https://www.geeksforgeeks.org/deep-learning-tutorial/>.
- [37] «deep-learning-applications,» [En ligne]. Available: <https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-applications>.
- [38] «Objectif IA : initiez-vous à l'intelligence artificielle,» openclassrooms. [En ligne]. Available: <https://openclassrooms.com/fr/courses/6417031-objectif-ia-initiez-vous-a-lintelligence-artificielle>.
- [39] «deep-learning-vs-machine-learning,» [En ligne]. Available: <https://flatironschool.com/blog/deep-learning-vs-machine-learning/>.
- [40] «demystifying-deep-learning-and-artificial-intelligence,» [En ligne]. Available: <https://thenewstack.io/demystifying-deep-learning-and-artificial-intelligence>.
- [41] «neural-network,» [En ligne]. Available: <https://aws.amazon.com/fr/what-is/neural-network/>.
- [42] J. Z. A. J. H. Laith Alzubaidi, «Review of deep learning: concepts, CNN architectures, challenges, applications, future,» 2021.
- [43] A. Plaet, Deep Reinforcement Learning, Springer, 2022.
- [44] «Advantages and Disadvantages of Deep Learning,» [En ligne]. Available: <https://www.ssttek.com/advantages-and-disadvantages-of-deep-learning/>.
- [45] «Introduction-to-recurrent-neural-network,» [En ligne]. Available: <https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/>.
- [46] «Understanding LSTM Networks,» 2015. [En ligne]. Available: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- [47] A. Mikami, «Long Short-Term Memory Recurrent Neural Network Architectures for Generating Music and Japanese Lyrics,» Computer Science Department, Boston College, 2016.
- [48] «Recurrent neural networks and Long-short term memory (LSTM),» [En ligne].

- Available: [https://people.cs.pitt.edu/~jlee/papers/cs3750\\_rnn\\_lstm\\_slides.pdf](https://people.cs.pitt.edu/~jlee/papers/cs3750_rnn_lstm_slides.pdf).
- [49] «Naive-bayes,» [En ligne]. Available: <https://deepai.org/machine-learning-glossary-and-terms/naive-bayes>.
- [50] P. K. Benjamin Obi Tayo, «Exploring Data Cleaning Techniques With Python,» 2023. [En ligne]. Available: <https://www.kdnuggets.com/2023/04/exploring-data-cleaning-techniques-python.html#:~:text=Data%20cleaning%20refers%20to%20the>.
- [51] geeksforgeeks, «vectorization-in-python,» [En ligne]. Available: <https://www.geeksforgeeks.org/vectorization-in-python/amp>.
- [52] datacamp, «python-string-split,» [En ligne]. Available: <https://www.datacamp.com/tutorial/python-string-split>.
- [53] datacamp, «what-is-tokenization,» [En ligne]. Available: <https://www.datacamp.com/blog/what-is-tokenization>.
- [54] «Bienvenue dans Colab !,» [En ligne]. Available: <https://colab.research.google.com/>.
- [55] scikit-learn, «scikit-learn Machine Learning in Python,» [En ligne]. Available: <https://scikit-learn.org/stable/index.html>.
- [56] «pandas,» [En ligne]. Available: <https://pandas.pydata.org/>.
- [57] «numpy,» [En ligne]. Available: <https://numpy.org/>.
- [58] «tensorflow,» [En ligne]. Available: <https://www.tensorflow.org/?hl=ar>.
- [59] «matplotlib,» [En ligne]. Available: <https://matplotlib.org/> (.).
- [60] J. Scott, *Social Network Analysis: A Handbook* (2nd edition), Thousand Oaks, CA: Sage Publications., 2000.
- [61] P. Blau, *Exchange and Power in Social Life*, 1964.
- [62] H. A. a. M. Gentzkow, «Social media and fake news in the 2016,» 2017.
- [63] A. S. S. W. J. T. a. H. L. Kai Shu, «Fake news detection on social media: A data mining perspective.,» 2017.
- [64] D. R. a. S. A. Soroush Vosoughi, «The spread of true and false news online,» 2018.
- [65] D. B. a. D. A. Schwartz, «Miley, CNN and The Onion: When fake news becomes realer than real. *Journalism Practice*,» 2016.
- [66] N. K. a. J. Voas, «The Economics of “Fake News”,» 2017.
- [67] A. Kucharski., «Post-truth: Study epidemiology of fake news,» 2016.
- [68] C. B. a. J. Golbeck., «Automatically Identifying Fake News in Popular Twitter Threads. In *Smart Cloud*,» 2017.